CODEN: JASMAN

The Journal of the Acoustical Society of America

ISSN: 0001-4966

Vol. 121, No. 1 January 2007

(Continued)

CONTENTS—*Continued from preceding page*

(Continued)

CONTENTS—*Continued from preceding page*

(Continued)

CONTENTS—*Continued from preceding page*

On the potential limitations of conventional sound metrics in quantifying perception of nonlinearly propagated noise

Kent L. Gee and S. Hales Swift
Department of Physics and Astronomy, N319 ESC, Brigham Young University, Provo, Utah 84602
kentgee@byu.edu; hales.swift@gmail.com

Victor W. Sparrow
Graduate Program in Acoustics, 201 Applied Science Building, The Pennsylvania State University, University Park, Pennsylvania 16802
vws1@psu.edu

Kenneth J. Plotkin
Wyle Laboratories, 2001 Jefferson Davis Highway, Suite 701, Arlington, Virginia 22202
kplotkin@wylelabs.com

J. Micah Downing
Blue Ridge Research and Consulting, LLC, 13 1/2 West Walnut Street, Asheville, North Carolina 28801
micah.downing@blueridgeresearch.com

Abstract: The use of conventional metrics to quantify the perception of nonlinearly propagated noise has been studied. Gaussian noise waveforms have been numerically propagated both linearly and nonlinearly, and from the resulting waveforms, several metrics are calculated. These metrics are overall, A-, C-, and D-weighted sound pressure levels, perceived noise level, Stevens Mark VII perceived loudness, Zwicker loudness, and sharpness. Informal listening demonstrations indicate that perceived differences in annoyance between linearly and nonlinearly propagated waveforms are substantial. Because the metrics studied seem inadequate in representing the perceived differences, rigorous subjective testing is encouraged to properly quantify and understand these differences.

PACS numbers: 43.50.Ba, 43.66.Lj, 43.25.Cb [MRS]
Date Received: August 14, 2006 **Date Accepted:** October 20, 2006

1. Introduction

The effects of noise radiated by high-performance jet aircraft on nearby communities and the environment is a research question that has been studied in a variety of contexts. One relevant issue related to the noise impact of these aircraft is the possible influence of nonlinear propagation effects. Pernet and Payne,[1] who performed experiments on nonlinear noise traveling in a long tube, were motivated by anomalous propagation effects previously observed in jet noise studies. Webster and Blackstock[2] later conducted outdoor experiments with high-amplitude noise propagation and offered evidence that nonlinearity also affected the propagation of noise from high-performance aircraft. More recent studies have provided verification that nonlinearity can significantly impact high-amplitude jet noise propagation (e.g., see Ref. 3) and efforts are underway to incorporate nonlinear effects in noise analysis models for these aircraft.[4]

The nonlinear propagation of a high-amplitude noise waveform may be described in the context of either the time or frequency domain. In the time domain, nonlinearity causes steepening of the high-amplitude portions of the waveform and the possible coalescence of acoustic shocks.[5] In the frequency domain, these time domain effects correspond to a spectral broadening as cascading sum- and difference-frequency generation occurs. These effects could

significantly alter the perception of a waveform from that predicted by linear propagation. Crighton[6] further contextualized the potential impact of nonlinear jet noise propagation when he pointed out that a nonlinear transfer of energy to high frequencies could impact calculations of noise metrics that penalize these frequencies.

The purpose of this letter is to begin to address the issue of the perception of nonlinear effects in the atmospheric propagation of broadband noise with the intent of motivating additional improvement in environmental impact modeling for high-performance aircraft noise. Specifically, two questions are addressed. First, if nonlinear effects are important in high-amplitude noise propagation, does perception of the noise change as a result of nonlinear propagation? Second, do common metrics adequately account for any change in perception due to nonlinearity?

In this letter, results are presented from a study in which a waveform with a shaped broadband spectrum has been numerically propagated both linearly and nonlinearly. The resultant waveforms may be heard by the reader. From the linearly and nonlinearly predicted waveforms, several single-number metrics commonly used in the environmental noise and sound quality communities have been calculated. The metrics calculated are overall, A-, C-, and D-weighted sound pressure level, perceived noise level, Stevens Mark VII perceived loudness, Zwicker loudness, and sharpness. Although the list is by no means exhaustive, the results demonstrate a possible shortcoming of traditional metrics when used to quantify the readily perceived difference between the nonlinearly and linearly propagated waveforms.

2. Metrics considered

The metric that is most commonly used to calculate the perceptual impact of an acoustic source is A-weighted sound pressure level, L_A. However, because the A-weighting curve is based on the 40-phon equal loudness contour and observed jet noise levels are often much greater than 40 phon, C-weighted sound pressure level (L_C), which is based on the 90-phon contour, has also been calculated. Another weighting that has been used specifically for aircraft noise is D-weighting, which is based on work by Kryter[7] and emphasizes annoyance or noisiness caused by high-frequency energy in the range of 1–12 kHz. Although D-weighting was formally standardized,[8] that standard has since been withdrawn and the weighting is not frequently used now. However, because of its intended application to aircraft noise and its particular emphasis on high-frequency content, which could be important for the perception of nonlinear effects, D-weighted sound pressure level (L_D) has also been calculated in this study. In addition, ordinary (nonweighted) overall sound pressure level, L, has also been calculated.

An additional metric that has been calculated is perceived noise level (PNL), which is also based on Kryter's work with the perception of noisiness.[9,10] (PNL) is currently used by the Federal Aviation Administration as part of the calculation of effective perceived noise level (EPNL), the standard metric for noise certification of commercial aircraft. PNL, rather than EPNL, has been calculated because the latter metric is intended for transient waveforms measured from flyovers.

Two other metrics calculated are Stevens Mark VII (Ref. 11) perceived loudness (PL) and Zwicker loudness[12] (ZL). Mark VII perceived loudness calculates the loudness in third-octave bands according to estimated inverses of the equal loudness contours. The loudness for each band is then summed with additional weight being given to the loudest band. Zwicker loudness calculates the specific loudness in each critical band and then sums them to find the total loudness. Both PL and ZL have been used extensively to assess loudness and annoyance in various sonic boom studies.[13]

The final metric that is calculated is sharpness (S), which is a psychoacoustical quality that increases according to the relative amount of high-frequency energy in a signal. Sharpness is viewed as a negative indicator of sensory pleasantness; thus sharp sounds are frequently felt to be more annoying or unpleasant. Although there are different proposed methods for calculation of sharpness,[14–17] the method selected for this study is that of Zwicker and Fastl,[14] who calculate sharpness based on a weighted sum of specific loudness.

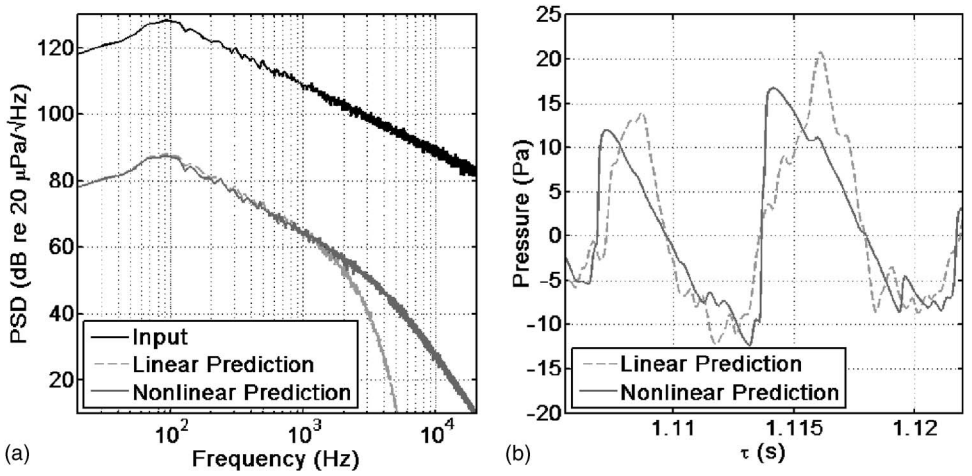


Fig. 1. (a) Input power spectral density at 10 m and predicted power spectral densities at 1000 m. (b) Short segments of the nonlinearly and linearly predicted waveforms at 1000 m as a function of retarded time, τ. Note the shock-like steepness of the nonlinearly predicted waveform at approximately 1.107 and 1.114 s.

Before proceeding, it is important to note that some of these metrics (L, L_A, L_C, PL, and ZL) are intended to quantify the subjective impression of loudness while the remaining metrics (L_D, PNL, and S) are designed to correlate with annoyance or perceived noisiness. However, loudness is a significant contributor to overall annoyance and L_A, PL, and ZL are often used to quantify the overall perceptual impact of sounds. The merits and shortcomings of many of these metrics (e.g., L_A) have been frequently discussed and debated for various applications and an extensive list of references could be given. However, for the purposes of this investigation, all the metrics have been placed on equal footing in their presumed ability to quantify an overall impression of a noise waveform, regardless of their original intent to quantify loudness or noisiness.

3. Test case description and results

3.1 Test case description

For this study, an initially Gaussian waveform with a shaped broadband spectrum has been numerically propagated with a model[18] that solves the generalized Burgers equation (GBE),[19] a widely used nonlinear model equation. The particular GBE used in this research accounts for the phenomena of quadratic nonlinearity, atmospheric absorption and dispersion, and spherical spreading. Free-field linear propagation of the waveform has also been carried out by simply removing the nonlinear term from the GBE. The shape of the power spectral density (PSD) for the input waveform, shown in Fig. 1(a), has been chosen to have a 6 dB/octave slope below the peak frequency of approximately 100 Hz and a −6 dB/octave slope above the peak frequency in order to simulate a jet mixing noise spectrum. The overall sound pressure level for the input waveform is $L=150$ dB re 20 μPa at an assumed input distance of 10 m, a reasonable level for a military jet aircraft.[20]

The input waveform, which consists of 524,288 points sampled at 200 kHz, has been propagated with both the nonlinear and linear numerical models out to a distance of 1000 m. Uniform atmospheric conditions of 1 atm, 20 °C, and 50% have been assumed for ambient pressure, temperature, and relative humidity. The calculated PSD (with a frequency resolution of 12 Hz) for each of the predicted waveforms is also shown in Fig. 1(a), where a nonlinear transfer of energy to high frequencies at the expense of the energy in peak-frequency region partially mitigates the expected roll-off due to ordinary linear atmospheric absorption. The high-frequency energy transfer is also evident in a comparison of short segments of the numeri-

Table 1. Calculated metrics and difference for the nonlinearly and linearly predicted waveforms. L denotes overall sound pressure level and the subscripts signify the type of weighting applied. PNL, PL, ZL, and S, respectively, represent perceived noise level, Stevens Mark VII perceived loudness, Zwicker loudness, and sharpness.

	L (dB)	L_A (dB A)	L_C (dB C)	L_D (dB D)	PNL (PN dB)	PL (PL dB)	ZL (phon)	S (acum)
Nonlinear	108.4	98.0	107.4	104.1	111.4	102.5	105.5	1.01
Linear	109.0	97.9	107.9	104.0	110.2	101.5	103.9	0.72
Difference	−0.6	0.1	−0.5	0.1	1.2	1.0	1.6	0.29

cally propagated waveforms, which are displayed in Fig. 1(b). The relatively steep portions of the nonlinearly predicted waveform compared to the linear waveform are the source of the additional high-frequency energy in the nonlinearly predicted PSD in Fig. 1(a).

3.2 Waveform playback

A critical aspect of this letter is the ability to listen to the input and output waveforms, which have been resampled at 44.1 kHz for convenience. As a reference, the shaped Gaussian input waveform may be heard by clicking on the link to Mm. 1. The linearly and nonlinearly predicted waveforms may be heard in Mm. 2 and Mm. 3, respectively. Note that the effect of spherical spreading has been removed from the predicted waveforms so that these and the input waveform may be heard and compared for a single audio playback level.

Mm. 1. Input.wav (225 kB). Shaped Gaussian waveform used as an input to the numerical models.

Mm. 2. Linear.wav (225 kB). Linearly predicted waveform at 1000 m.

Mm. 3. Nonlinear.wav (225 kB). Nonlinearly predicted waveform at 1000 m.

Playback of these three waveforms reveals two noteworthy points. First, the lowpass-filter effect of atmospheric absorption is heard when the linearly predicted waveform in Mm. 2 is compared to the input waveform in Mm. 1. Second, the perceptual impact of nonlinear propagation may be clearly heard by comparing Mm. 3 with the linear waveform in Mm 2. There is a staccato-like, impulsive quality of the nonlinearly propagated waveform that is also present in some far-field jet noise recordings and appears to be related to the presence of shock-like structures in the waveform.[21] In informal listening demonstrations carried out by the authors, the nonlinear waveform is usually perceived to be somewhat louder but significantly more annoying than the linear waveform, although these are admittedly subjective statements that should eventually be quantified by rigorous jury testing. However, because there appears to be a considerable difference in perception between the nonlinearly and linearly propagated waveforms, regardless of subjective descriptors used, a study of which metrics also exhibit significant difference between the waveforms has been performed.

4. Metric calculations and discussion

The metrics described previously in Sec. II have been calculated for the linearly and nonlinearly predicted waveforms and are displayed in Table 1. Also shown in Table 1 is the difference between the nonlinear and linear predictions for each of the metrics. The result for overall sound pressure level, L, indicates that the nonlinear spectrum has slightly less energy than the linear spectrum. This result is generally expected for atmospheric nonlinear propagation because energy is transferred to higher frequencies where absorption coefficients are greater. Although the difference in L for this test case (0.6 dB) is likely negligible, one could potentially argue that nonlinear effects might cause an aircraft to be perceived as quieter than for ordinary linear

propagation as a consequence of the additional energy losses. Of course, L does not account for the nonuniform response of the ear and so calculation of the other metrics better addresses this issue.

Table 1 also shows the calculated results for the weighted sound pressure levels, L_A, L_C, and L_D, and the more sophisticated measures, PNL, PL, and ZL. Because of its greater weighting of low frequencies, L_C gives an almost identical difference between nonlinear and linear propagation as L. The other weighted sound pressure levels, L_A and L_D, only show differences of 0.1 dB and therefore indicate that the nonlinear and linear waveforms should be perceived essentially the same. Finally, PNL, PL, and ZL, all exhibit somewhat greater differences between the nonlinearly and linearly predicted waveforms (1.2 PN dB, 1.0 PL dB and 1.6 phon, respectively) and suggest that the nonlinear waveform should be perceived as slightly more annoying or louder. However, these differences are still relatively small and appear to be insufficient to accurately represent the perceived difference between the two waveforms.

The final metric shown in Table 1 is that of sharpness, S. Because the unit of S, the acum, is intended to vary linearly with perception of sharpness, the nonlinear waveform's sharpness is approximately 40% greater than that of the linear waveform. This difference could likely be perceived as significant. However, before sharpness is considered to be a candidate quantifier of nonlinearly propagated noise, there is another shortcoming of S (and, in fact, of all the metrics) that should be discussed.

A fundamental difficulty with sharpness and the other metrics calculated is that they rely solely on averaged power spectral calculations. Because phase is neglected, these metrics do not uniquely characterize perception of a waveform. Although phase is typically ignored when quantifying perception of noise, formation of an acoustic shock through nonlinear propagation requires a specific phase relationship between Fourier spectral components of a complex waveform. If this phase relationship is altered, so are the time waveform and the perception of an acoustic shock. In other words, if the shock-like structures are critical to the overall perception of the nonlinearly propagated waveform, modification of the Fourier phase spectrum of the waveform could significantly alter how it is perceived.

To illustrate this point, the Fourier phase spectrum for the nonlinearly propagated waveform in Mm. 3 has been randomized with uniform probability. This process results in a time waveform that has a Gaussian probability density function but possesses the same PSD as the original nonlinearly propagated waveform. This phase-modified waveform may be heard in Mm. 4 and is perceived to be different than the original nonlinearly propagated waveform. Because each of the metrics discussed, including sharpness, responds identically to Mm. 3 and Mm. 4, none can completely quantify the perceptual impact of nonlinear propagation.

Mm. 4. Phase-modified Nonlinear.wav (225 kB). Nonlinearly predicted waveform at 1000 m with its Fourier phase randomized.

One final point of discussion regards the "staccato-like impulsive" quality of the nonlinearly propagated waveform in Mm. 3. For the traditional classification of impulsive sounds (explosions, sonic booms, gunfire, etc.), an annoyance penalty is given when assessing perceptual impact due to their short duration and high intensity.[22] If the nonlinearly propagated waveform can be treated as conventional impulsive noise, then there is already a standardized protocol for addressing its perception. However, the nonlinear noise phenomenon differs from traditional impulsive noise because the pressure fluctuations are part of a steady-state noise signal and not a sudden deviation from ambient levels (as occurs in an explosion). In addition, nonlinearity actually causes the pressure excursions in the nontransient nonlinearly propagated waveform to be of lower amplitude than the linearly propagated waveform [see Fig. 1(b)], which does not possess an impulsive quality. Finally, the perceptual difference between the nonlinearly propagated waveform in Mm. 3 and the phase randomized version in Mm. 4 is clear, but conflicts with the conclusion of Fidell *et al.* that phase randomization of impulsive sounds does not alter perception of noisiness.[23] Additional research is therefore needed to better quantify the perception of the unique phenomenon of nonlinearly propagated noise.

5. Conclusion

This letter has sought to increase awareness regarding the potential limitations of various noise and sound quality metrics when used to quantify perception of nonlinearly propagated noise. Both the waveforms and the metric comparisons have direct implications regarding environmental impact assessment of high-performance jet aircraft. Playback of the waveforms indicates that nonlinear propagation could cause these aircraft to be perceived as more annoying than predicted by models that assume ordinary linear propagation. However, this perceptual difference may not be observed with conventional single number metrics because the weighted sound pressure levels, perceived noise level, Stevens Mark VII loudness, and Zwicker loudness do not appear to adequately penalize the annoyance caused by the high-frequency energy present in the nonlinear waveform. Although the sound quality metric, sharpness, does indicate some appreciable perceived difference between nonlinearly and linearly predicted waveforms, the neglect of phase information important to nonlinear propagation limits its utility as a viable metric for this application.

It is clear that additional research regarding the perceptual impact of nonlinearly propagated noise is needed. Although the multimedia content of this letter makes possible informal listening demonstrations, rigorous subjective testing is required to quantify perceived differences between nonlinearly and linearly propagated noise. Subjective testing could result in the formulation of an alternative metric and thereby improve environmental impact assessments of high-power jet aircraft.

Acknowledgment

This work was supported in part by the Strategic Environmental Research and Development Program.

References and links

[1]D. F. Pernet and R. C. Payne, "Non-linear propagation of signals in air," J. Sound Vib. **17**, 383–396 (1971).
[2]D. A. Webster and D. T. Blackstock, "Experimental investigation of outdoor propagation of finite-amplitude noise," *NASA Contractor Report 2992* (Applied Research Laboratories, The University of Texas at Austin, 1978).
[3]K. L. Gee, V. W. Sparrow, M. M. James, J. M. Downing, C. M. Hobbs, T. B. Gabrielson, and A. A. Atchley, "Measurement and prediction of noise propagation from a high-power jet aircraft," AIAA Pap. 2006–2931 (2006).
[4]B. J. Iklheimer, "Advanced simulation noise model for modern fighter aircraft," in *CD-ROM: Proceedings of Noise-Con 05*, edited by J. S. Bolton, P. Davies, and G. C. Maling, Jr. (Institute of Noise Control Engineering of the USA, Washington D.C., 2005), Paper No. nc05_209.
[5]S. N. Gurbatov and O. V. Rudenko, "Statistical phenomena," in *Nonlinear Acoustics*, edited by M. F. Hamilton and D. T. Blackstock (Academic, San Diego, 1998), Chap. 13, pp. 377–398.
[6]D. G. Crighton, "Nonlinear acoustic propagation of broadband noise," in *Recent Advances in Aeroacoustics*, edited by A. Krothapalli and C. A. Smith (Springer, New York, 1986), pp. 411–454.
[7]K. D. Kryter, "Concepts of perceived noisiness, their implementation and application," J. Acoust. Soc. Am. **43**, 344–361 (1968).
[8]"Frequency weighting for the measurement of aircraft noise (D-weighting)," IEC 537 (International Electrotechnical Commission, Geneva, Switzerland, 1976).
[9]K. D. Kryter, "Scaling human reaction to sound from aircraft," J. Acoust. Soc. Am. **31**, 1415–1429 (1959).
[10]K. D. Kryter and K. S. Pearsons, "Some effects of spectral content and duration on perceived noise level," J. Acoust. Soc. Am. **35**, 866–883 (1963).
[11]S. S. Stevens, "Perceived level of noise by Mark VII and decibels (E)," J. Acoust. Soc. Am. **51**, 575–601 (1972).
[12]"Acoustics—Method for calculating loudness level," ISO 532 (International Organization for Standardization, Geneva, Switzerland, 1975).
[13]J. D. Leatherwood, B. M. Sullivan, K. P. Shepherd, D. A. McCurdy, and S. A. Brown, "Summary of recent NASA studies of human response to sonic booms," J. Acoust. Soc. Am. **111**, 586–598 (2002).
[14]E. Zwicker and H. Fastl, *Psychoacoustics: Facts and models*, 2nd ed. (Springer, New York, 1999).
[15]G. von Bismarck, "Sharpness as an attribute of the timbre of steady state sound," Acustica **30**, 159–172 (1974).
[16]W. Aures, "Berechnungsverfahren für den sensorischen Wohlklang beliebiger Schallsignale," Acustica **59**, 130–141 (1985).
[17]A. Marui and W. L. Martens, "Predicting perceived sharpness of broadband noise from multiple moments of the specific loudness distribution," J. Acoust. Soc. Am. **119**, EL7–EL13 (**2006**).

[18]K. L. Gee, "Prediction of nonlinear jet noise propagation," Ph.D. thesis, Graduate Program in Acoustics, The Pennsylvania State University, University Park, PA, 2005.
[19]M. F. Hamilton and C. L. Morfey, "Model equations," in *Nonlinear Acoustics*, edited by M. F. Hamilton and D. T. Blackstock (Academic, San Diego, 1998), Chap. 3, pp. 41–63.
[20]K. L. Gee, T. B. Gabrielson, A. A. Atchley, and V. W. Sparrow, "Preliminary analysis of nonlinearity in military jet aircraft noise propagation," AIAA J. **43**, 1398–1401 (2005).
[21]K. L. Gee and V. W. Sparrow, "Quantifying nonlinearity in the propagation of noise from military jet aircraft," in *CD-ROM: Proceedings of Noise-Con 05*, edited by J. S. Bolton, P. Davies, and G. C. Maling, Jr. (Institute of Noise Control Engineering of the USA, Washington D.C., 2005), Paper No. nc05_194.
[22]"Quantities and procedures for description and measurement of environmental sound—Part 4: Noise assessment and prediction of long-term community response," ANSI S12.9-2005/Part 4 (American National Standards Institute, Washington, DC, 2005).
[23]S. Fidell, K. S. Pearsons, M. Grignetti, and D. M. Green, "The noisiness of impulsive sounds," J. Acoust. Soc. Am. **48**, 1304–1310 (1970).

Bayesian estimation of the underlying bone properties from mixed fast and slow mode ultrasonic signals

Karen R. Marutyan
Department of Physics, Washington University in St. Louis, One Brookings Drive, St. Louis, Missouri 63130
marutyan@hbar.wustl.edu

G. Larry Bretthorst
Department of Radiology, Washington University, One Brookings Drive, St. Louis, Missouri 63130
larry@bayes.wustl.edu

James G. Miller
Department of Physics, Washington University in St. Louis, One Brookings Drive, St. Louis, Missouri 63130
james.g.miller@wustl.edu

Abstract: We recently proposed that the observed apparent negative dispersion in bone can arise from the interference between fast wave and slow wave modes, each exhibiting positive dispersion [Marutyan *et al.*, J. Acoust. Soc. Am. **120**, EL55–EL61 (2006)]. In the current study, we applied Bayesian probability theory to solve the inverse problem: extracting the underlying properties of bone. Simulated mixed mode signals were analyzed using Bayesian probability. The calculations were implemented using the Markov chain Monte Carlo with simulated annealing to draw samples from the marginal posterior probability for each parameter.

PACS numbers: 43.80.Ev, 43.80.Jz, 43.80.Qf [CC]
Date Received: September 21, 2006 **Date Accepted:** October 20, 2006

1. Introduction

Measurements of the ultrasonic characteristics of bone to assess the risk of fracture due to osteoporosis are often based on the frequency dependence of attenuation (Droin *et al.*, 1998; Langton *et al.*, 1984; Wear and Armstrong, 2001). Bone attenuates ultrasound in a manner that is approximately linear with frequency (Droin *et al.*, 1998; Wear, 2000a) so that the attenuation characteristics can be reported as a rate of change with frequency of the attenuation coefficient times the distance traveled, known as broadband ultrasound attenuation (BUA), or, if normalized with bone thickness, the normalized broadband ultrasound attenuation (nBUA). This parameter is also known as the slope of attenuation. For media with an attenuation coefficient that can be approximated as increasing linearly with frequency, such as the attenuation coefficient in bone, the nearly local approximation to the Kramers-Kronig relations with one subtraction (Waters *et al.*, 2000) suggests that an increase of velocity with frequency would be expected. If the attenuation coefficient were strictly proportional to frequency, the dispersion would increase approximately logarithmically with frequency. In contrast with the anticipated results, the frequency dependence of phase velocity (dispersion) measured in a number of laboratories indicates that, on average, the phase velocity of ultrasonic waves propagating in cancellous bone decreases with increasing frequency (negative dispersion). However, a positive dispersion is observed in approximately 10% to 20% of sites measured (Droin *et al.*, 1998; Nicholson *et al.*, 1996; Strelitzki and Evans, 1996; Wear, 2000b). Although it is known that in general bone supports two compressional modes, a fast wave and a slow wave (Fellah *et al.*, 2004; Hosokawa and Otani, 1997), in almost all studies that reported negative dispersion only a single wave was observed. We recently proposed that apparent negative dispersion can result from the interference between fast wave and slow wave modes, each of which propagates with the anticipated

positive dispersion (Marutyan *et al.*, 2006). Results from simulations based on this hypothesis appear to be consistent with measurements of both negative and positive dispersions depending upon specific details of the site measured.

The purpose of this letter is to demonstrate that the true ultrasonic characteristics underlying the fast and slow modes, including their frequency-dependent attenuations and positive dispersions, can be recovered from the received mixed mode signal using a Bayesian approach. One motivation for these investigations is the assumption that identifying the true material properties that underlie the mixed mode signal will provide a more reliable estimate of bone quality and fracture risk than the apparent properties derived from the mixed mode signal.

The model for ultrasonic propagation in bone, which represents the "forward problem" that was addressed in a previous publication (Marutyan *et al.*, 2006), is described in Sec. 2. The methods used to address the "inverse problem," which is the subject of the current work, are outlined in Sec. 3. There we describe the basis of the Bayesian analysis applied and outline some specifics of the computational implementation. In Sec. 4 the results retrieved by Bayesian analysis are examined for a range of input parameters. We close with a discussion of the limitations of the current study.

2. Model for ultrasonic propagation in bone

We model ultrasonic propagation in bone by considering two independent modes: one with the characteristics of a fast wave and the other with the characteristics of a slow wave. As such, our model is not restricted to any specific bone orientation relative to the load bearing direction. Consequences of the anisotropy of bone are therefore accounted for by the corresponding numerical values of the attenuation coefficient and the phase velocity.

Input values for the slope of attenuation for both the fast and slow modes were selected from the published literature (Waters and Hoffmeister, 2005). In that study cancellous bone specimens were obtained from bovine tibia. Values for nBUA in our simulations were set to be $\beta_{\text{fast}}=2.3\times10^{-4}$ Np/m/Hz$\approx$20.0 dB/cm/MHz and $\beta_{\text{slow}}=8.0\times10^{-5}$ Np/m/Hz$\approx$6.9 dB/cm/MHz for the fast and slow wave, respectively. These values are for a specimen oriented in the superior-inferior direction. Logarithmic dispersions proportional to those values of slopes of attenuation were calculated using the nearly local approximation of the Kramers-Kronig relations. Because the Kramers-Kronig relations are ambiguous to a constant offset, the values of the phase velocity at 300 kHz (v_{fast} and v_{slow}) were arbitrarily set to 2100 and 1500 m/s for the fast and the slow wave, respectively. Other choices for the phase velocities were also explored. Variations in v_{fast} from 1900 to 2500 m/s were shown not to significantly alter the outcome of the simulations. These values for velocity are consistent with the results reported by Hosokawa and Otani (1997).

Once the ultrasonic parameters for sound propagation in bone were established, the frequency domain representation of the output waveform assuming linear plane propagation was obtained in the following manner:

$$\text{Output}(\omega) = w \cdot \text{Input}(\omega) \cdot H_{\text{fast}}(\omega,l) + (1-w) \cdot \text{Input}(\omega) \cdot H_{\text{slow}}(\omega,l) \tag{1}$$

where $\omega=2\pi f$ is a positive angular frequency, l is a bone thickness, w represents a weighting parameter that varies from 0 to 1, and Input (ω) is the frequency spectrum of the input waveform with center frequency of 550 kHz and −6 dB bandwidth spanning the frequency range of 250 to 850 kHz typical of that used for measurements on bone. The transfer functions $H_{\text{fast}}(\omega,l)$ and $H_{\text{slow}}(\omega,l)$ for the fast and slow modes are given by

$$H_{\text{fast/slow}}(\omega,l) = \exp(-\alpha_{\text{fast/slow}}(\omega)l)\exp(i\omega l/c_{\text{fast/slow}}(\omega)), \tag{2}$$

where $\alpha_{\text{fast/slow}}(\omega)=\beta_{\text{fast/slow}}\omega/2\pi$ denotes the (linear with frequency) attenuation coefficient and $c_{\text{fast/slow}}(\omega)$ denotes the (logarithmically increasing) phase velocity for the fast and the slow wave, respectively. The time-domain radiofrequency (rf) sample traces were then obtained from the inverse Fourier transform of the propagated spectrum. As discussed in

Table 1. Prior probabilities assigned to model parameters.

	β_{fast} (dB/cm/MHz)	β_{slow} (dB/cm/MHz)	v_{fast} (m/s)	v_{slow} (m/s)	w
Low	0	0	2000	1000	0
Mean	26	26	2500	1700	0.5
High	43	43	4000	2000	1
Standard deviation	26	26	1000	1000	0.5

Sec. 4, Gaussian noise was added to the received signal in order to test the consequences of varying signal-to-noise ratios.

3. Bayesian inference

Bayesian methods are described in some detail in a volume by Jaynes and Bretthorst (2003). In the present brief report, we make use of Bayesian methods similar to those used in Bretthorst *et al.* (2005a) and Bretthorst *et al.* (2005b) to solve a somewhat different problem in nuclear magnetic resonance. Briefly, the calculation proceeded by applying Bayes' theorem to evaluate the joint posterior probability for all the parameters in the model

$$P(\beta_{\text{fast}}\beta_{\text{slow}}v_{\text{fast}}v_{\text{slow}}w|DI) = \frac{P(\beta_{\text{fast}}\beta_{\text{slow}}v_{\text{fast}}v_{\text{slow}}w|I)P(D|\beta_{\text{fast}}\beta_{\text{slow}}v_{\text{fast}}v_{\text{slow}}wI)}{P(D|I)}, \quad (3)$$

where D is a simulated data vector containing 4096 data points.

Equation (3) uses the conventional notation for conditional probabilities in which the vertical bar '|' means "given." Correspondingly, probabilities of the form $P(\text{left}|\text{right})$ should be understood as the joint probabilities for the items contained in *left* given that all items contained in *right* are true. The term $P(\beta_{\text{fast}}\beta_{\text{slow}}v_{\text{fast}}v_{\text{slow}}w|I)$ is the prior probability for the model parameters conditional on the information at hand, denoted as I. The prior probability for the data, $P(D|I)$, is a normalization constant and is not needed in the Monte Carlo simulations used to implement this calculation.

Assuming the prior probabilities for each parameter are logically independent, the joint prior was factored using the product rule of probability theory

$$P(\beta_{\text{fast}}\beta_{\text{slow}}v_{\text{fast}}v_{\text{slow}}w|I) = P(\beta_{\text{fast}}|I)P(\beta_{\text{slow}}|I)P(v_{\text{fast}}|I)P(v_{\text{slow}}|I)P(w|I). \quad (4)$$

All prior probabilities were assigned using bounded Gaussians to represent what is known about these parameters. (See Table 1 for assignments used in the simulations.) Changing the bounds, means, and standard deviations of the priors has little effect on the resulting parameter estimates.

The calculations were implemented using Markov chain Monte Carlo with simulated annealing. The simulation ran through two phases: an annealing phase during which the posterior probability for the model was computed using thermodynamic integration, and a parameter estimation phase during which samples were drawn from the joint posterior probability. Fifty Markov chain Monte Carlo simulations were run in parallel. After the completion of the annealing phase, 50 independent samples were drawn from each simulation, resulting in a total of 2500 samples for the joint posterior probability. Finally, the marginal posterior probability for each parameter was obtained by Monte Carlo integration of the joint posterior probability over all of the parameters except the one of interest. For example, the marginal posterior probability for β_{fast} was computed as

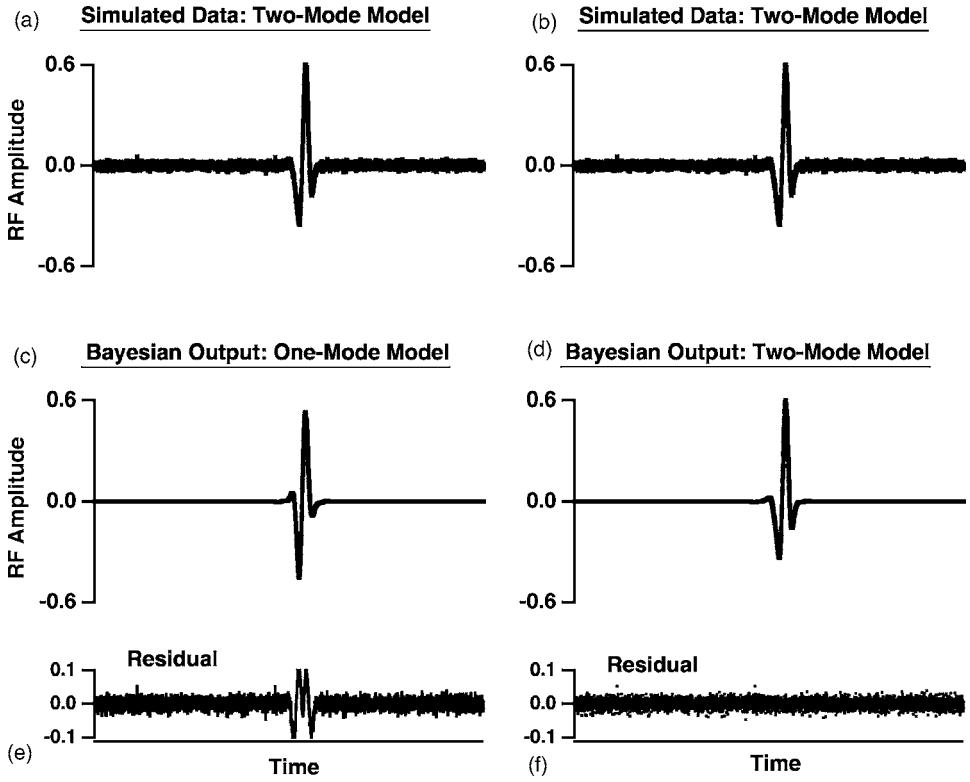


FIG. 1. Comparison of simulated data [a radio-frequency signal that is identical in panels (a) and (b)] with the Bayesian output for one-mode model [panel (c)] and two-mode model [panel (d)] computed from the parameters that maximized the joint posterior probability. Simulated data correspond to ultrasonic propagation over a 3-mm sample path for 30% fast wave (70% slow wave) mixed waveform content with added Gaussian noise (50:1 peak signal-to-noise ratio). The residuals for one- and two-mode models are plotted in panels (e) and (f), respectively.

$$P(\beta_{\text{fast}}|DI) = \int P(\beta_{\text{fast}}\beta_{\text{slow}}v_{\text{fast}}v_{\text{slow}}w|DI)d\beta_{\text{slow}}dv_{\text{fast}}dv_{\text{slow}}dw. \quad (5)$$

The program that implements the Bayesian calculations was run on a Sun Enterprise 250 dual 400-MHz workstation and took about 100 min to complete the analysis for one data set. Similar calculations took about 3 min using 32 processors of an SGI Altix 3000 with Itanium2 processors running at 900 MHz.

4. Results and discussion

The analysis was applied to the simulated sample rf traces for which the propagated fast and slow waves were significantly overlapped. For these rf signals, only a single *mixed* wave that corresponded to an interference field was apparent. There are presently no objective criteria for resolving the question of whether an observed rf signal is the result of the propagation of a single mode or two interfering modes (excluding trivial cases in which the fast and the slow waves are separated in time).

It is progressively easier to differentiate the fast and the slow wave when their overlap is minimized. Consequently, we simulated mixed rf signals for different propagation distances corresponding to strong, moderate, and weak overlaps between the fast and the slow wave. Specifically the specimen thickness, l, was set to 3, 6, and 12 mm, which correspond to approximately 85%, 70%, and 40% overlap for a simulated pulse of temporal length of about 3.8 μs.

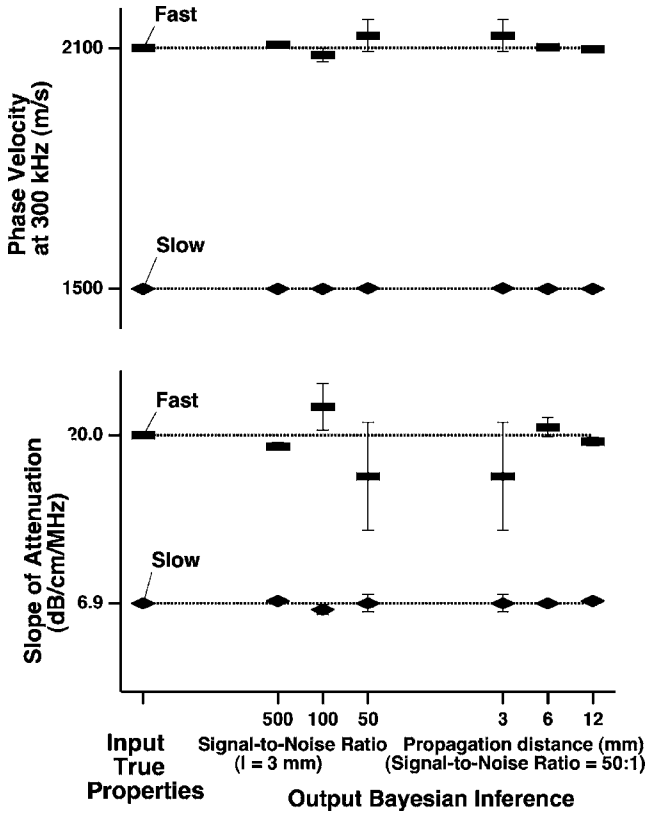


Fig. 2. Bayesian parameter estimates. The true values (that served as input to the data simulation) for the phase velocity at 300 kHz (top) and the slope of attenuation (bottom) for the fast and the slow wave are compared to the values estimated from the marginal posterior probabilities. The results for a range of noise levels (signal-to-noise ratio equal to 500:1, 100:1, and 50:1 for a fixed propagation distance of $l=3$ mm) and a range of propagation distances [l equal to 3, 6, and 12 mm for fixed signal-to-noise ratio (50:1)] are shown as mean$\pm$standard deviation. For all cases the mixed wave is composed of 30% fast wave and 70% slow wave.

The case for $l=3$ mm is depicted in Fig. 1. A visual examination of the simulated data [shown in both panels (a) and (b)] for two-mode propagation that served as an input to the Bayesian analysis reveals no signs of interference effects. However, evidence suggesting the presence of two modes rather than one is provided by comparing the input rf with the rf traces predicted from the parameter estimates that maximized joint posterior probability based on one-mode propagation [panel (c)] and two-mode propagation [panel (d)]. The residuals computed from the two-mode model are random and consistent with the characteristics of Gaussian noise that was introduced to the simulated data (signal-to-noise ratio of 50:1). In contrast, the residuals for a single mode fit exhibit a systematic variation, a strong indication that this is an improper model.

Our primary objective was to estimate the true material properties of the sample under investigation from the mixed mode study. We investigated data sets with a range of signal-to-noise ratios (500:1, 100:1, and 50:1), a range of propagation distances l (3, 6, and 12 mm), and different compositions of the overlapped wave [from 10% fast (90% slow) wave to 90% fast (10% slow) wave]. Representative comparisons of the parameter values inferred from Bayesian probability theory and the true values are depicted in Figs. 2 and 3. Unless otherwise specified, in these simulations the signal-to-noise ratio was set to 50:1. This value is smaller than that typically encountered in experimental studies, especially in laboratory-based through-transmission measurements.

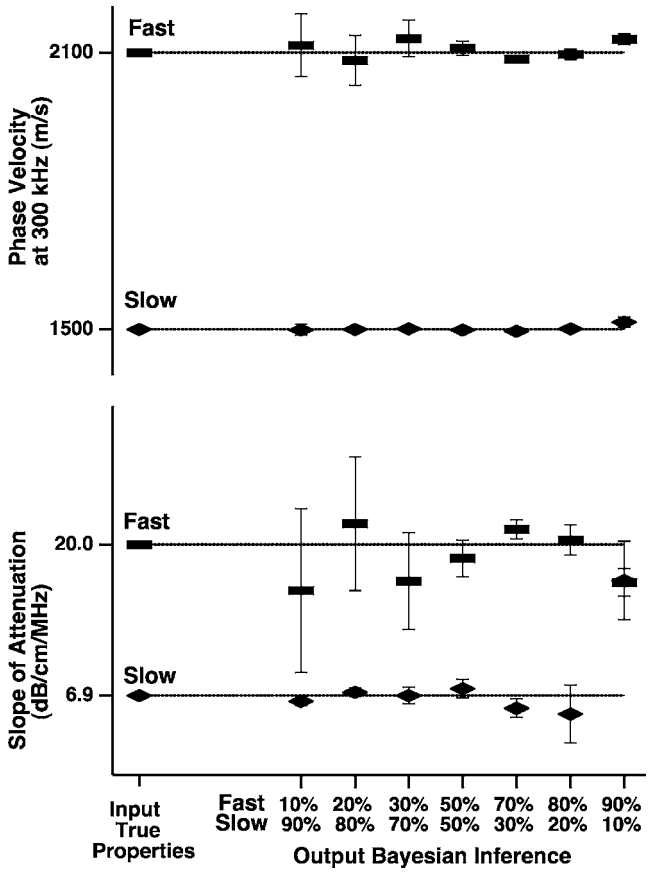


Fig. 3. Bayesian parameter estimates as a function of the mixed wave content. The Bayesian inferences for the phase velocity at 300 kHz (top) and the slope of attenuation (bottom) for the fast and the slow waves are obtained along with estimates of error from the marginal posterior probability. For all cases the propagation distance and the signal-to-noise ratio were fixed to 3 mm and 50:1, respectively. Parameter estimates are plotted as mean±standard deviation.

The parameter estimates in Figs. 2 and 3 correspond to the mean values of the marginal posterior probability with uncertainties represented by standard deviations. As expected, the parameter estimates improved as the noise was suppressed or the amount of overlap was reduced (by increasing bone thickness). The uncertainty bars in Fig. 2 are inversely proportional to the signal-to-noise ratio. This relationship between uncertainty in the parameters estimation and the noise level is a general feature of the Bayesian probability.

4.1 Limitations of the study

There are a number of limitations to the preliminary work described in this brief report. We identify some of the pertinent issues here. In some cases, we have carried out additional studies, the results of which are not presented here because of space limitations. When these additional studies are completed, a longer and more detailed manuscript will be submitted. In the current study we assumed that the attenuation coefficient increases approximately linearly with frequency in accordance with many results in the published literature (Droin *et al.*, 1998; Wear, 2000a). However, some studies reported attenuation that was proportional to a power law or had other functional forms (Chaffai *et al.*, 2000; Hoffmeister *et al.*, 2000; Langton *et al.*, 1984). We suggest that the interference effects, which are the subject of the present manuscript, may play a

role in measurements of the apparent frequency dependence of attenuation, making it difficult to ascertain whether the reported deviations from approximate linearity with frequency are true properties of bone or are a consequence of interference.

Another question that might be asked about the study reported here arises from the assumption of independence of the model parameters. It is known from the published literature, for example, that broadband ultrasound attenuation and speed of sound in bone are correlated. The consequences of such correlations can best be explored in studies of actual experimental data. In the approach used here, the parameters are logically independent of each other. Logical independence does not mean that the estimated parameters cannot be correlated. Correlations are present in the model and are incorporated into the inference through the likelihood.

Although we have successfully explored the use of a range of bandwidths and center frequencies in our simulations, the potential impact of specific experimental measurement effects such as diffraction, frequency-dependent transmission coefficients, and the like must be evaluated by processing actual experimental data. Regarding signal-to-noise ratio limitations, our results suggest that the accuracy with which the Bayesian approach can recover the true material properties is reduced as the signal-to-noise ratio decreases. It will not be clear what minimum signal-to-noise ratio is required, and, therefore, whether *in vivo* application of this approach will be feasible, until a range of experimental data has been analyzed. It is conceivable that meaningful values can still be recovered under challenging signal-to-noise ratio conditions by carrying out computationally more intensive processing, providing that the amplitudes of the received signals remain distinct from the noise level.

5. Summary

We have applied Bayesian probability theory to analyze the ultrasonic radio frequency signals produced by the interference of independently propagated fast and slow waves. Bayesian analysis was applied to obtain the estimates of the true parameter values that underlie ultrasonic propagation. For the cases and ranges of model parameters studied, Bayesian analysis resulted in accurate estimations, suggesting that Bayesian analysis might be effective in extracting true bone properties from mixed mode experimental data.

Acknowledgments

This work was supported in part by NIH Grants No. R37HL40302 and R24CA83060.

References and links

Bretthorst, G. L., Hutton, W. C., Garbow, J. R., and Ackerman, J. J. H. (**2005a**). "Exponential model selection (in NMR) using Bayesian probability theory," Concepts in Magnetic Resonance Part A **27A**, 64–72.

Bretthorst, G. L., Hutton, W. C., Garbow, J. R., and Ackerman, J. J. H. (**2005b**). "Exponential parameter estimation (in NMR) using Bayesian probability theory," Concepts in Magnetic Resonance Part A **27A**, 55–63.

Chaffai, S., Padilla, F., Berger, G., and Laugier, P. (**2000**). "*In vitro* measurement of the frequency-dependent attenuation in cancellous bone between 0.2 and 2 MHz," J. Acoust. Soc. Am. **108**, 1281–1289.

Droin, P., Berger, G., and Laugier, P. (**1998**). "Velocity dispersion of acoustic waves in cancellous bone," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **45**, 581–592.

Fellah, Z. E. A., Chapelon, J. Y., Berger, S., Lauriks, W., and Depollier, C. (**2004**). "Ultrasonic wave propagation in human cancellous bone: Application of Biot theory," J. Acoust. Soc. Am. **116**, 61–73.

Hoffmeister, B. K., Whitten, S. A., and Rho, J. Y. (**2000**). "Low-megahertz ultrasonic properties of bovine cancellous bone," Bone **26**, 635–642.

Hosokawa, A., and Otani, T. (**1997**). "Ultrasonic wave propagation in bovine cancellous bone," J. Acoust. Soc. Am. **101**, 558–562.

Jaynes, E. T. (**2003**). *Probability Theory: The Logic of Science*, edited by G. L. Bretthorst Cambridge, U.P., Cambridge, UK.

Langton, C. M., Palmer, S. B., and Porter, R. W. (**1984**). "The measurement of broadband ultrasonic attenuation in cancellous bone," Eng. Med. **13**, 89–91.

Marutyan, K. R., Holland, M. R., and Miller, J. G. (**2006**). "Anomalous negative dispersion in bone can result from the interference of fast and slow waves," J. Acoust. Soc. Am. **120**, EL5–EL61.

Nicholson, P. H. F., Lowet, G., Langton, C. M., Dequeker, J., and VanderPerre, G. (**1996**). "A comparison of time-domain and frequency-domain approaches to ultrasonic velocity measurement in trabecular bone," Phys. Med. Biol. **41**, 2421–2435.

Strelitzki, R., and Evans, J. A. (**1996**). "On the measurements of the velocity of ultrasound in the os calcis using short pulses," Eur. J. Ultrasound **4**, 205–213.

Waters, K. R., and Hoffmeister, B. K. (**2005**). "Kramers-Kronig analysis of attenuation and dispersion in trabecular bone," J. Acoust. Soc. Am. **118**, 3912–3920.
Waters, K. R., Hughes, M. S., Mobley, J., Brandenburger, G. H., and Miller, J. G. (**2000**). "On the applicability of Kramers-Kronig relations for ultrasonic attenuation obeying a frequency power law," J. Acoust. Soc. Am. **108**, 556–563.
Wear, K. A. (**2000a**). "The effect of frequency-dependent attenuation and dispersion on sound speed measurements: Application in human trabecular bone," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **47**, 265–273.
Wear, K. A. (**2000b**). "Measurements of phase velocity and group velocity in human calcaneus," Ultrasound Med. Biol. **26**, 641–646.
Wear, K. A., and Armstrong, D. W., III (**2001**). "Relationships among calcaneal backscatter, attenuation, sound speed, hip bone mineral density, and age in normal adult women," J. Acoust. Soc. Am. **110**, 573–578.

The fluid and elastic nature of nucleated cells: Implications from the cellular backscatter response

Ralph E. Baddour
Department of Medical Biophysics, University of Toronto, 610 University Avenue, Toronto, Ontario M5G 2M9, Canada
rbaddour@uhnres.utoronto.ca

Michael C. Kolios
Department of Physics, Ryerson University, 350 Victoria Street, Toronto, Ontario M5B 2K3, Canada
mkolios@ryerson.ca

Abstract: In a previous experiment [Baddour *et al.*, J. Acoust. Soc. Am. **117**(2), 934–943 (2005)] it was shown that it is possible to deduce the ultrasound backscatter transfer function from single, subresolution cells *in vitro*, across a broad, continuous range of frequencies. Additional measurements have been performed at high frequencies (10–65 MHz) on cells with different relative nucleus sizes. It was found that for cells with a nucleus to cell volume ratio of 0.50, the backscatter response was better modeled as an elastic sphere. For the cells in which the ratio was 0.33, the backscatter showed good agreement with the theoretical solution for a fluid sphere.

PACS numbers: 43.80.Cs, 43.80.Gx [CC]
Date Received: September 1, 2006 **Date Accepted:** October 20, 2006

1. Introduction

A large body of work has been published that shares the common objective of characterizing biological tissues by interrogation with ultrasound, whether it be by analysis of radio-frequency (rf) backscatter echo signals (Lizzi *et al.*, 1983) or of brightness-mode (B-mode) images (Waag, 1984). Although the advent of high-frequency, pulse-echo ultrasound devices (operating in the range of 10–65 MHz) has permitted higher resolution imaging of tissues, the associated wavelengths of sound (25–150 μm)—the resolution limit for imaging applications—are still greater than the size of most eukaryotic cells (5–30 μm, when in suspension). Nevertheless, techniques have been developed that exploit global alterations in acoustic properties, such as rf echo anisotropy (Insana *et al.*, 1991) or spectral slope (Lizzi *et al.*, 1996), to deduce average structural parameters in the size regime of cells, such as effective scatterer size or distribution. However, the actual nature of the scattering interaction at the cellular level remains not well understood. Not only is it difficult to measure acoustic scattering from single cells, due to problems with localization and low signal strength, but knowledge of the precise mechanical properties of cells, required to model the scattering process, is still an ongoing area of study.

The first successful backscatter measurement from individual cells was performed by Baddour *et al.* (2005) *in vitro* and used to evaluate a proposed model of acoustic scattering from single cells. Because of the large changes in backscatter intensity observed from cell ensembles during the process of apoptotic cell death (Kolios *et al.*, 2002, 2003), and since most of the significant gross structural changes that occur in apoptosis are related to the nucleus, it was hypothesized that the nucleus is primarily responsible for the scattering from a cell. Although the validity of the model of a cell as a simple elastic sphere with nucleuslike properties was unconvincing, this study provided a useful method to deduce the backscattering from individual subresolution scatterers. Here we show the backscatter response from two different cell types, measured using a refinement of this method, to higher frequencies, achieving a higher signal-

to-noise ratio (SNR) than in previous measurements. These new results provide evidence for the dual fluid and elastic character of cells.

2. Methods

Very sparse suspensions of cells (less than 10 000 cells/cm^3) were prepared in a degassed, dilute phosphate buffered saline (PBS) solution at room temperature, as described in Baddour *et al.* (2005). Suspensions of two different types of cells were prepared: OCI-AML-5 (Wang *et al.*, 1991), a line of human acute myeloid leukemia cells, and PC-3 (Kaighn *et al.*, 1979), a line of human prostate cancer cells. These two cell types were selected because of their significant difference in nucleus to cell volume ratios. Since time in culture was found to affect cell size (Taggart *et al.*, 2006), aliquots of both suspensions were taken on the day of the experiment for sizing by optical confocal microscopy. OCI-AML-5 cells were found to have mean nucleus and cell diameters of 9.1 and 11.5 μm, respectively (nucleus to cell volume ratio=0.50). For PC-3, nucleus and cell had respective mean diameters of 17.9 and 26.0 μm (volume ratio=0.33).

Data acquisition was performed using a VisualSonics VS40b (VisualSonics Inc., Canada) high-frequency ultrasound device using three single-element transducers, each with different resonant frequencies, *f* numbers, and focal lengths—20 MHz polyvinylidene fluoride: *f*2.35, 20 mm (VisualSonics Inc., Canada); 40 MHz polyvinylidene fluoride: *f*3, 9 mm (VisualSonics Inc., Canada); and 80 MHz lithium niobate: *f*3, 6 mm (Resource Center for Medical Ultrasonic Transducer Technology, University of Southern California). Since the device only generates short, broadband pulses with preset center frequencies, the transducers were excited with 19-, 40-, and 55-MHz pulses. The performance of the 80-MHz transducer pulsed at 55 MHz was satisfactory, with a 6-dB intensity bandwidth ranging from approximately 40 to 65 MHz. With each transducer, 30 independent sets of 255 linearly separated (30-μm spacing), raw rf echo signals (unprocessed A-scans) were acquired from both cell suspensions and then thresholded, discarding all scan lines not containing any data value greater than 80% of the maximum data value detected in all the echo signals (for that particular cell type and transducer). This threshold step, to remove empty acquisition lines and indirect cell hits, was set higher than in the previous study (Baddour *et al.*, 2005) in an attempt to obtain a higher "purity" of backscattered signals (i.e., where the scattered angle is exactly 180°). It is assumed that cells centered in the focal beam width of the transducer produce a higher amplitude echo for two reasons: any deviation from the transducer axis lowers the insonation intensity and, for spherical scatterers, scattering efficiency for the pure backscatter condition is always higher than for small angles off-axis.

Using the 10 to 30 rf signals that remain from each data set after thresholding, the backscatter transfer function (BSTF) was calculated as

$$\mathrm{BSTF} = \left| \frac{R(\omega,r)}{R_{\mathrm{ref}}(\omega,r)} \right|^2, \tag{1}$$

where $R(\omega,r)$ is the Fourier transform of the scattered signals from the transducer's depth of field. In this study, signals were windowed with a tight Hamming window (0.4 μs in width), centered relative to the maximum amplitude of the individual scattered pulses, prior to being Fourier transformed. $R_{\mathrm{ref}}(\omega,r)$ is the Fourier transform of a reference measurement: the perpendicular specular reflection from a flat, polished SiO_2 crystal (Edmund Industrial Optics Inc., part 43424; ρ=2.20 g/cm^3, c=5720 m/s) placed at the transducer focus in degassed PBS at room temperature. It has been shown that such a measurement is a valid approximation of the transducer's electromechanical response (Szabo *et al.*, 2004).

3. Results

Figure 1 shows representative backscattered pulses from individual OCI-AML-5 and PC-3 cells. For each transducer, the incident (transmitted) pulse presented is the reference pulse. Although the amplitude scales are in arbitrary units (output from the 8-bit digital sampler in the

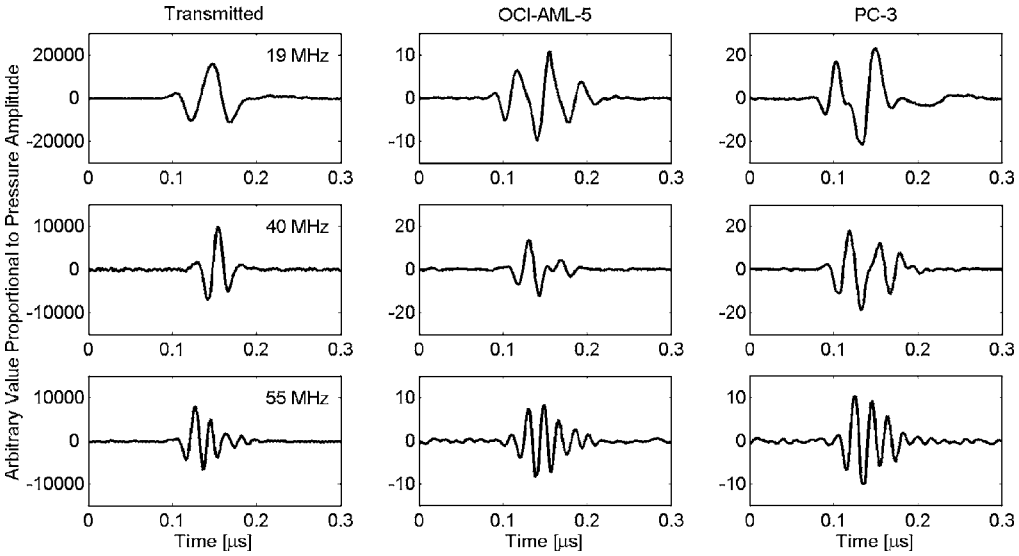


Fig. 1. Representative backscattered signals from individual cells of two different cell lines (OCI-AML-5, PC-3) and the corresponding incident acoustic pulses transmitted from three wideband transducers: an *f*2.35 20 MHz excited at 19 MHz, an *f*3 40 MHz excited at 40 MHz, and an *f*3 80 MHz excited at 55 MHz.

VS40b device), all signals were corrected for gain variations and are directly comparable. It is evident from the low noise in the signals that a higher SNR was achieved compared to Baddour *et al.* (2005), where different transducers were employed; SNRs computed here (after signal thresholding) were in the range of 11–15 dB, compared to 8–11 dB in the previous study.

The results of the method described to deduce the BSTF from individual OCI-AML-5 and PC-3 cells is presented in the form of spectral plots [Figs. 2(a) and 3(a)]. The BSTFs, expressed in decibels relative (dBr) to the backscatter intensity from the reference, are shown for the frequency ranges corresponding to the 6-dB bandwidths of each transducer.

Four theoretical backscatter frequency responses were calculated for both cell types: elastic sphere with cell or nucleus properties, and fluid sphere with cell or nucleus properties. The Faran-Hickling solution (Faran, 1951; Hickling, 1962) was employed for the calculation of the elastic sphere responses and the Anderson solution (Anderson, 1950) for the fluid sphere responses. To generate meaningful results, reasonable values for the bulk properties of cells and their nuclei were obtained and are summarized in Table 1. Speed of sound values were obtained from recent measurements of the acoustic properties of whole cells and isolated nuclei (Taggart *et al.*, 2006). For PC-3, which were not included in these measurements, the sound speeds measured from human embryonic kidney (HEK) cells and nuclei were chosen for use here; like PC-3, HEK cells are adherent in culture and are approximately similar in size. As derived in Baddour *et al.* (2005), 1.43 g/cm^3 was used as the effective density of the nuclei. The bulk density of a whole cell was calculated as a volume-weighted sum of the nucleus density and 1.05 g/cm^3, the assumed mean density of the cytoplasm (chosen to be marginally higher than typical sea water). Finally, the Poisson's ratio (a parameter only necessary in the elastic sphere calculations) of both OCI-AML-5 and PC-3 nuclei was assumed to be the same as the one of nuclei from chondrocytes (connective tissue cells), which has been measured to be 0.42 (Knight *et al.*, 2002). For whole cells, the value of 0.487 is from experiments by Boudou *et al.* (2006) with a polyacrylamide gel developed to simulate a thin layer of cells.

Figures 2(a) and 3(a) show the theoretical frequency response of the backscatter pressure intensity that best agrees with the corresponding measured BSTF as determined by least squares analysis. To avoid figure clutter, the remaining three other theoretical solutions are pre-

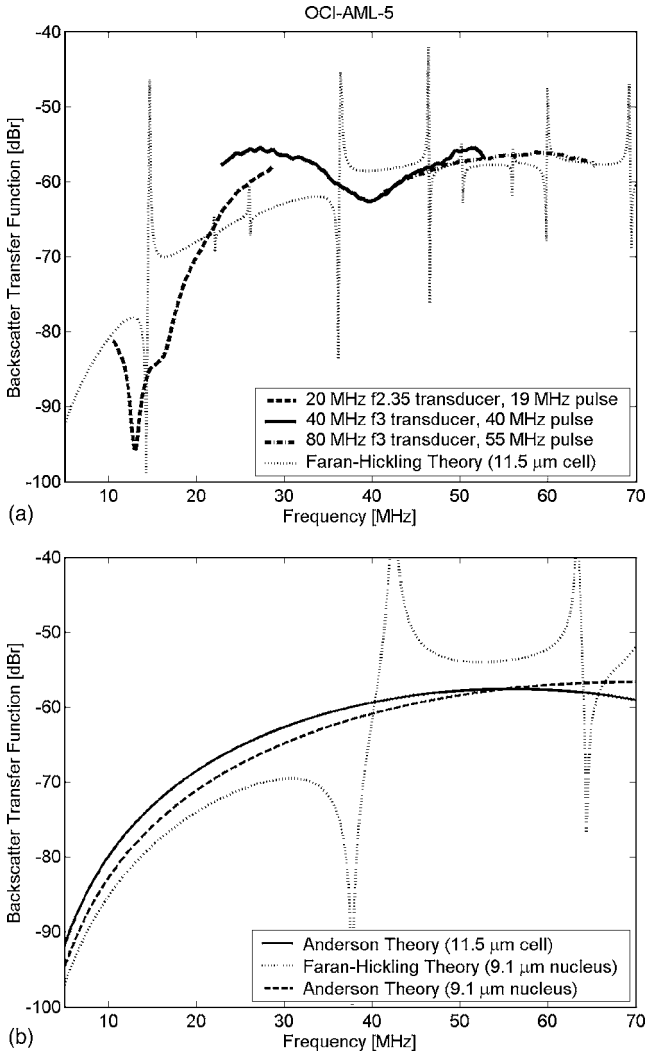


Fig. 2. Theoretical and measured backscatter frequency responses of single OCI-AML-5 cells in PBS using three different wideband transducers. (a) Measured BSTF and elastic sphere theory (properties of a whole cell). (b) Theoretical BSTFs for a fluid sphere with the properties of a whole cell, an elastic sphere with nucleus properties, and a fluid sphere with nucleus properties.

sented in Figs. 2(b) and 3(b). Note that the theoretical curves have not been shifted or scaled in any way, except to take into account the geometric effect of diminishing solid angle with distance.

4. Discussion and conclusions

The backscattered pulses from the smaller cells (OCI-AML-5) consistently had smaller peak amplitudes than those from the larger cell type (PC-3). This disparity became especially significant during 19-MHz broadband insonation. If one assumes that a PC-3 cell can best be modeled as a fluid sphere, this result is not surprising as the fundamental frequency for an air bubble in water, referred to as the Minnaert frequency if surface tension effects are disregarded (Minnaert, 1933), is $ka=0.01335$ (where a is the bubble radius and $k=\omega/c$, the incident wave number). This works out to 25 MHz for a bubble the same diameter as a PC-3 cell. Unlike the sharp

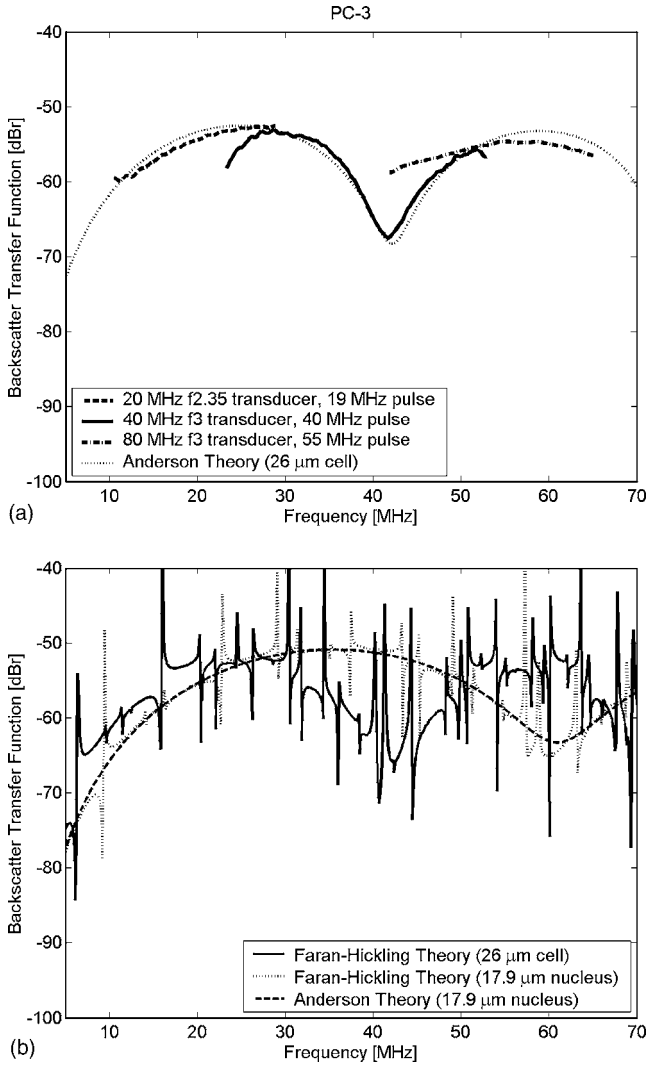


Fig. 3. Theoretical and measured backscatter frequency responses of single PC-3 cells in PBS using three different wideband transducers. (a) Measured BSTF and fluid sphere theory (properties of a whole cell). (b) Theoretical BSTFs for an elastic sphere with the properties of a whole cell, an elastic sphere with nucleus properties, and a fluid sphere with nucleus properties.

resonant peak predicted for a bubble, the theory for a liquid sphere in another immiscible liquid (a closer approximation to our experimental conditions) predicts a broader resonant response of smaller amplitude (Temkin, 1999). A rounded peak centered at 25 MHz is clearly evident in both the measured and theoretical results shown in Fig. 3(a). As can be seen in Fig. 3(b), the only other theoretical curve that predicts the measured peak in BSTF at 25 MHz is the elastic sphere whole cell theory. However, the numerous sharp resonant peaks predicted by the Faran-Hickling solution were not observed experimentally.

By visual inspection alone, it is obvious from Figs. 3(a) and 3(b) that the backscatter spectra for PC-3 cells are best fitted by the scattering theory of a fluid sphere with whole cell properties. For OCI-AML-5, the results are not as striking; the best fit appears to be with elastic sphere theory (whole cell properties). As seen in Fig. 2(a), there is some broad agreement in spectral features between the experimental BSTFs and the theoretical curve, such as the sharp

Table 1. Summary of the parameters used in the theoretical backscatter response calculations.

	Diameter d (μm)	Speed of sound c (m/s)	Density ρ (g/cm^3)	Poisson's ratio σ
OCI-AML-5 nucleus	9.1	1503	1.43	0.42
OCI-AML-5 cell	11.5	1535	1.24	0.487
PC-3 nucleus	17.9	1493	1.43	0.42
PC-3 cell	26.0	1523	1.18	0.487

dip between 10 and 20 MHz, the trough near 40 MHz, and the mean intensity level between 50 and 65 MHz. It is possible that the many sharp resonant spikes predicted by the Faran-Hickling solution might not be seen experimentally because of the normal biological variation of sizes and deviations from spherical shape for cells and nuclei in a given population, whereas the theory is defined for a geometrically perfect sphere. However, these can only be part of the cause; when the spectra from individual A-scans were examined, sharp resonances were never observed, implying the existence of additional factors.

Although analysis of the backscatter frequency response is somewhat indicative of the less fluid nature of OCI-AML-5 cells, there is further support for this hypothesis in the raw rf backscattered signals from these cells. When using a broadband 55-MHz insonation pulse (shown in Fig. 1), the scattered pulse from a typical OCI-AML-5 cell has an extent of approximately 0.1 μs and six cycles that are clearly identifiable above the noise floor. For a typical PC-3 cell, only five cycles can be identified over a similar period, although one might expect a longer backscattered signal for these cells due to their much larger size. The relatively longer backscattered pulse from the smaller OCI-AML-5 could be evidence of more elastic behavior, as elastic scatterers have a longer ring-down time.

By contrasting our analysis the ultrasonic backscatter responses measured from individual OCI-AML-5 and PC-3 cells, it is clear that the OCI-AML-5 cells behave more like elastic scatterers and that the PC-3 cells are very well modeled as fluid spheres. The results of this study imply that the nucleus behaves chiefly as an elastic scatterer and that the cytoplasm that surrounds it is well approximated by a fluid. We postulate that for cells with low nucleus to cell volume ratios (like PC-3), the effect of the nucleus is less important and the backscatter frequency response of the whole cell can best be modeled as a fluid sphere. Conversely, for cells in which the nucleus occupies a larger proportion of the cell volume (like OCI-AML-5), the elastic properties of the nucleus can no longer be ignored. It is possible that this class of cells might be best modeled as an elastic sphere surrounded by a fluid shell.

Acknowledgments

This work was supported by operating grants from the Canadian Institutes of Health Research and the Natural Sciences and Engineering Research Council of Canada. The ultrasound imaging device was purchased with funds from the Canadian Foundation for Innovation, the Ontario Innovation Trust, and Ryerson University.

References and links

Anderson, V. C. (**1950**). "Sound scattering from a fluid sphere," J. Acoust. Soc. Am. **22**(4), 426–431.
Baddour, R. E., Sherar, M. D., Hunt, J. W., Czarnota, G. J., and Kolios, M. C. (**2005**). "High-frequency ultrasound scattering from microspheres and single cells," J. Acoust. Soc. Am. **117**(2), 934–943.
Boudou, T., Ohayon, J., Arntz, Y., Finet, G., Picart, C., and Tracqui, P. (**2006**). "An extended modeling of the micropipette aspiration experiment for the characterization of the Young's modulus and Poisson's ratio of adherent thin biological samples: Numerical and experimental studies," J. Biomech. **39**(9), 1677–1685.
Faran, J. J. (**1951**). "Sound scattering by solid cylinders and spheres," J. Acoust. Soc. Am. **23**(4), 405–418.
Hickling, R. (**1962**). "Analysis of echoes from a solid elastic sphere in water," J. Acoust. Soc. Am. **34**(10), 1582–1592.
Insana, M. F., Hall, T. J., and Fishback, J. L. (**1991**). "Identifying acoustic scattering sources in normal renal parenchyma from the anisotropy in acoustic properties," Ultrasound Med. Biol. **17**(6), 613–626.
Kaighn, M. E., Narayan, K. S., Ohnuki, Y., Lechner, J. F., and Jones, L. W. (1979). "Establishment and

characterization of a human prostatic-carcinoma cell-line (PC-3)," Investig. Urol. **17**(1), 16–23.
Knight, M. M., Bravenboer, J. V. D. B., Lee, D. A., van Osch, G. J. V. M., Weinans, H., and Bader, D. L. (**2002**). "Cell and nucleus deformation in compressed chondrocyte-alginate constructs: Temporal changes and calculation of cell modulus," Biochim. Biophys. Acta **1570**(1), 1–8.
Kolios, M. C., Czarnota, G. J., Lee, M., Hunt, J. W., and Sherar, M. D. (**2002**). "Ultrasonic spectral parameter characterization of apoptosis," Ultrasound Med. Biol. **28**(5), 589–597.
Kolios, M. C., Taggart, L., Baddour, R. E., Foster, F. S., Hunt, J. W., Czarnota, G. J., and Sherar, M. D. (**2003**). "An investigagion of backscatter power spectra from cells, cell pellets and microspheres," Proc.-IEEE Ultrason. Symp. **2003**, 752–757.
Lizzi, F. L., Greenebaum, M., Feleppa, E. J., Elbaum, M., and Coleman, D. J. (**1983**). "Theoretical framework for spectrum analysis in ultrasonic tissue characterization," J. Acoust. Soc. Am. **73**(4), 1366–1373.
Lizzi, F. L., Astor, M., Kalisz, A., Liu, T., Coleman, D. J., Silverman, R., Ursea, R., and Rondeau, M. (**1996**). "Ultrasonic spectrum analysis for assays of different scatterer morphologies: Theory and very-high frequency clinical results," Proc.-IEEE Ultrason. Symp. **1996**, 1155–1159.
Minnaert, F. M. (**1933**). "On musical air-bubbles and the sounds of running water," Philos. Mag. **16**, 235–248.
Szabo, T. L., Karbeyaz, B. U., Cleveland, R. O., and Miller, E. L. (**2004**). "Determining the pulse-echo electromechanical characteristic of a transducer using flat plates and point targets," J. Acoust. Soc. Am. **116**(1), 90–96.
Taggart, L., Baddour, R. E., Giles, A., Czarnota, G., and Kolios, M. C. (**2007**). "Ultrasonic characterization of whole cells and isolated nuclei," Ultrasound Med. Biol. **33**, in press.
Temkin, S. (**1999**). "Radial pulsations of a fluid sphere in a sound wave," J. Fluid Mech. **380**, 1–38.
Waag, R. C. (**1984**). "A review of tissue characterization from ultrasonic scattering," IEEE Trans. Biomed. Eng. **31**(12), 884–893.
Wang, C., Koistinen, P., Yang, G. S., Williams, D. E., Lyman, S. D., Minden, M. D., and McCulloch, E. A. (**1991**). "Mast-cell growth-factor, a ligand for the receptor encoded by c-kit, affects the growth in culture of the blast cells of acute myeloblastic-leukemia," Leukemia **5**(6), 493–499.

Why do transposed stimuli enhance binaural processing?: Interaural envelope correlation vs envelope normalized fourth moment

Leslie R. Bernstein and Constantine Trahiotis
Department of Neuroscience and Department of Surgery (Otolaryngology), University of Connecticut Health Center, Farmington, Connecticut 06030
les@neuron.uchc.edu; tino@neuron.uchc.edu

Abstract: High-frequency, "transposed" stimuli have been shown to yield enhanced processing of ongoing interaural temporal disparities (ITDs). This paper concerns determining which aspect or aspects of the envelopes of such stimuli mediate enhanced resolution of ITD. Behavioral measures and quantitative analyses utilizing special classes of transposed stimuli show that the "internal" interaural envelope correlation accounts both qualitatively and quantitatively for the enhancement. In contrast, the normalized fourth moment of the envelope (Y), which provides an index of the degree to which the envelopes of high-frequency stimuli fluctuate, does not lead to a successful accounting of the data.

PACS numbers: 43.66.Pn, 43.66.Ba [QF]
Date Received: September 18, 2006 **Date Accepted:** October 13, 2006

1. Introduction

A number of recent studies has demonstrated that the processing of ongoing interaural temporal disparities (ITDs) at high frequencies can be enhanced via the use of "transposed" stimuli. van de Par and Kohlrausch (1997) created such stimuli in an effort to provide high-frequency auditory channels with envelope-based information made to mimic waveform-based information normally available only in low-frequency channels. The enhanced processing of ITDs via transposed stimuli has been demonstrated in terms of lower thresholds of binaural detection (van de Par and Kohlrausch, 1997), smaller threshold ITDs (Bernstein and Trahiotis, 2002), larger extents of laterality (Bernstein and Trahiotis, 2003), and resistance to binaural interference effects produced by the addition of simultaneously presented, diotic low-frequency energy (Bernstein and Trahiotis, 2004; 2005).

The similarities between the internal representations of high-frequency transposed stimuli and their low-frequency counterparts notwithstanding, it remains an open question just which aspect or aspects of the envelopes of high-frequency stimuli, be they transposed or conventional, foster enhancements of ITD processing. One possibility has been suggested by Dye *et al.* (1994) who measured threshold ITDs after manipulating the phase spectrum of equal-amplitude, multi-component wave forms centered at 4 kHz. They suggested that their listeners' performance and the binaural processing of ITDs conveyed by high-frequency stimuli, in general, could be accounted for by considering the degree to which the envelopes of the stimuli fluctuate. The metric Dye *et al.* used to quantify the degree of fluctuation of the envelope was one described in two earlier papers by Hartmann (1987) and Hartmann and Pumplin (1988). It is the fourth moment of the envelope normalized by the square of the power (the second moment) of the envelope:

$$Y = \langle E^4 \rangle / \langle E^2 \rangle^2.$$

Dye *et al.*'s (1994) appeal to the utility of *Y* was that it accounted qualitatively for the fact that their listeners achieved lower threshold ITDs with stimuli having greater values of *Y*. For example, their three-component-in-phase stimuli, which had a constant *Y* of 2.11, yielded substantially lower threshold ITDs than did the same three-component stimuli having their respective phases assigned randomly. For the latter stimuli, the ensemble average *Y* of the stimuli was 1.67.

It occurred to us that Dye *et al.*'s (1994) sine-phase and random-phase stimuli differed not only in terms of their *Y* values, but also in terms of the respective spectra of their envelopes. In order to understand why this is so, consider that rotating by 180° the phase angle of the one of the sidebands of a 100% sinusoidally amplitude-modulated (SAM) tone results in what is referred to as a quasi-frequency modulated (QFM) signal. Although the power spectrum of the *waveform* is unaltered by the phase shift, the power spectrum of the envelope is not. Whereas the spectrum of the envelope of the SAM tone contains a component at dc and a component at the frequency of modulation (*fm*), the spectrum of the envelope of the QFM signal contains a relatively larger dc component and a component at twice the frequency of modulation of the SAM tone (i.e., 2*fm*). Consequently, the two types of stimuli also differ in the interaural correlation functions of their envelopes.

Therefore, logically, it seemed possible that the differences in sensitivity to ITDs conveyed by the two types of stimuli in the experiment of Dye *et al.* stemmed from, or could be accounted for, by differences in their interaural envelope correlation functions rather than by differences in their values of *Y*, per se. In order to investigate this issue, we utilized a method of stimulus generation that allowed us to manipulate, independently, the *Y* values and the interaural correlation functions of the envelopes of the high-frequency signals conveying the ITD. It will be seen that changes in the "internal" interaural envelope correlation provide a quantitative account for threshold ITDs and changes in *Y* do not.

In order to generate the stimuli utilized in our experiment, we began with a four-component, equal-amplitude tonal complex consisting of 25, 50, 75, and 100 Hz. For one type of stimulus, the four components were in cosine phase. The four-component complex was then transposed to 4 kHz in the manner described in detail by Bernstein and Trahiotis (2002). For a second type of stimulus, what we refer to as the "max-flat" signal, the phases of the components were chosen, via a computer-based iterative technique, to yield the minimum *Y* value of the envelope of the stimulus after the low-frequency components were transposed to 4 kHz. The *Y* value of the cosine-phase signal was 7.9, the *Y* value for the max-flat signal was 2.7. Nevertheless, and crucial for our purposes, the normalized interaural envelope correlation functions (and the envelope power spectra) of these two types of stimuli were virtually identical.

A third type of stimulus was generated by beginning with the cosine-phase complex, transposing it to 4 kHz, and perturbing the phases of the components of the transposed *waveform*. The perturbation was performed by adding to the phase of each component a randomly chosen value between 0 and 1.1π. The value of 1.1π was determined, empirically, to yield an ensemble of waveforms having an expected value of *Y* of 2.7. A final ensemble of stimuli having *Y* values falling between 2.5 and 2.9 was used in the behavioral experiments. It is important to recognize that the *Y* value of 2.7 that characterized this set of "phase-scrambled" signals was the same value of *Y* that characterized the max-flat stimuli. In contrast, and again crucial for our purposes, the normalized interaural correlation functions of the envelopes of the phase-scrambled stimuli *differed* from the (virtually identical) correlation functions of the envelopes of the max-flat and cosine-phase transposed stimuli. The correlation functions of the envelope of the phase-scrambled signals were more "shallow" than those of the other two types of stimuli. That is, a larger change in the value of delay was required to bring about a given change in the value of the interaural correlation, as compared to the delay required with the other two stimuli. In summary, we constructed three sets of stimuli such that one pair (the cosine-phase and max-flat) had differing *Y* values and essentially identical interaural correlation functions, while a second pair (the max-flat and phase-scrambled) had identical *Y* values but differing interaural envelope correlation functions.

2. Experiment

2.1 Procedure

Detection of ongoing ITD was measured using the three types of transposed stimuli described above. Briefly, each of the four-component low-frequency tonal complexes was (linearly) half-wave rectified and transformed to the frequency domain where components above 2 kHz were filtered out by setting their magnitudes to zero. The resulting signal was transformed back to the time domain and multiplied by a 4 kHz sinusoidal carrier, yielding the desired high-frequency transposed stimulus (see Bernstein and Trahiotis, 2002; 2003 for further details concerning the process of transposition).

All stimuli were generated digitally with a sampling rate of 20 kHz (TDT AP2), were low-pass filtered at 8.5 kHz (TDT FLT2), and were presented via Etymotic ER-2 insert earphones at a level of 70 dB SPL. The duration of each stimulus was 300 ms including 20 ms $\cos^2$ rise-decay ramps. A continuous diotic noise low pass filtered at 1300 Hz (No equivalent to 30 dB SPL) was presented to preclude the listeners' use of any information at low spectral frequencies (e.g., Nuetzel and Hafter, 1976, 1981; Bernstein and Trahiotis, 1994).

Threshold ITDs were determined using a two-cue, two-alternative, forced choice, adaptive task. Each trial consisted of a warning interval (500 ms) and four 300 ms observation intervals separated by 400 ms. Each interval was marked visually by a computer monitor. Feedback was provided for approximately 400 ms after the listener responded. The stimuli in the first and fourth intervals were diotic. The listener's task was to detect the presence of an ITD (left-ear leading) that was presented with equal *a priori* probability in either the second or the third interval. The remaining interval, like the first and fourth intervals, contained diotic stimuli.

Because of the time required to generate the stimuli, 410-ms-long exemplars of each type of waveform were calculated prior to a trio of adaptive runs. Then, the exact waveform destined for each observation interval was determined by selecting, at random, the starting point of a 300-ms-long sample. This was done independently and repeatedly for all four observation intervals composing a trial within the adaptive run.

Ongoing ITDs were imposed by applying linear phase shifts to the representation of the signals in the frequency domain and then gating the signals destined for the left and right ears coincidentally, after transformation to the time domain. The ITD for a particular trial was determined adaptively in order to estimate 70.7% correct (Levitt, 1971). The initial step size for the adaptive track corresponded to a factor of 1.584 (equivalent to a 2 dB change of ITD) and was reduced to a factor of 1.122 (equivalent to a 0.5 dB change of ITD) after two reversals. A run was terminated after 12 reversals and threshold was defined as the geometric mean of the ITD across the last ten reversals.

Three normal-hearing adults (including the first author) served as listeners and three consecutive thresholds were obtained from each listener for each of the three types of stimuli. The ordering of the conditions was chosen randomly and was the same for all three listeners. After three thresholds were obtained for each type of stimulus, the conditions were revisited in reverse order. For two of the listeners, a total of 12 thresholds was obtained for each type of stimulus and the two smallest and two largest thresholds were excluded. For the third listener (the first author), a total of six thresholds was obtained for each type of stimulus and the smallest and largest of those thresholds were excluded. For two of the listeners, final threshold ITDs for each type of stimulus were based on the eight remaining estimates; for the third listener, final threshold ITDs for each type of stimulus were based on the four remaining estimates. All listeners received substantial practice before formal collection of data began.

2.2 Results and discussion

Figure 1 displays the mean of the threshold ITDs obtained across the three listeners for each type of stimulus, with an error bar representing its standard error. The asterisks will be discussed after the data are described. Exemplars of each of the three types of stimuli are depicted above their corresponding threshold ITDs. Note that the threshold ITDs obtained with the co-

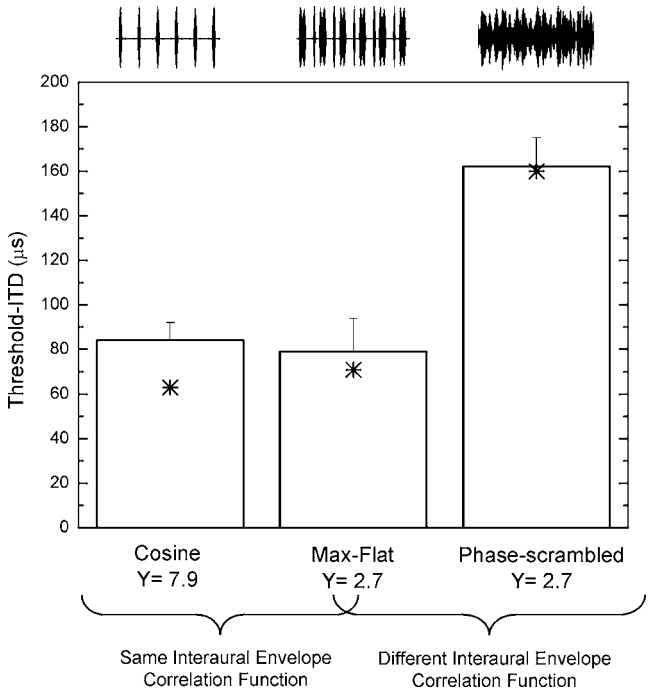


Fig. 1. Average threshold ITDs (in μs) across three listeners for the cosine, max-flat, and phase-scrambled stimuli, respectively. The error bars represent +1 standard error of the means. Asterisks represent predicted threshold ITDs calculated via a cross-correlation-based model. Exemplars of each of the three types of stimuli are depicted above their corresponding threshold ITDs.

sine and max-flat stimuli were both approximately 80 μs, while the threshold ITD obtained with the phase-scrambled stimulus was approximately twice that value. The reader is reminded that the cosine and max-flat stimuli had virtually identical interaural envelope correlation functions and differing values of Y while the max-flat and phase-scrambled stimuli had identical values of Y and differing interaural envelope correlation functions.

The patterning of the data suggests that the interaural envelope correlations, and not their Y values, account for sensitivity to ITD. This interpretation of the data was tested by calculating predictions of threshold ITDs from a cross-correlation based model that incorporates stages of peripheral processing representing bandpass filtering, envelope compression, and half-wave, square-law rectification. The model also includes a stage of low-pass filtering designed to attenuate spectral components of the envelope above 150 Hz (see Bernstein and Trahiotis, 2002 for details). Predictions of threshold ITD were made by minimizing the least-squared error between the data and the predictions assuming a constant-criterion change in interaural correlation at the output of the model. The predictions, indicated by the asterisks, accurately capture the findings in that they account for 88% of the variance in the data.

The success of the model notwithstanding, it seemed important to consider whether predictions stemming from changes in the interaural correlations of the envelopes of the physical stimuli (i.e., without taking into account effects resulting from peripheral processing) could also account for the data. Those predicted threshold-ITDs were 111 μs for the cosine-phase stimulus, 108 μs for the max-flat stimulus, and 127 μs for the phase-scrambled stimulus. Although those predictions conform to the ordering of the obtained threshold ITDs, they only account for 37% of the variance in the data. Clearly, as we have argued in the past (e.g., Bernstein and Trahiotis, 2002; 2003; Bernstein *et al.*, 1999), peripheral auditory processing must be taken into account.

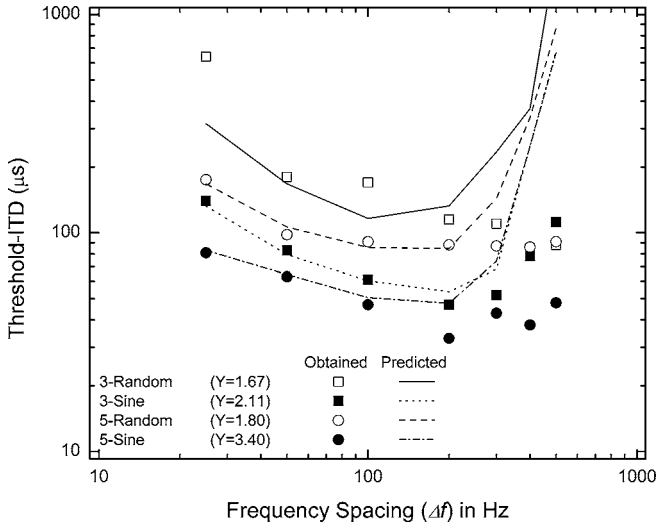


Fig. 2. Threshold ITDs (in μs) as a function of the frequency separation (Δf in Hz) between adjacent components of three- and five-component multi-tonal stimuli centered at 4 kHz. The data (symbols) were transcribed from Fig. 3 of Dye *et al.* (1994). Lines represent predictions calculated via a cross-correlation-based model.

In the same spirit, it seemed only fair to determine whether values of Y calculated *subsequent* to processing via the model might also account for the data. In order to do so, we calculated the normalized fourth moment of the output of the model (the envelope of the stimulus as it is represented "internally" by the model). The resulting values of Y were 5.5 for the cosine-phase stimulus, 2.1 for the max-flat stimulus, and 1.3 for the phase-scrambled stimulus. The ordering of these values of Y is not in accord with the notion that larger values of Y are associated with, or lead to, smaller threshold ITDs. Thus, the Y statistic, whether calculated "externally" or "internally," supports neither a qualitative nor a quantitative account of the data.

Figure 2 displays the mean threshold ITDs obtained by Dye *et al.* (1994) using their three- and five-component stimuli. The data were transcribed from their Fig. 3. The lines in the figure are predicted threshold ITDs calculated via the model described above. The predictions were made using a single criterion value of change in interaural correlation that was determined iteratively by evaluating visually how well predictions of the model seemed to fit the data. Changes in the criterion value of correlation affected only the vertical positioning of the four lines and not the ordering among them. Note that the predictions account for both how the threshold ITDs are ordered across the four types of stimuli and the fact that sensitivity to ITD improved as the separation between the components composing the stimuli increased from 25 to about 200 Hz. The major differences between the predictions and the data occurred for values of Δf above that value. We believe such differences do not represent a failure of the model. The threshold ITDs obtained by Dye *et al.* (1994) with those large values of Δf are abnormally small as compared to those commonly reported for high-frequency envelope-based ITD processing (e.g., Henning and Ashton, 1981; Nuetzel and Hafter, 1976, 1981, Bernstein and Trahiotis, 1994; Bernstein and Trahiotis, 2002). Because Dye *et al.* did not report using a low-pass masking noise, it may have been the case that their listeners used information stemming from physical or aural distortion products at low spectral frequencies.

In conclusion, it appears that the normalized interaural correlation of high-frequency complex stimuli can account quantitatively for differences in the resolution of ITDs found with stimuli having widely differing temporal signatures of their envelopes. In contrast, it appears that consideration of another potentially useful index, the normalized fourth moment of the envelope, does not lead to a successful explanation of the data.

Acknowledgments

This research was supported by research Grant Nos. NIH DC-04147, DC-04073, and DC-00234 from the National Institute on Deafness and Other Communication Disorders, National Institutes of Health.

References and links

Bernstein, L. R., and Trahiotis, C. (**1994**). "Detection of interaural delay in high-frequency SAM tones, two-tone complexes, and bands of noise," J. Acoust. Soc. Am. **95**, 3561–3567.

Bernstein, L. R., and Trahiotis, C. (**2002**). "Enhancing sensitivity to interaural delays at high frequencies by using transposed stimuli," J. Acoust. Soc. Am. **112**, 1026–1036.

Bernstein, L. R., and Trahiotis, C. (**2003**). "Enhancing interaural-delay-based extents of laterality at high frequencies by using 'transposed stimuli,'" J. Acoust. Soc. Am. **113**, 3335–3347.

Bernstein, L. R., and Trahiotis, C. (**2004**). "The apparent immunity of high-frequency transposed stimuli to low-frequency binaural interference," J. Acoust. Soc. Am. **116**, 3062–3069.

Bernstein, L. R., and Trahiotis, C. (**2005**). "Measures of extents of laterality for high-frequency 'transposed' stimuli under conditions of binaural interference," J. Acoust. Soc. Am. **118**, 1626–1635.

Bernstein, L. R., Par, Steven van de, and Trahiotis, C. (**1999**). "The normalized correlation: Accounting for $NoS\pi$ thresholds obtained with Gaussian and "low-noise" masking noise," J. Acoust. Soc. Am. **106**, 870–876.

Dye, R. H., Jr., Niemiec, A. J., and Stellmack, M. A. (**1994**). "Discrimination of interaural envelope delays: The effect of randomizing component starting phase," J. Acoust. Soc. Am. **95**, 463–470.

Hartmann, W. M. (**1987**). "Temporal fluctuations and the discrimination of spectrally dense signals by human listeners," in *Auditory Processing of Complex Sound*, edited by W. A. Yose and C. S. Watson (Erlbaum, Hillsdale, NJ).

Hartmann, W. M. and Pumplin, J. (**1988**). "Noise power fluctuations and the masking of sine signals," J. Acoust. Soc. Am. **83**, 2277–2289.

Henning, G. B., and Ashton, J. (**1981**). "The effect of carrier and modulation frequency on lateralization based on interaural phase and interaural group delay," Hear. Res. **4**, 186–194.

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

Nuetzel, J. M., and Hafter, E. R. (**1976**). "Lateralization of complex waveforms: Effects of fine-structure, amplitude, and duration," J. Acoust. Soc. Am. **60**, 1339–1346.

Nuetzel, J. M., and Hafter, E. R. (**1981**). "Discrimination of interaural delays in complex waveforms: Spectral effects," J. Acoust. Soc. Am. **69**, 1112–1118.

Par, S. van de and Kohlrausch, A. (**1997**). "A new approach to comparing binaural masking level differences at low and high frequencies," J. Acoust. Soc. Am.0001-4966 **101**, 1671–1680.

Auditory efferent activation in CBA mice exceeds that of C57s for varying levels of noise

Robert D. Frisina, S. R. Newman, and Xiaoxia Zhu
Department of Otolaryngology-Head and Neck Surgery, University of Rochester School of Medicine and Dentistry, 601 Elmwood Avenue, Rochester, New York 14642-8629
robert_frisina@urmc.rochester.edu; srnewm@yahoo.com; xiaoxiazhu2@yahoo.com

Abstract: The medial olivocochlear efferent (MOC) system enhances signals in noise and helps mediate auditory attention. Contralateral suppression (CS) of distortion product otoacoustic emissions (DPOAEs) has revealed age-related MOC declines. Here, differences in CS as a function of contralateral noise intensity (43–67 dB sound pressure level) were measured; 2f1-f2 DPOAE grams were recorded for young adult CBA and C57 mice. In CBAs, CS was a monotonic function of contralateral noise level. The C57s showed normal hearing, measured with DPOAE amplitudes and auditory brainstem response thresholds, but showed little CS, suggesting a loss of efferent dynamics preceding any deficiencies of the afferent auditory system.

PACS numbers: 43.64.Jb [BLM]
Date Received: October 4, 2006 **Date Accepted:** October 31, 2006

1. Introduction

In addition to the auditory afferent system, which carries auditory information to the central nervous system, there is an efferent feedback system whose role is only partially understood. There is likely a dynamic feedback regulation of the afferent system via the efferents utilizing the cochlear outer hair cell system, as well as a protective role, and involvement in auditory attention. Contralateral suppression (CS) of distortion product otoacoustic emission (DPOAE) amplitudes has been used as a means of measuring the medial olivocochlear efferent (MOC) system's activity. When noise is presented to a nontest ear, the test ear's DPOAEs are suppressed by efferent fiber activity that innervates OHCs. Past studies have demonstrated changes in the MOC system with aging. Specifically, it has been shown in both humans and mice there is a decrease in DPOAE amplitudes that is preceded by deficits in CS during aging (Kim *et al.*, 2002; Jacobson *et al.*, 2003).

Gorga *et al.* (2002) and James *et al.* (2002) demonstrated in humans that CS magnitude increased with increasing contralateral noise intensity levels. The CBA and C57 mouse strains are two of the most useful animal models for investigating the neural bases of age-related hearing loss—presbycusis (Willott, 1991; Frisina and Walton, 2001; 2006). The capabilities of the MOC in cases of changing sound conditions and as a function of different stimulus parameters are not well understood in the presbycusis mouse models. In a pioneering study, Sun and Kim (1999), utilizing an ipsilateral adaptation paradigm, found that C57 medial olivocochlear bundle responses were less than those of CBA mice. The goal of the present investigation was to identify differences in the strength of the MOC efferent response for CBA and C57 mice, with variations in the sound intensity level of a contralateral wideband noise.

2. Methods

Auditory brainstem response (ABR), DPOAE and CS testing procedures were similar to previous reports (Probst *et al.*, 1991; Jacobson *et al.*, 2003; Candreia *et al.*, 2004; Guimaraes *et al.*, 2004; Varghese *et al.*, 2005; Tadros *et al.*, 2006), so these methods are summarized here. All animal procedures were approved by the University of Rochester (Rochester, NY) Committee on Animal Resources.

2.1 Subjects

Young adult CBACa/J and C57BL/6 mice were used. The CBA strain has a gradual hearing loss that parallels that of presbycusis in humans, when one corrects for absolute differences in the lifespans of mice and men. C57 mice possess the *ahl* gene, which results in a rapid and early high-frequency hearing loss through young adulthood into middle age. CBA and C57 mice were bred in house, housed according to institutional protocols, with original breeding pairs obtained from Jackson Labs. Young adult CBA mice, 2.7–3.5 months of age, ($N=12$, 7 female, 5 male), and a group of C57 mice, all 2 months of age ($N=12$, 6 female, 6 male), were tested.

Prior to inclusion in the study group, each animal underwent testing to assess base line auditory function. Mice with clearly visualized, healthy tympanic membranes were included. Mice were anesthetized with a mixture of ketamine/xylazine (120 and 10 mg/kg body weight, respectively, intraperitoneal injection) prior to all experimental sessions, and normal body temperature was maintained with a heating pad. All recording sessions were completed in a soundproof acoustic chamber lined with Sonex. Prior to recording, the stimulus probe and microphone were placed near the tympanic membrane with the aid of the operating microscope.

2.2 DPOAE-grams

DPOAEs were recorded using the TDT BioSig III System. All stimuli were digitally synthesized at 200 kHz using SigGen software applications. A constant frequency ratio of 1.25 for frequencies 1 ($F1$) and 2 ($F2$) was used, and the tones were simultaneously presented as continuous tones, sampled every 84 ms, for 30 s. L1 was set to 65 dB sound pressure level (SPL) and L2 was set to 50 dB SPL, calibrated using a 0.1 ml coupler, which simulated the mouse ear canal. The level of the contralateral noise was randomly varied between 43 and 67 dB SPL for each DPOAE gram. The wideband noise had a bandwidth of 6–30 kHz, and the intensity levels were chosen to be below the middle ear muscle threshold in mice (e.g., Sun and Kim, 1999). Trials were alternated 2–3 times between increasing and decreasing orders of presented frequency levels (with random order of intensity levels for each subject), with the recording session duration limited by depth of anesthesia. Duration of testing was approximately 1 h per animal.

2.3 Data analyses and statistics

Results for each animal were averaged by frequency and applied towards group mean data. CS magnitude was defined and calculated as DP level with contralateral noise minus DP level in quiet, computed for each frequency and each intensity level (and reported as an absolute value). Statistical analysis utilized one-way analysis of variance (ANOVA, Prism, GraphPad, Inc.). When an ANOVA was statistically significant at the $p<0.05$ level or better, then Bonferroni's multiple comparison *post hoc t* tests were run to determine if statistically significant differences existed between specific conditions. For some analyses, data for different $F1/F2$ stimulus frequencies were averaged together, for three frequency ranges: $F1$ (low frequencies: 5–15 kHz), $F2$ (middle frequencies: 15–30 kHz), and $F3$ (high frequencies: 30–50 kHz).

After all testing was completed, the $F1$ and $F2$ values were compared throughout all intensity levels to determine if they had remained constant. ANOVAs were used to demonstrate that no statistically significant differences existed between the $F1$ and $F2$ magnitudes across the entire frequency range tested. This was done to rule out the involvement of the middle ear muscle response for the contralateral noise levels of the present investigation, which were selected to be below the middle ear reflex threshold in mice for wideband stimuli.

3. Results

Young adult CBA mice demonstrated normal ABR thresholds (Fig. 1(a)) and healthy DPOAE amplitudes in quiet (Fig. 1(b)), indicating that they had normal hearing (Willott, 1991; Jacobson *et al.*, 2003; Guimaraes *et al.*, 2004; Varghese *et al.*, 2005). In addition, a statistically significant increase in CS magnitude as a function of intensity was observed across the CBA mouse frequency range of hearing, as presented in Fig. 2. When results were analyzed based upon

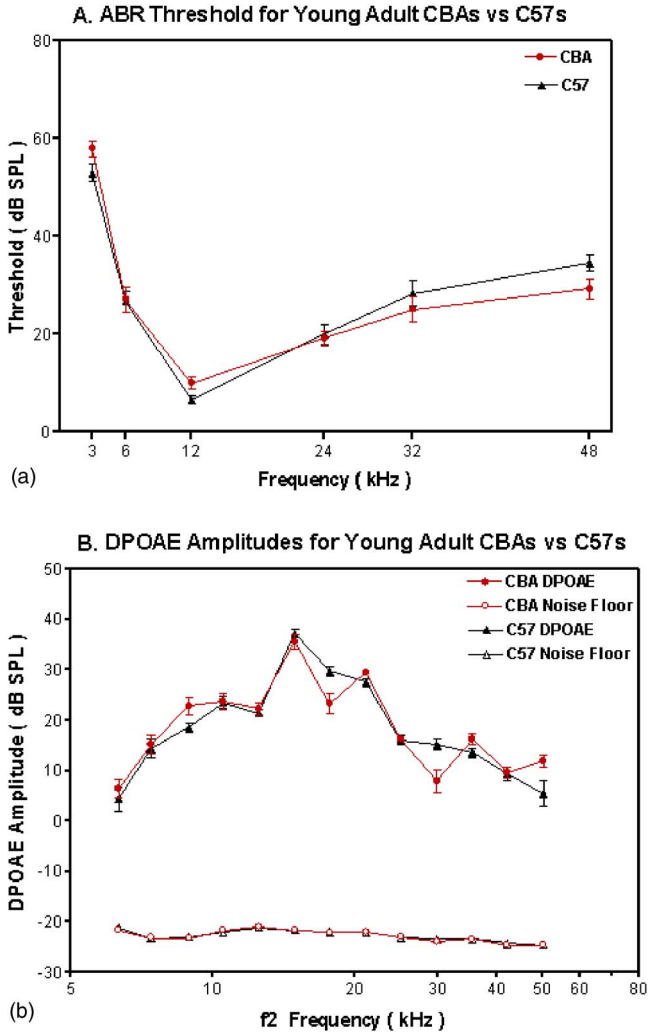


Fig. 1. (Color online) (A) Both subject groups demonstrated normal ABR thresholds with no significant differences between the two groups. Error bars represent standard deviation. (B) Both young adult C57 and CBA mice demonstrated good DPOAE amplitudes, well above the noise floor, with no significant differences between groups. Error bars represent standard error of the mean. This figure shows mean data for the two subject groups.

DPOAE frequency bands, a stronger efferent response was noted in the lower frequencies compared to intermediate and high frequencies of the mouse audibility range (Fig. 3).

Young adult C57 mice demonstrated normal ABRs (Fig. 1(a)) and DPOAE responses (Fig. 1(b)), indicating that, like the CBAs, they had normal hearing and healthy cochleae at this age, consistent with previous studies (Jimenez *et al.*, 1999). However, they displayed very low levels of CS, and little change in CS magnitude as intensity was increased, regardless of where in the mouse hearing range the measurements were made (Figs. 2 and 3). Note in Fig. 2 that there is a small CS response for C57s at some of the low frequencies, but significantly less than the CBAs.

4. Discussion

The results of this study support previous investigations in both human and animal subjects. The present experimental findings in CBA mice are consistent with the Gorga *et al.* (2002) and

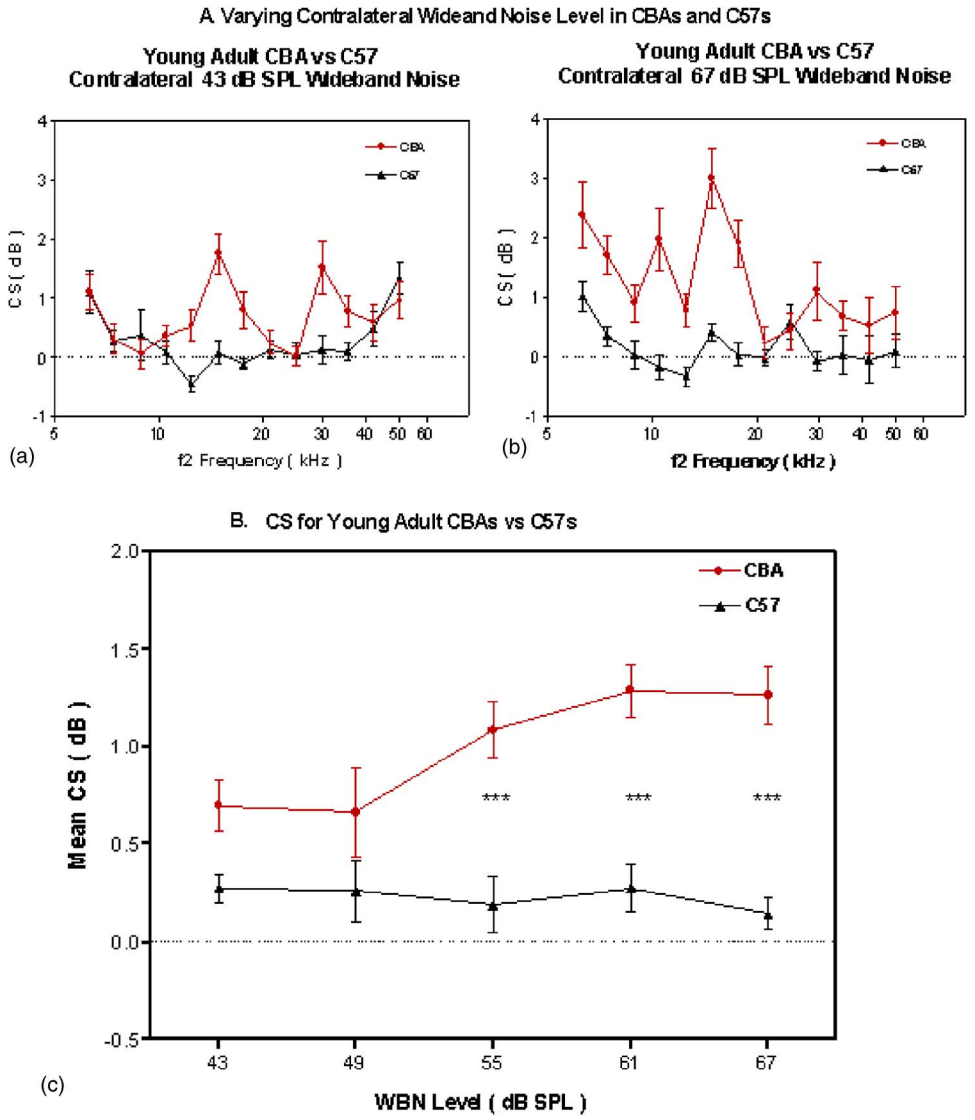


Fig. 2. (Color online) (A) CBA mice showed significantly more activation of the MOC system, especially at low frequencies, for both levels of the contralateral wideband noise (43 [left] and at 67 [right] dB SPL). (B) Comparison of CS in CBA vs C57 mice as the nontest ear was exposed to wideband noise ranging from 43 to 67 dB SPL. At 55, 61, and 67 dB SPL, the difference in CS between CBA and C57 subject groups was statistically significant ($p < 0.0001^{***}$). Error bars represent standard error of the mean. This figure shows mean data for the two subject groups.

James *et al.* (2002) demonstration of a heightened efferent response as contralateral noise intensity was raised for human listeners. Interestingly, the C57 mice, who go on to lose a significant portion of their auditory afferent function during the first year of life, show a very small efferent response at moderate sound levels, consistent with Sun and Kim (1999), and almost a complete lack of ability to increase their efferent activity in response to elevated contralateral noise levels. These results indicate that a poorly functioning MOC system precedes the age-related hearing loss in C57 mice, a hypothesis put forth previously by Kim *et al.* (2002) in humans and Jacobson *et al.* (2003) in rodents. An impaired efferent system could also contrib-

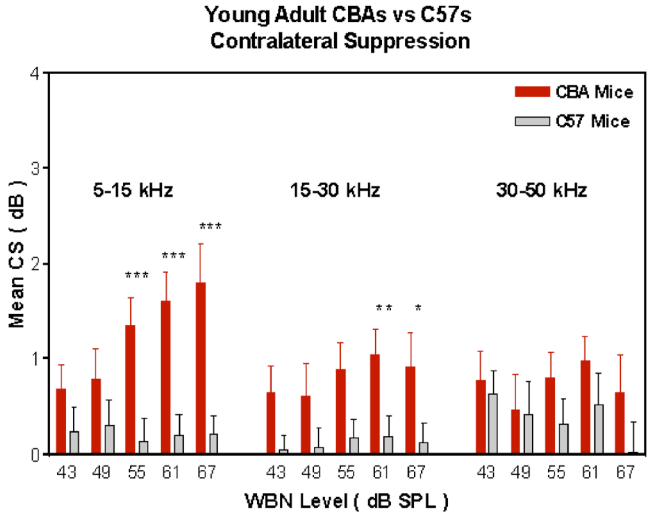


Fig. 3. (Color online) Comparison of CS in CBA vs C57 mice as the contralateral ear was exposed to wideband noise ranging from 43 to 67 dB SPL. Data are further broken down into stimulating frequency bands (of *F*1 and *F*2), with statistically significant differences indicated (*, $p<0.05$; **, $p<0.01$; ***, $p<0.001$). Error bars represent standard error of the mean. *F*1: 5–15 kHz, *F*2: 15–30 kHz, *F*3: 30–50 kHz. This figure shows mean data for the two subject groups.

ute to the increased susceptibility to noise-induced hearing loss inherent in C57s relative to CBAs and certain other mouse strains (Jimenez *et al.*, 2001; Vazquez *et al.*, 2004).

The presence of a more vigorous efferent response in the lower frequencies of DPOAEs is interesting. The question of why the auditory system provides more attenuation or protection for lower frequency hearing may be related to the fact that most environmental background noise has more energy in the low frequencies. In addition, attenuating the transmission of low- frequency information to the central auditory system is a way to diminish the effects of upward spread of masking in conditions of high levels of environmental background noise. Last, it may be that preferential attenuation of low frequencies by the auditory efferent feedback system contributes to the earlier loss of high-frequency hearing, and the preservation of low-frequency hearing to a relatively later age in the C57 strain.

5. Summary and conclusions

In young CBA mice, the magnitude of the contralateral suppression—*or strength of the medial olivocochlear efferent system*—increases with contralateral intensity of a wideband noise. In young C57 mice, the magnitude of the contralateral suppression is small, and does not increase significantly with contralateral stimulus intensity. These findings support the idea that a poorly functioning MOC efferent system precedes the rapid, age-related loss of hearing in C57 mice. The MOC efferent system has its most dramatic action in the lower frequencies of the mouse hearing range. As these results are compared to results of mice in different age groups, the role of the MOC efferent system in the etiology of age-related hearing loss—presbycusis—will likely be better understood.

Acknowledgments

We thank John Housel and Enza Daugherty for technical and project support. Supported by NIH Grants NIA P01 AG09524 from the National Institute on Aging, NIDCD P30 DC05409 from the National Institute on Deafness & Communication Disorders, and the International Center for Hearing Speech Research, Rochester, NY.

References and links

Candreia, C., Martin, G. K., Stagner, B. B., and Lonsbury-Martin, B. L. (**2004**). "Distortion product otoacoustic emissions show exceptional resistance to noise exposure in MOLF/Ei mice," Hear. Res. **194**, 109–117.
Frisina, R. D., and Walton, J. P. (**2001**). "Aging of the mouse central auditory system," in *Handbook of Mouse Auditory Res: Behavior to Molecular Biology*, edited by J. P. Willott (CRC, New York), pp. 339–379.
Frisina, R. D., and Walton, J. P. (**2006**). "Age-related structural and functional changes in the cochlear nucleus," Hear. Res. **217**, 216–233.
Gorga, M. P., Neely, S. T., Dorn, P. A., Dierking, D., and Cyr, E. (**2002**). "Evidence of upward spread of suppression in DPOAE measurements," J. Acoust. Soc. Am. **112**, 2910–2920.
Guimaraes, P., Zhu, X., Cannon, T., Kim, S. H., and Frisina, R. D. (**2004**). "Sex differences in distortion product otoacoustic emissions as a function of age in CBA mice," Hear. Res. **192**, 83–89.
Jacobson, M., Kim, S. H., Romney, J., Zhu, X., and Frisina, R. D. (**2003**). "Contralateral suppression of distortion-product otoacoustic emissions declines with age: A comparison of findings in CBA mice with human listeners," Laryngoscope **113**, 1707–1713.
James, A. L., Mount, R. J., and Harrison, R. V. (**2002**). "Contralateral Suppression of DPOAE measured in real time," Clin. Otolaryngol. Allied Sci. **27**, 106–112.
Jimenez, A. M., Stagner, B. B., Martin, G. K., and Lonsbury-Martin, B. L. (**1999**). "Age-related loss of distortion product otoacoustic emissions in four mouse strains," Hear. Res. **138**, 91–105.
Jimenez, A. M., Stagner, B. B., Martin, G. K., and Lonsbury-Martin, B. L. (**2001**). "Susceptibility of DPOAEs to sound overexposure in inbred mice with AHL," J. Assoc. Res. Otolaryngol. **2**, 233–245.
Kim, S. H., Frisina, D. R., and Frisina, R. D. (**2002**). "Effects of age on contralateral suppression of distortion product otoacoustic emissions in human listeners with normal hearing," Audiol. Neuro-Otol. **7**, 348–357.
Probst, R., Lonsbury-Martin, B. L., and Martin, G. K. (**1991**). "A review of otoacoustic emissions," J. Acoust. Soc. Am. **89**, 2027–2067.
Sun, X. M., and Kim, D. O. (**1999**). "Adaptation of 2f1-f2 distortion product otoacoustic emissions in young-adult and old, CBA and C57 mice," J. Acoust. Soc. Am. **105**, 3399–3409.
Tadros, S. F., D'Souza, M., Zettel, M. L., Zhu, X., Lynch-Erhardt, M., and Frisina, R. D. (**2007**). "Serotonin 2B receptor: Upregulated with age and hearing loss in mouse auditory system," Neurobiol. Aging (in press).
Varghese, G. I., Zhu, X., and Frisina, R. D. (**2005**). "Age-related declines in contralateral suppression of distortion product otoacoustic emissions utilizing pure tones in CBA/CaJ mice," Hear. Res. **209**, 60–67.
Vazquez, A. E., Jimenez, A. M, Martin, G. K., Luebke, A. E., and Lonsbury-Martin, B. L. (**2004**). "Evaluating cochlear function and the effects of noise exposure in the B6. CAST+Ahl mouse with distortion product otoacoustic emissions," Hear. Res. **194**, 87–96.
Willott, J. F. (**1991**). *Aging and the Auditory System: Anatomy, Physiology, and Psychophysics* (Singular San Diego), p. 286.

The acoustic signature of marine seep bubbles

Ira Leifer
Marine Science Institute, University of California, Santa Barbara, California 93106-1080
ira.leifer@bubbleology.com

Dajun Tang
Applied Physics Laboratory, University of Washington, Seattle, Washington 98105
djtang@apl.washington.edu

Abstract: Observations of seabed bubbles (62 m depth) at a natural marine hydrocarbon seep by passive acoustic and optical approaches are compared. The acoustic and optical methods observed a bimodal distribution with peaks at 1500 and 1750 Hz, and 2200 and 2800 μm radius, respectively. Radii were ~20% lower than predicted by the Minnaert formula. Frequency shifts were observed for bubbles emitted within a few milliseconds and were attributed to coupling between nearby bubbles. Surfactants also may have played a role.

PACS numbers: 43.30.−k, 43.20.Px, 43.30.Xm [GD]
Date Received: July 20, 2006 **Date Accepted:** October 3, 2006

1. Introduction

Bubble size measurements in nature (Medwin and Breitz, 1989), and industry (Boyd and Varley, 2001) is a problem of critical importance to a wide range of disciplines that has long challenged researchers. For example, to predict air-sea gas exchange or to optimize mass transfer in chemical reactors, as well as for many other geophysical and industrial applications, requires knowing the bubble size distribution. Approaches have been developed including optics (Leifer *et al.*, 2003), laser (Asher *et al.*, 1997), and active acoustics (Medwin and Breitz, 1989). Each approach has advantages and disadvantages (Leifer *et al.*, 2003). Optical measurements have intensive processing requirements (high bandwidth) and typically are invasive, particularly in turbid water. Active acoustic (sonar) techniques can operate at a distance but are difficult to interpret at high bubble densities or for large (i.e., nonspherical) bubbles, and have high power requirements. Passive acoustics has been investigated for bubble measurement in chemical engineering (Boyd and Varley, 2001) and geophysical systems (Leighton and Walton, 1987). However, uniquely determining the bubble size distribution for some sources, such as breaking waves (Deane, 1997) remains a significant challenge. Passive acoustics has advantages of low cost, power, and bandwidth, remote sensing potential, and highly compressible data. In this study, passive acoustic and optical seabed bubble observations at a natural marine hydrocarbon seep are compared for a short data set, selected as typical of many vents in the seep field.

Marine seeps are of interest because their emission of methane—an important greenhouse gas—may be important to atmospheric budgets (Kvenvolden and Rogers, 2005). However, bubble dissolution presents a potentially significant barrier to transit of the water column (Clark *et al.*, 2003). Critical to understanding the fate of seep methane is the bubble-emission size distribution; however, few distributions have been published (Leifer and Boles, 2005; Leifer and MacDonald, 2003). This study investigates passive acoustics as a useful tool for seep studies.

Passive acoustic bubble measurement uses the bubble formation sound (volume oscillations). The peak frequency (the zeroth oscillatory mode) f and the bubble equivalent spherical radius r for a given pressure P are (Minnaert, 1933):

$$f = 1/(2\pi r)\sqrt{(3\gamma P/\rho)}, \tag{1}$$

where γ is the ratio of gas specific heat at constant pressure to constant volume and is either 1-isothermal, or 5/3—adiabatic—and ρ is the water density. For a $r \sim 1$ mm bubble at atmo-

spheric pressure, $f \sim 3.5$ khz. After formation, the acoustic signal rapidly decreases with the envelope shape partly determined by the driving pressure creating the bubble, which deforms the bubble from spherical shape (Strasberg, 1955). Some laboratory data show a frequency shift within the pulse and are proposed to result from nonlinear, hydrodynamic factors affecting the volume oscillations (Manasseh *et al.*, 2001). Coupling between adjacent bubbles increases the peak frequency with decreasing separation for stationary bubbles (Hsiao *et al.*, 2001) and shifts with increasing bubble production rate (Nikolovska *et al.*, 2003) and walls (Payne *et al.* 2004). Furthermore, Eq. (1) was derived using a linear approach that may be inappropriate for larger bubbles (Manasseh *et al.*, 2001), whose initial shape oscillations are asymmetric (Longuet-Higgins, 1989).

2. Measurement techniques

A hydrophone (Bioacoustic Probe, GreeneRidge Sciences, Inc. Santa Barbara, CA) recorded seep bubble formation at the seabed and is described in detail in Tang (2005). The hydrophone was mounted to a bubble measurement system (described briefly below) above the camera's field of view and was shielded from rising bubbles. The hydrophone is self-contained and was $\sim$70 cm distant from seep vents during observations. For this study, the sampling frequency was 12 kHz, which allowed $\sim$5 h continuous recording on the internal battery. The hydrophone also records depth and temperature every second. Because the hydrophone and bubble measurement system were mounted directly to the submarine, the substantial vehicle noise reduced the signal to noise ratio (SNR) of bubble formation.

A video bubble measurement system (BMS) was submarine mounted and deployed in the Coal Oil Point seep field to quantify seep bubble size distributions during a submarine cruise on 21 September 2005. Optical bubble measurement and analysis are reviewed in Leifer *et al.* (2003) with details of the BMS for seep measurements in Leifer and Boles (2005). Recent BMS improvements included two transparent screens to delineate the measurement volume and prevent along-axis bubble advection. Thus bubbles are within a clearly defined range of distances or size scales. Bubble blockers with downward lips prevented bubbles from rising between the camera and measurement volume or behind the screen and casting shadows.

Field video was recorded on a mini-DV recorder and then digitized at 60 fields s^{-1} and full digital resolution and extracted to 60 images s^{-1} (720×480 pixels). Bubbles were tracked between frames allowing 10–15 measurements of bubble position, radius r, and time t for each bubble. Vertical velocities $V_Z(r)$ and trajectory angles were calculated for each bubble. Trajectory angle varied due to submarine and BMS tilt, surge, and currents. Histograms with logarithmically spaced r bins were calculated and combined with $V_Z(r)$ to calculate the emission size distribution, $F(\sharp \mu m^{-1}\ s^{-1})$.

3. Setting

Bubble vents were observed both optically and acoustically at the informally named La Goleta seeps (major plume, La Goleta A at 34° 22.503′ N, 119° 51.193′ W, at the seabed, 62 m deep) in the Coal Oil Point (COP) seep field during a series of Delta submarine dives. The COP seep field is arguably the best-studied seep field in the world, and releases significant gas to the ocean and atmosphere, $\sim$1.15 m^3 gas s^{-1} (Hornafius *et al.*, 1999) escapes to the atmosphere from $\sim$3 km^2 of seafloor. The seeps release $\sim$80 barrels oil/day^{-1} -5×10^6 l yr^{-1} (Clester *et al.*, 1996) with oil slicks a perennial channel feature (Leifer *et al.*, 2006).

The study area seabed is comprised of fine-grained sediment with vents separated on meter-scale distances emitting bubble chains or small bubble streams. Some vents were associated with small (5 cm length scale) pockmarks in the sediment that contained tar. Seabed gas collected from the main plume of La Goleta Seeps was 76% methane, 4% ethane, 1.6% butane, with trace higher *n*-alkanes, 9.2% carbon dioxide and $\sim$8% air (unpublished data). Bubbles were slightly oily, although extensive oil slicks generally are observed at the sea surface of La Goleta Seeps.

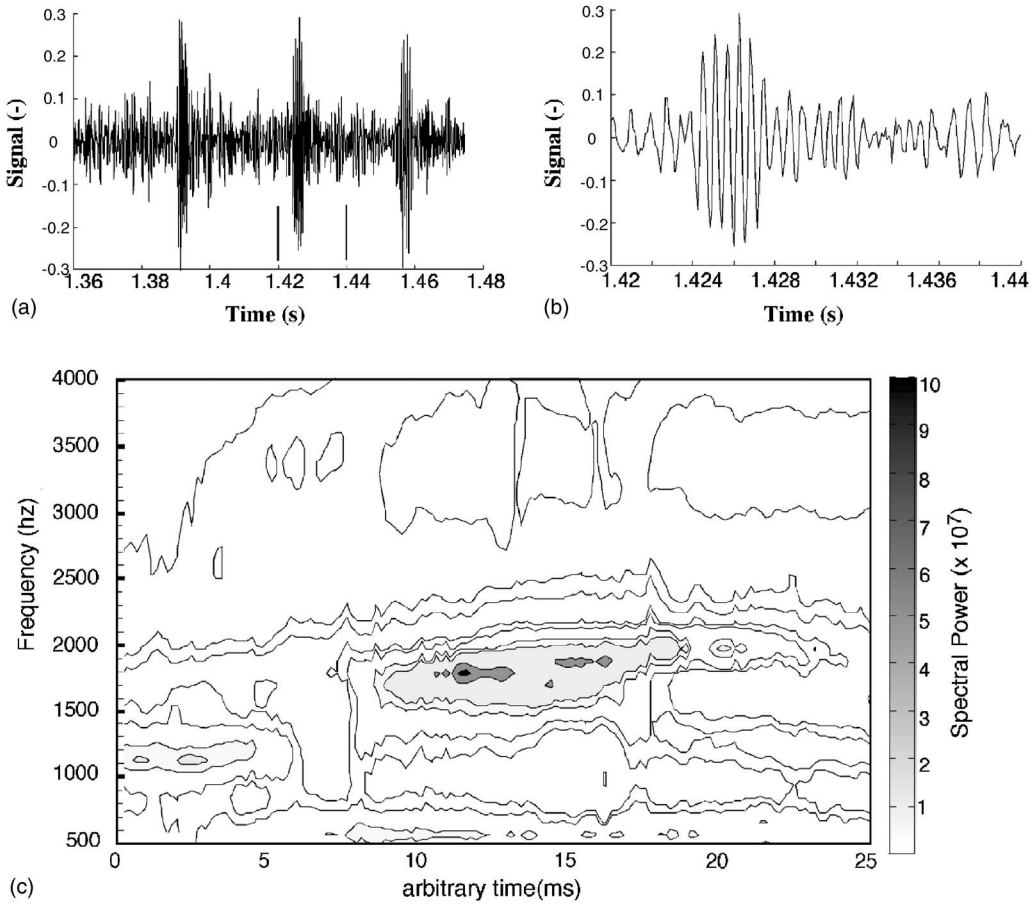


Fig. 1. (a) Short segment of recorded acoustic signal. (b) Expanded subset of (a) between the two vertical marks on (a). (c) Spectragram (order 1238, 14-pole, Burg spectral method on 128 point data sequences after a Hanning window) showing two bubbles produced almost simultaneously.

4. Results

A portion of the acoustic time series was analyzed. Four seconds of data was chosen corresponding to when submarine noise was minimal. Bursts of sound from the formation of individual bubbles clearly are visible (Fig. 1). The audio signal was high-pass filtered (fifth-order Butterworth, 1.2 kHz cutoff frequency) to remove submarine noise below 1 kHz and allow the bubble sounds to be clearly audible. A short time sequence shows the typical oscillatory pattern (Fig. 1). However, due to the poor signal to noise ratio ($\sim$10–20 db) and the bubble sound transiency ($\sim$6 ms or less), Fourier transforms provided inadequate spectral resolution. Instead, spectral analysis was conducted using the Burg algorithm, a parametric spectral estimation method (Marple, 1988). Spectragrams were calculated on 128 point (10.6 ms) data sequences after applying a Hanning window. Spectragrams of the entire data set were manually inspected to identify bubbles (256 order, 24-pole, Burg method), e.g., Fig. 1(c). Once a bubble was identified, 8192-order ($\sim$1 Hz resolution), 24-pole spectra were calculated on 256-data point sequences and the peak frequency f, identified for the strongest signal. Note, that different data lengths cause variations in f due to shifts in f during the pulse, thus uncertainty in f is $\sim$20 Hz. Given the sharpness of the spectral peak of bubble sound (Fig. 1(c)), unevenness in the hydrophone response likely affects f less than the 20 Hz uncertainty. A total of 64 bubbles were identified.

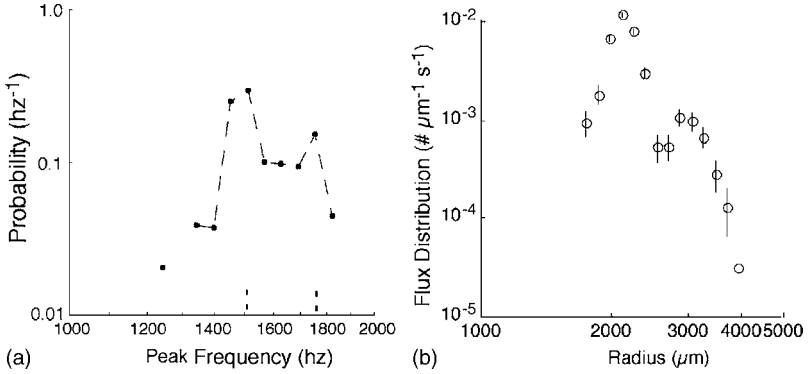


Fig. 2. (a) Probability distribution of peak acoustic frequencies from bubbles recorded by the bioprobe hydrophone. (b) Optical bubble emission size distribution for the same vent.

The probability distribution of f was calculated by generating histograms with logarithmically spaced f bins, normalized to Hz^{-1}, and shows two emission modes centered at 1500 and 1750 Hz (Fig. 2(a)). The optical F also was bimodal with peaks at $r=2200$ and 2800 μm (Fig. 2(b)). The ratio between the two frequency modes was 1.2, similar to the ratio between the peaks in F, 1.27. Direct comparison of individual bubbles between the acoustic and optical data was not feasible as the camera was mounted ~70 cm above the seabed and thus observed the bubbles after they had risen from the seabed, on the order of 3 s based on measured V_Z. V_Z showed significant variability and were slower than V_Z for clean bubbles indicating surfactant contamination.

In some cases, bubbles escaped from the vent in pairs or trios—based on overview video of the vent taken after the BMS sequence was recorded. Where bubbles escaped very closely in time, frequency shifts were observed from the initial frequency towards the frequency of the later bubble(s). Shifts on the order of 100 Hz were observed over a few milliseconds (Fig. 3(a)). Bubbles produced more than a few milliseconds apart did not evidence a frequency shift (Fig. 3(b)). Overall, bubbles were emitted at about 25 per second although erratically.

5. Discussion and conclusion

This study presents the first field data of passive acoustic and optical bubble measurements. Bubble radii calculated from f were ~20% greater than the optical derived radii. There are several possible explanations, including surfactants—or acoustic coupling between adjacent bubbles. Sinha (2003) noted that surfactants increased f. Bubble surface and path oscillations were significantly damped compared to similar size clean bubbles and V_Z were much slower than for clean bubbles, indicating that even the larger bubbles ($r \sim 3000$ μm) were significantly contaminated. Note, bubbles this large generally behave as though hydrodynamically clean in natural waters (Patro *et al.*, 2002). The effect of surfactants on bubble dynamics during and shortly after formation is poorly understood.

This has been proposed to result from bubble-bubble interactions where bubbles are emitted in a bubble chain (Manasseh *et al.*, 1998). Another significant difference was the shift in f during a bubble pulse where the bubbles were rapidly in sequence. The most likely explanation is coupling between adjacent bubbles. Several studies suggest that frequency shifts occur for bubbles closer than $20r$ (Payne *et al.*, 2004), ~4–6 cm for these seep bubbles. The change in frequency during the pulse could result from the preceding bubble's acceleration increasing the distance to the second bubble, coupling of bubbles of different sizes, or a surfactant effect. Further, the progression of the acoustic signal along a bubble chain causes a phase shift that changes the signal envelope from a sharp onset and exponential decrease—e.g., Longuet-Higgins (1989)—to a hill-shaped envelope (Nikolovska *et al.*, 2003) as in Fig. 1(b). However, in some cases where the bubbles were produced 5 or more ms apart—i.e., 1–2 mm—they did not

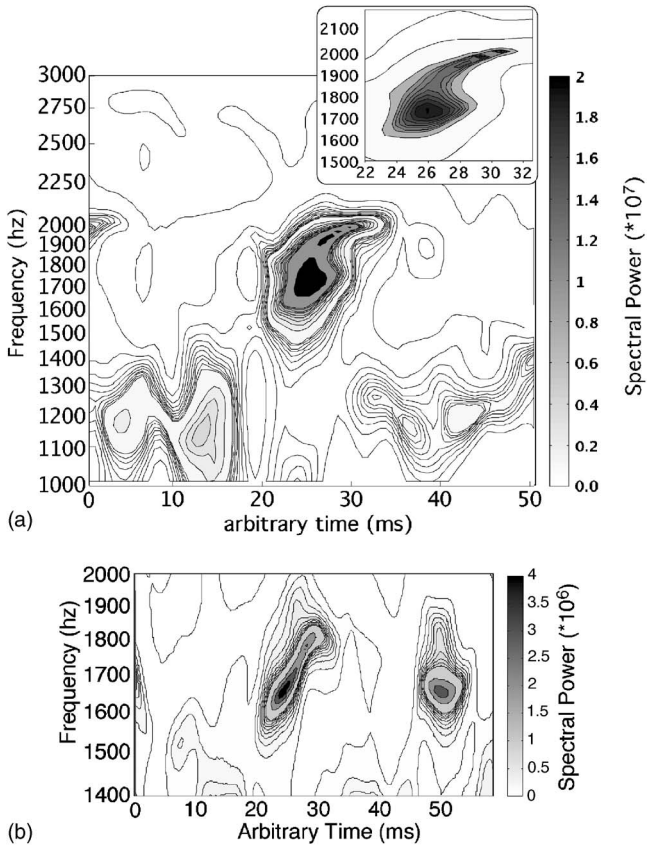


Fig. 3. (a) Spectragram for two or more bubbles produced sequentially separated by a few milliseconds. Inset shows details of bubble spectragram. (b) Spectragram of a bubble pair production followed by a second bubble 20 ms later.

show a frequency shift (Fig. 1). In the latter case, the amplitude had decreased significantly before the next bubble was formed. Also, in some cases, similarly sized bubbles may have been produced nearly simultaneously. Unfortunately, the field SNR was too poor to identify similarly sized bubble pair production.

Seep bubbles are highly likely to be contaminated. Seawater contains numerous surfactants and for marine seep bubbles, significant contamination from sediment, oil, surface-active bacteria and algae, and other marine particles is likely. Also likely important is the geometry and dynamics of seep bubble formation. Seep bubbles often escape from sandy seabed (Leifer and Boles, 2005) or cracks/fractures in rock or hydrate (Leifer and MacDonald, 2003). Thus, seep orifices are typically asymmetric, in contrast to what generally pertains in the laboratory. Another potentially important factor is that currents and surge can lead to asymmetric bubble pinch-off at the vent orifice, which may affect the resultant sound.

Acknowledgments

The authors would like to thank the support under Award No. NA03OAR4300104, subcontract UAF-05-0140 from the National Oceanic and Atmospheric Administration (NOAA), U.S. Department of Commerce, and the Office of Naval Research. Further, this work could not have been completed without the help of Timothy Wen, Tonya del Sontro, Frank Kinnamin, and Dave Farrar. Thanks also to the Crew of the R/V Velero and the Delta submarine. The statements,

findings, conclusions, and recommendations are those of the authors and do not necessarily reflect the views of NOAA, the Department of Commerce, or the University of California, Santa Barbara.

References and links

Asher, W. E., Karle, L. M., and Higgins, B. J. (**1997**). "On the difference between bubble-mediated air-water transfer in freshwater and sea water," J. Mar. Res. **55**, 1–34.

Boyd, J. W. R., and Varley, J. (**2001**). "The uses of passive measurement of acoustic emissions from chemical engineering processes," Chem. Eng. Sci. **56**, 1749–1767.

Clark, J. F., Leifer, I., Washburn, L., and Luyendyk, B. P. (**2003**). "Compositional changes in natural gas bubble plumes: Observations from the Coal Oil Point marine hydrocarbon seep field," Geo-Mar. Lett. **23**, 187–193.

Clester, S. M., Hornafius, J. S., Scepan, J., and Estes, J. E. (**1996**). "Quantification of the relationship between natural gas seepage rates and surface oil volume in the Santa Barbara Channel (abstract)," EOS Trans. Am. Geophys. Union **77**, F419.

Deane, G. B. (**1997**). "Sound generation and air-entrainment by breaking waves in the surf zone," J. Acoust. Soc. Am. **101**, 233–240.

Hornafius, J. S., Quigley, D. C., and Luyendyk, B. P. (**1999**). "The world's most spectacular marine hydrocarbons seeps (Coal Oil Point, Santa Barbara Channel, California): Quantification of emissions," J. Geophys. Res., [Oceans] **104**, 20,703–20, 711.

Hsiao, P.-Y., Devaud, M., and Bacri, J.-C. (**2001**). "Acoustic coupling between two air bubbles in water," Eur. Phys. J. E **4**, 5–10.

Kvenvolden, K. A., and Rogers, B. W. (**2005**). "Gaia's breath-Global methane exhalations," Mar. Pet. Geol. **22**, 579–590.

Leifer, I., and Boles, J. (**2005**). "Measurement of marine hydrocarbon seep flow through fractured rock and unconsolidated sediment," Mar. Pet. Geol. **22**, 551–568.

Leifer, I., and MacDonald, I. R. (**2003**). "Dynamics of the gas flux from shallow gas hydrate deposits: Interaction between oily hydrate bubbles and the oceanic environment," Earth Planet. Sci. Lett. **210**, 411–424.

Leifer, I., De Leeuw, G., and Cohen, L. H. (2003). "Optical measurement of bubbles: System, design and application," **20**, 1317–1332.

Leifer, I., Luyendyk, B. P., and Broderick, K. (**2006**). "Tracking an oil slick from multiple natural sources, Coal Oil Point, California," Mar. Pet. Geol. **23**, 621–630.

Leighton, T. G., and Walton, A. J. (**1987**). "An experimental study of the sound emitted by gas bubbles in a liquid," Eur. J. Phys. **8**, 98–104.

Longuet-Higgins, M. S. (**1989**). "Monopole emission of sound by asymmetric bubble oscillations. I. Normal modes," J. Fluid Mech. **201**, 525–541.

Manasseh, R., Yoshida, S., and Rudman, M. (**1998**). "Bubble formation processes and bubble acoustic signals," in *Proceedings of the Third International Conference on Multiphase Flow*, Lyon, France, 8–12 June.

Manasseh, R., LaFontaine, R. F., Davy, J., Shepherd, I., and Zhu, Y.-G. (**2001**). "Passive acoustic bubble sizing in sparged systems," Exp. Fluids **30**, 672–682.

Marple, S. L. Jr. (**1988**). *Digital Spectral Analysis with Applications* (Prentice-Hall, Englewood Cliffs, NJ).

Medwin, H., and Breitz, N. D. (**1989**). "Ambient and transient bubble spectral densities in quiescent seas and under spilling breakers," J. Geophys. Res., [Oceans] **94C**, 12,751–12,759.

Minnaert, M. (**1933**). "On musical air bubbles and the sound of running water," Philos. Mag. **16**, 235–248.

Nikolovska, A., Manasseh, A., and Ooi, A. (**2003**). "Temporal evolution of the frequency of a bubble chain," in *Proceedings of the Fifth International Symposium on Cavitation*, Osaka Japan, 1–4, Nov.

Patro, R., Leifer, I., and Bowyer, P. (**2002**). "Better bubble process modeling: Improved bubble hydrodynamics parameterization," in *Gas Transfer and Water Surfaces*, edited by M. Donelan, W. Drennan, E. S. Salzman, and R. Wanninkhof, AGU Monograph, **127**, 315–320.

Payne, E. M. B., Manasseh, R., and Ooi, A. (**2004**). "On collective oscillations of bubble chains and arrays," in *Proc. 15th Australasian Fluid Mechanics Conference*, Sydney, Australia, 31–17, Dec.

Sinha, N. N. (**2003**). "Bubble-based chemical sensing for the process industries," Philos. Mag. **83**, 2815–2827.

Strasberg, M. (**1955**). "Gas bubbles as sources of sound in liquids," J. Acoust. Soc. Am. **28**, 20–26.

Tang, D. (**2005**). "Inverting for sandy sediment sound speed in very shallow water using boat noise," in *Boundary Influences in High Frequency, Shallow Water Acoustics*, edited by N. G. Pace and P. Blondel (University of Bath, UK), pp. 13–20.

The use of resonant scattering to identify stone fracture in shock wave lithotripsy

Neil R. Owen, Michael R. Bailey, and Lawrence A. Crum
Center for Industrial and Medical Ultrasound, Applied Physics Laboratory, University of Washington, 1013 NE 40th Street, Seattle, Washington 98105
nowen@apl.washington.edu; bailey@apl.washington.edu; lac@apl.washington.edu

Oleg A. Sapozhnikov and Leonid A. Trusov
Department of Acoustics, Faculty of Physics, M.V. Lomonosov Moscow State University, Leninskie Gory, Moscow 119992, Russia
oleg@acs366.phys.msu.ru; leotrusov@gmail.com

Abstract: There is currently little feedback as to whether kidney stones have fractured during shock wave lithotripsy. Resonant scattering of the lithotripter shock wave was used here to differentiate intact and fractured stone models in water. Scattering, including reflection and radiation due to reverberation from within the stone, was calculated numerically with linear elasticity theory and agreed well with measurements made with a focused receiver. Identification of fracture was possible through frequency analysis, where scatter from fractured stones was characterized by higher energy in distinct bands. High-speed photography concurrent with measurement indicated the effect was not due to cavitation.

PACS numbers: 43.20.Fn, 43.20.Gp, 43.80.Gx, 43.25.Vt [ANN]
Date Received: August 9, 2006 **Date Accepted:** October 2, 2006

1. Introduction

Indication that a kidney stone has fractured during shock wave lithotripsy (SWL) would aid a urologist in deciding to continue or cease treatment. Stopping treatment on a stone that is not breaking, and a good percentage of stones do not break, would potentially spare the patient acute injury and reduce the risk of chronic complications.[1,2] Our goal was to detect resonant scattering from kidney stone models to see if a stone fractured into two pieces could be distinguished from an intact stone. Resonant scattering is the term we use to describe the radiation caused by reverberations within a stone struck by a shock wave.

Although there is a vast field of techniques to detect calcifications in the body[3–6] and an equally large field of research on acoustic scattering,[7–9] clinical lithotripsy relies on B-mode ultrasound or fluoroscopy to target the stone. Little feedback is provided by these methods as to whether a stone has fractured. Generally, in the confines of the body to the resolution of the systems, it is difficult to distinguish an intact stone from a collection of stone fragments. Some research on feedback to target and track the stone has been completed. Thomas, Wu, and Fink[10] used time-reversal techniques in a peizoceramic lithotripter to target the stone. Borhis, Bayer, and Lechner[11] used spectral Doppler ultrasound to determine if a stone had been hit by a shock wave. Owen *et al.*[12] reported the use of vibro-acoustography to track the stone in water and thereby improved comminution. Chang *et al.*[13] developed a real-time tracking method to improve comminution, where analysis of ultrasound images was used to track the stone and to move the lithotripter accordingly. As for feedback on whether a stone has fractured, Fedele *et al.*[14] described a diagnostic sensor and signal processing technique, both under patent review, to assess the degree of stone fragmentation and stone location. The technique uses cavitation signals, where the secondary acoustic emission from bubble collapse is dependent on stone size.

Like Fedele *et al.*, we listen remotely with a broadband receiver during SWL, however our objective is to detect a difference in resonant scattering due to fractured stone pieces being

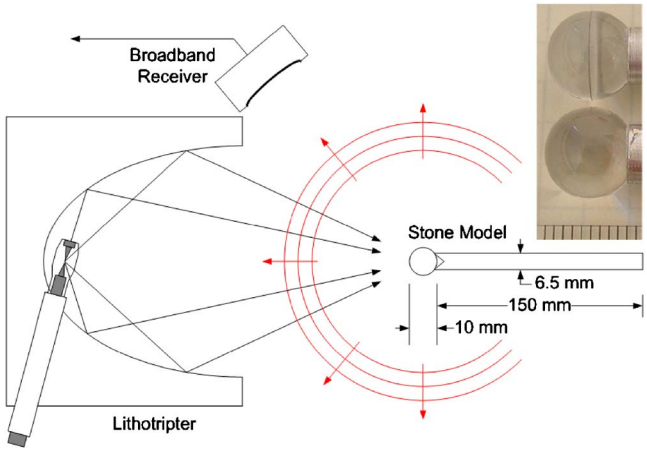


Fig. 1. (Color online) Illustration of the experiment for the calculation and measurement of acoustic scatter. The inset is a photograph of the intact and fractured stone models (marks on the scale are millimeters).

smaller than the intact stone. In this work, resonant scatter signals were calculated numerically and measured experimentally, and then compared to determine the scattering from stone models. A frequency analysis method was developed to distinguish between the intact and fractured stone models. During the *in vitro* experiments, high-speed photography was used to visualize cavitation.

2. Theory

Numerical simulation of the experiment (Fig. 1) was accomplished in three steps. First, a finite-difference time-domain numerical model for elastic wave propagation, developed originally to study stresses within kidney stones due to the impact of shock waves,[15] was used to calculate pressure in the liquid and stress in the stone model. An axisymmetric grid cast in cylindrical coordinates was initialized with a planar lithotripter shock pulse[16] with 40 MPa peak positive pressure, and with elastic coefficients (Table 1) representing the stone model and surrounding water. Results from this step were rendered into a movie (Mm. 1) to display the evolution of pressure and stress. Second, radiation from the stone model was calculated with the Helmholtz-Kirchhoff integral taken on a cylindrical surface surrounding the stone model.[17] The temporal wave form of acoustic pressure and its normal derivative in each grid point along this cylindrical surface was calculated during the first step. The Helmholtz-Kirchhoff integral allowed calculating the scattered pressure wave form at every point on the receiver. Third, the receiver signal was calculated by averaging acoustic pressure over a spherical bowl surface representing the focused receiver. Results from this final step were both compared to measurements of acoustic scatter and analyzed to determine fracture.

Mm. 1. Results from the linear elastic model that was used in the first step of calculation. Acoustic scatter comprises reflection and reverberation, which are most easily observed to the left of the stone models and along their axes (2.1 Mb). This is a file of type "mov".

Table 1. Elastic properties used in calculation (λ and μ are the Lamé constants)

	Sound speed m/s	Density kg/m^3	λ GPa	μ GPa
Glass	5600	2300	23.1	25
Epoxy	2720	1430	4.36	···
Aluminum	6300	2700	51.8	24
Water	1500	1000	···	···

3. Methods

In degassed water, intact and fractured stone models were placed individually at the focus of an electrohydraulic research lithotripter[18] and subjected to 20 shock waves. The experimental arrangement is illustrated in Fig. 1. Each shock wave was generated with a lithotripter charging potential of 17 kV, corresponding to 25 MPa peak positive pressure and 10 MPa peak negative pressure. The intact stone was modeled with a glass sphere of 5 mm radius that was bonded with epoxy to a counter-sunk aluminum rod for rigid placement in the acoustic field. Similarly, the fractured stone was modeled with two glass hemispheres of 5 mm radius that were bonded together with a 1 mm film of epoxy. A photograph of both models is shown in Fig. 1. Acoustic scatter from each shock wave was measured with a focused receiver, high-pass filtered at 100 kHz (Model 3202, Krohn-Hite Corporation, Brockton, MA), digitized at 50 MS/s (TDS744, Tektronix, Inc., Beaverton, OR) and saved in a computer. The filter was necessary to remove low frequency noise, presumably a combination of a structural resonance of the receiver and transient electrical signals from the high voltage discharge.

The remote broadband receiver was designed to measure acoustic scatter without blocking the acoustic path between the lithotripter and the stone models (Fig. 1). It was fabricated with a polyvinylidene fluoride (PVDF) film of 25 μm thickness and 50 mm diameter that was formed to a spherical curvature with a radius of 150 mm. With the addition of a differential amplifier, the −3 dB pass band was between 100 kHz and 10 MHz. The receiver was fabricated at the Center for Industrial and Medical Ultrasound as an extension of recent literature.[19] Focal properties were characterized by wiring the PVDF film as a source, exciting it at 3.6 MHz, and measuring the resulting acoustic field with a hydrophone (GL-0150-1A, Specialty Engineering Acoustics, Soquel, CA). The axial and lateral dimensions of the pressure focus measured at −6 dB were 45 and 2 mm, respectively.

Effort was made to reduce cavitation in the experiments. Stone models were made of smooth, lens quality glass, because glass wets and adheres well to water. The lithotripter was operated at a low charging voltage of 17 kV to reduce the negative pressure of the shock wave. Each shock wave was triggered 1 min apart to allow bubbles to dissolve.[20] Last, digital photography was used to visualize cavitation. A high-speed camera (Imacon 200, DRS Technologies, Parsippany, NJ), 105 mm lens, and 27.5 mm extender produced a 2.0 cm $\times$ 2.4 cm field of view in 980 $\times$ 1200 pixels.

Calculated and measured signals were processed using MATLAB (The Mathworks, Inc., Natick, MA). Measured signals were deconvolved with a sampled approximation of the high-pass filter's impulse response. Deconvolution removed the transient effects of the filter but also added high frequency noise, caused in part by division in frequency, which was removed with a 10 MHz low-pass filter. Twenty measured wave forms were then aligned in time to correct for jitter in the shock wave trigger timing, and then averaged. Averaging was intended to compensate for shot-to-shot variation in the shock pulse, but had little effect on the signals since they were quite repeatable. Time domain processing was not performed on signals obtained from calculation.

Several steps were used to analyze calculated and measured data in the frequency domain. First, specular reflection of the shock wave, a sequence lasting about 5 μs, was found to be similar for data from both the intact and fractured stone models. To isolate the differences in wave forms from the two models, the sequence of specular reflection was truncated and 15 μs of resonant scattering, resulting from reverberation in the stone and directly following the specular reflection, was retained. Second, to compare the data based on frequency alone, each sequence was normalized by its root-mean-square value. Third, the power spectral density of each sequence was found. Fourth, energy values, E, were calculated, based on Parseval's theorem, by integrating the power spectra over two distinct bands in frequency, from $f_1 = 0$ to $f_2 = 0.5$ MHz and from $f_1 = 0.5$ MHz to $f_2 = 2.0$ MHz

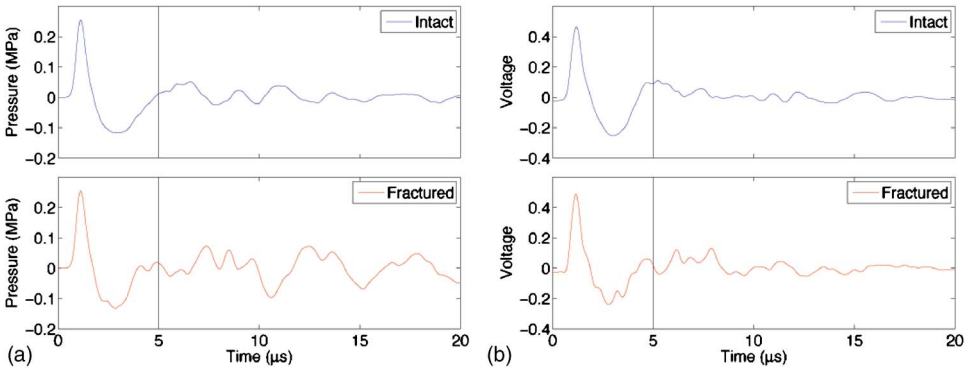


Fig. 2. (Color online) Acoustic scatter from (a) calculation and (b) measurement. The vertical line at 5 μs roughly separates the segments of reflection of the shock wave and resonant scattering caused by reverberation. Segments of reflection (left of the vertical line) were truncated and segments of reverberation (right of the vertical line) were retained for analysis in the frequency domain. Since the PVDF receiver was not calibrated, the amplitude of measured signals are presented in volts.

$$E = \sum_{k=f_1/\Delta f}^{f_2/\Delta f} W_k \cdot \Delta f, \tag{1}$$

where W_k is the discrete power spectral density sequence at frequency $f=k\Delta f$, k is an index variable, f_1 and f_2 are end points of the aforementioned frequency bands, and Δf is the difference between adjacent samples in frequency. Last, in each frequency band a ratio was found by

$$R = \frac{E_F}{E_I}, \tag{2}$$

where R is the ratio of energies, E_F is energy calculated with Eq. (1) using data from the fractured stone model, and E_I is energy calculated with Eq. (1) using data from the intact stone model. A ratio using Eq. (2) was found separately for both calculated and measured data. The expectation was that resonant scattering from the fractured stone model would be characterized by higher frequencies, therefore the ratio of energies is meant to distinguish between the two stone models.

4. Results

Mm. 1 shows calculated acoustic scattering from the two stone models. Glass spheres and hemispheres are traced in black and the aluminum rods are traced in white; all other colors represent a stress or pressure value. Upon impingement, the shock wave is both reflected by the surface of the stone models and transmitted into the stone models. A clear difference in the scattered field is visible when the movie is paused at $t=15$ μs. At that time, to the left of the stone models and along their axes, the reflected shock waves are followed by pressure fluctuations in the water that correspond to reverberations within the stone models. These reverberations represent nothing other than vibrational resonance of the stone after excitation by a broadband load associated with a short lithotripter pulse. The result can therefore be called resonant scattering. As is always the case for vibration of an elastic object, the resonance frequency is inversely proportional to the object size. Compared to the intact stone model, pressure fluctuations emitted by the fractured stone model are closer together and of higher amplitude. This effect is caused by the glass hemispheres being smaller than the glass sphere and by the strong reflection at the planar surfaces between the hemispheres. The difference is less off axis.

Measured and calculated wave forms from intact and fractured stone models are shown in Fig. 2. The vertical lines separate the sequences of reflection, to the left, and resonant scattering, to the right. Results generally agree well when comparing calculation and measure-

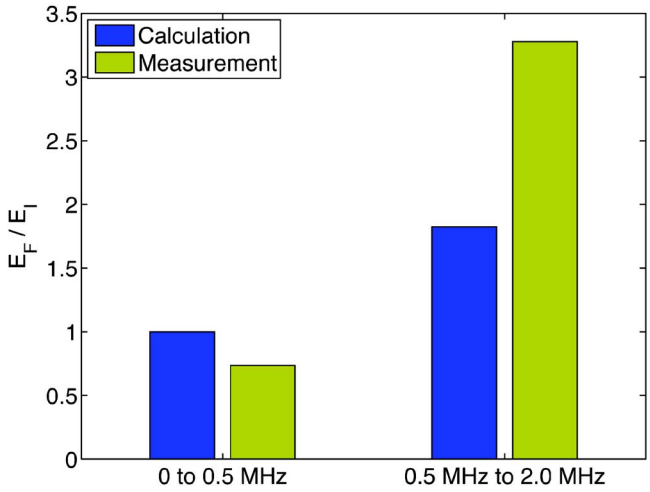


Fig. 3. (Color online) The ratio of energies between fractured and intact stones, $R=E_F/E_I$, in frequency bands for signals from calculation and measurement. Both calculation and measurement show increased energy in the frequency band between 0.5 and 2.0 MHz, which indicates that the intact and fractured stone models can be distinguished.

ment, especially the reflection of the shock wave between times 0 and 5 μs. Wave forms from intact and fractured stone models have noticeable differences. For the intact stone model, one reverberation across the axial diameter of the sphere is 3.6 μs using a sound speed of 5600 m/s (Table 1). Accordingly, pressure fluctuations corresponding to waves within the stone model begin approximately 3.6 μs after the peak positive pressure of the reflected shock wave. Similarly for the fractured stone model, pressure fluctuations begin sooner, at 1.8 μs after the peak positive pressure of the shock wave, and are of higher amplitude. The respective periods of resonant acoustic scatter, 3.6 and 1.8 μs, are evident.

Figure 3 shows the ratios, $R=E_F/E_I$ (see Eq. (2)), found with frequency-domain processing for calculated and measured results. The frequency bands were chosen based on reverberation times for the stone models, where a period of 3.6 μs corresponds to 280 kHz and a period of 1.8 μs corresponds to 560 kHz. Both calculation and measurement show increased energy in the frequency band between 0.5 and 2.0 MHz, indicating that the fractured stone model can be distinguished from the intact stone model.

Figure 4 shows high-speed images of the intact stone model taken simultaneously with acoustic scattering detection. In the photographs, there are no bubbles visible before the shock wave arrives at 0 μs and only tiny bubbles growing at 26 μs, which is after our data have been recorded. The signal below the photographs is the scattering signal recorded not only for the first 20 μs, as in Fig. 2, but for hundreds of microseconds following shock wave arrival. Together the photographs and signal indicated that little noise was received as the bubbles grew and a detectable pressure spike was generated when the bubbles collapsed and rebounded at about 250 and 350 μs. Photographic and recorded results from the fractured stone model were similar.

5. Discusssions and conclusions

Calculated and measured scattering from an intact stone model and a model with a fracture normal to the shock wave axis agree well in the time domain. We speculate minor discrepancies are primarily due to neglecting shear waves in the epoxy in the model; an exact value for the Lamé constant, μ, was not known. Both calculation and measurement show the fractured stone reverberated at a higher frequency and that fractured stones could be distinguished from intact

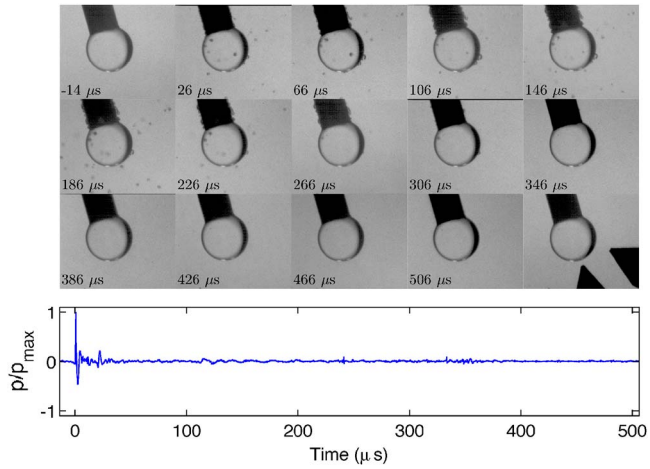


Fig. 4. (Color online) High-speed photography of the intact stone model subjected to a single shock wave and the corresponding measured acoustic scatter wave form. Each photograph is 2.0 cm × 2.4 cm and labeled with a time relative to $t=0$ μs, when the shock wave arrived at the stone. The bottom right photograph illustrates with black pointers the axes of the shock wave (bottom) and the receiver (right). Below the photographs is the measured wave form, where the propagation time between the stone and the receiver has been subtracted. The period of acoustic scatter used for analysis is between times 0 and 20 μs. Bubble collapse and rebound events generally happen much later, at about 250 and 350 μs here, and do not influence the period of resonant scatter.

stones. A method using the ratio of energies in distinct bands, related to stone size, was developed to clearly and simply display the difference between the intact and fractured stone models.

This work was a carefully controlled *in vitro* study, and although clearly obstacles exist, many results are encouraging for further *in vivo* study. Cavitation may occur *in vivo*, and its effect on the signals is not known. However, it was possible to avoid disruption of the measured signals, as evidenced by good agreement with the model which does not account for cavitation, by making an effort to reduce cavitation such as giving bubbles time to dissipate before taking measurements. Real kidney stones can have irregular nonspherical shape, but the dominant effects observed here, which were used to identify fracture, are related to the reverberation time across the length of the stone. Scatter measured remotely from stones *in vivo* may have lower amplitude than glass stones in water, but the signals obtained here are very strong, hundreds of millivolts, so some attenuation or reduction in signal amplitude might be tolerated. Measured signals were intentionally shown as voltage to show this amplitude. For pressure amplitudes, the receiver could be calibrated by comparing the scattered signals to measurements taken with a hydrophone at the location of the scatterer. The calibrated system can then be used to measure high intensity ultrasound fields without damage to a hydrophone.[21,22]

The geometry of the counter-sunk aluminum rods may have affected the experiment by increasing the scattering amplitude over another arrangement. They were meant to hold the model stones rigidly in order to study shot to shot variability in the receive signal and to facilitate averaging, which, for example, would remove random noise from cavitation. However, results indicate the measurement was repeatable and that cavitation was not the dominant effect. Choosing the counter-sunk end was also influenced by the numerical experiment, which was axisymmetric, and intended for the surface area of contact between the glass, epoxy, and rod to be consistent for both model stones. Any effect in frequency caused by the rods would have been the same for both model stones, and therefore would not influence the ratio of energies used to distinguish the model stones. Also, a long length was chosen for the rod to prevent reflection from the distal end from contaminating the measurement. Work in progress involves measuring acoustic scatter from breakable model stones held within a plastic pipette, which allows the stones to move.

The study here looked at resonant acoustic scatter as feedback to identify fracture; however, because the technique essentially detects a change in frequency related to the change in size of the stone, our technique lays the groundwork for assessment of complete comminution. Potentially this concept could be used to determine if fragments are sufficiently small to pass from the body and thus that treatment is completed.

Acknowledgments

The authors thank Brian MacConaghy and Adam Maxwell for design and fabrication of the spherical receiver; the investigators, staff, and students at the Center for Industrial and Medical Ultrasound; and the Consortium for Shock Waves in Medicine. This work was supported by: NIH DK43881, DK55674, NSBRI SMS00402, and RFBR 05-02-16987.

References and links

[1]J. E. Lingeman and J. R. Newmark, "Adverse bioeffects of shock-wave lithotripsy," in *Kidney Stones: Medical and Surgical Management*, edited by F. L. Coe *et al.* (Lippencott-Raven, Philidelphia, 1996), pp. 605–614.
[2]A. P. Evan, A. R. Willis, J. E. Lingeman, and J. A. McAteer, "Renal trauma and the risk of long-term complications in shock wave lithotripsy," Nephron **78**, 1–8 (1998).
[3]P. N. Rao, "Imaging for kidney stones," World J. Urol. **22**, 323–327 (2004).
[4]L. Gao, K. J. Parker, R. M. Lerner, and S. F. Levinson, "Imaging of the elastic properties of tissue—A review," Ultrasound Med. Biol. **22**, 959–977 (1996).
[5]M. Fatemi and J. F. Greenleaf, "Ultrasound-stimulated vibro-acoustic spectrography," Science **280**, 82–85 (1998).
[6]K. Nightingale, M. S. Soo, R. Nightingale, and G. Trahey, "Acoustic radiation force impulse imaging: In vivo demonstration of clinical feasibility," Ultrasound Med. Biol. **28**, 227–235 (2002).
[7]A. D. Pierce, *Acoustics: An Introduction to its Physical Principles and Applications* (Acoustical Society of America, New York, 1991).
[8]A. Ishimaru, *Wave Propagation and Scattering in Random Media* (IEEE, Piscataway, NJ, 1997).
[9]P. M. Morse, *Vibration and Sound* (Acoustical Society of America, New York, 1981).
[10]J. L. Thomas, F. Wu, and M. Fink, "Time reversal focusing applied to lithotripsy," Ultrason. Imaging **18**, 106–121 (1996).
[11]C. Bohris, T. Bayer, and C. Lechner, "Hit/miss monitoring of ESWL by spectral Doppler ultrasound," Ultrasound Med. Biol. **5**, 705–712 (2003).
[12]N. R. Owen, M. R. Bailey, A. Maxwell, B. MacConaghy, T. D. Khokhlova, and L. A. Crum, "Vibro-acoustography for targeting kidney stones during lithotripsy," J. Acoust. Soc. Am. **116**, 2509 (2004).
[13]C. C. Chang, I. Manousakas, Y. R. Pu, M. Liang, C. H. Chen, F. M. Yu, W. H. Yang, Y. C. Tong, and C. L. Kuo, "In vitro study of ultrasound based real-time tracking for renal stones in shock wave lithotripsy: Part II—A simulated animal experiement," J. Urol. (Baltimore) **6**, 2594–2957 (2002).
[14]F. Fedele, A. J. Coleman, T. G. Leighton, P. R. White, and A. M. Hurrell, "Development of a new diagnostic sensor for extra-corporeal shock-wave lithotripsy," in *Proceedings of the First Mediterranean Conference on Medical and Biological Engineering*, "Health in the Information Society," pp. 134–139, (Naples, Italy, 2004).
[15]R. O. Cleveland and O. A. Sapozhnikov, "Modeling elastic wave propagation in kidney stones with application to shock wave lithotripsy," J. Acoust. Soc. Am. **118**, 2667–2676 (2005).
[16]C. C. Church, "A theoretical study of cavitation generated by an extracorporeal shock wave lithotripter," J. Acoust. Soc. Am. **86**, 215–227 (1989).
[17]O. A. Sapozhnikov, L. A. Trusov, N. R. Owen, M. R. Bailey, and R. O. Cleveland, "Detecting fragmentation in kidney stones in lithotripsy by means of shock wave scattering," in *Fifth International Symposium on Therapeutic Ultrasound*, edited by G. T. Clement, N. J. McDonald, and K. Hynynen (American Institute of Physics, New York, 2005), pp. 308–312.
[18]R. O. Cleveland, M. R. Bailey, N. Fineberg, B. Hartenbaum, M. Lokhandwalla, J. A. McAteer, and B. Sturtevant, "Design and characterization of a research electrohydraulic lithotripter patterned after the Dornier HM3," J. Acoust. Soc. Am. **71**, 2514–2525 (2000).
[19]J. A. Ketterling, O. Aristiżabal, D. H. Turnbull, F. L. Lizzi, "Design and fabrication of a 40 MHz annular array transducer," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **52**, 672–681 (2005).
[20]T. G. Leighton, *The Acoustic Bubble* (Academic, London, 1994).
[21]P. Kaczkowski, B. Cunitz, V. Khokhlova, and O. A. Sapozhnikov, "High resolution mapping of nonlinear MHz ultrasonic fields using a scanned scatterer," in *IEEE Ultrasonics Symposium* (Honolulu, Hawaii2003), pp. 982–985.
[22]M. E. Schafer, J. Gessert, and W. Moore, "Development of a high intensity focused ultrasound (HIFU) hydrophone system," in *Fifth International Symposium on Therapeutic Ultrasound*, edited by G. T. Clement, N. J. McDonald, and K. Hynynen (American Institute of Physics, New York, 2005), pp. 609–613.

Excitation of polymer-shelled contrast agents with high-frequency ultrasound

Jeffrey A. Ketterling and Jonathan Mamou
Riverside Research Institute, Frederic L. Lizzi Center for Biomedical Engineering, New York, New York 10006
ketterling@rrinyc.org; mamou@rrinyc.org

John S. Allen III
Department of Mechanical Engineering, University of Hawaii-Manoa, 2540 Dole Street Honolulu Hawaii 96822
alleniii@hawaii.edu

Orlando Aristizábal, Rene G. Williamson, and Daniel H. Turnbull
Skirball Institute of Biomolecular Medicine, New York University School of Medicine, 540 First Avenue New York, New York 10016
oarist@saturn.med.nyu.edu; rene_williamson@yahoo.com; turnbull@saturn.med.nyu.edu

Abstract: Few experimental and complementary theoretical studies have investigated high-frequency (>20 MHz) nonlinear responses from polymer-shelled ultrasound contrast agents. Three polymer agents with different shell properties were examined for their single-bubble backscatter when excited with a 40 MHz tone burst. Higher-order harmonic responses were observed for the three agents; however, their occurrence was at least partly due to nonlinear propagation. Only one of the agents (1.1 μm mean diameter) showed a subharmonic response for longer excitations ($\approx$10–15 cycles) and midlevel pressure excitations ($\approx$2.5 MPa). Theoretical calculations of the backscattered spectrum revealed behavior similar to the experimental results in specific parameter regimes.

PACS numbers: 43.80.Qf, 43.25.Yw [CC]
Date Received: September 12, 2006 **Date Accepted:** October 31, 2006

1. Introduction

Applications involving microcirculation, such as ophthalmic disease diagnosis and small-animal imaging, particularly the normal and abnormal vascular development of genetically engineered mouse embryos and neonates, are ideal for high-frequency ultrasound (HFU, >20 MHz) contrast agents. The acoustic contrast agents in current use were designed to have an optimal response in the 1–10 MHz range of standard clinical ultrasound probes. These agents usually have a mean diameter of 2–4 μm where a larger bubble typically has a lower resonance frequency. Agents specifically designed for use with HFU have yet to become available, but many existing and experimental agents have been studied for HFU responses while utilizing a broadband pulse.[1–3] The results typically showed backscatter centered at the transducer center frequency but no harmonic content was reported.

Goertz *et al.*[4] recently reported the generation of harmonic and subharmonic backscatter components for the lipid-shelled agent Definity (Bristol-Myers Squibb, New York, NY). The contrast agents were excited with a tone burst consisting of 4–10 cycles at either 20 or 30 MHz. The observation of harmonics is surprising because few agents are typically present in the submicron size range in a given dose and, from the theory of free bubbles, the expectation would be that relatively high pressures would be required to excite a bubble nonlinearly with HFU. Previous theoretical work by Allen *et al.*[5] suggests that shell waves may contribute to an enhanced high-frequency dipole response. Deng *et al.*[2] reported experimental results that suggest a high-frequency response from conventional-size protein-shelled agents may be due to

elastic properties of the shell. A new generation of polymer agents is becoming available, but most studies with these agents have been done with low MHz ultrasound[6–8] and not much is known about their high-frequency response.

In this paper we report experimental and theoretical backscatter results for nitrogen-filled, polycaprolactone-shelled agents (POINT Biomedical, San Carlos, CA) with mean diameters of 3.4 (Point 1668, P1), 1.1 (Point 1466, P2) and 0.56 (Point 1470, P3) μm. These agents were excited with tone bursts from 1–15 cycles using a 40 MHz, focused transducer. The backscatter signals from single agents were examined for harmonic content. Experimental results were compared to theoretical calculations for the case of the polymer shell remaining intact during excitation.

2. Methods

2.1 Theory

For polymer-shelled agents, experimental evidence indicates the elasticity and thickness of the shell are important components to the overall dynamics.[9] The model of Church,[10] developed for protein-shelled agents, and a more-recent model based on finite elasticity,[11] are the most appropriate models for these considerations. Both models were examined for this study but here, for brevity, we report initial results only from the Church model.

The model was modified to include compressibility corrections for the surrounding liquid as has been done for free bubble equations. Adiabatic behavior of the interior gas was assumed using a polytropic exponent relationship (γ=1.4) and van der Waals hard core terms for air. Shell thickness was estimated to be 5 nm and surface tension values of 50 and 5 dyn/cm were used for the shell/liquid and shell/gas interfaces, respectively. The density of the shell was assumed to be 1.1 gm/cm^3 and the shell shear viscosity was approximated as 5.0 Poise. We estimated a shear modulus of 8.0 MPa for the shell under the assumption that the polymer behaved as a rubbery material. The radius-time, $R(t)$, curve of the bubble response was calculated and the radiated pressure was computed from $p(t) \propto R^2\ddot{R}+2R\dot{R}^2$.[12] The theoretical calculations did not account for the bandpass-filtering effect of the transducer or acoustic attenuation in water.

2.2 Experimental system

Experiments were performed with a dilute solution of contrast agent passing through a flow phantom. The flow phantom had a 3 mm channel with fluid ports on each side. The rate of flow was controlled by a planetary gear-driven digital pump (Cole Parmer, Vernon Hills, IL). The focal region of a 40 MHz probe attached to a Vevo 770 (Visual Sonics, Toronto, Canada) ultrasound backscatter microscope was placed within the flow phantom, and the passage of the contrast agents was visualized in real time. A pc containing a PCI-based digitizer (Acqiris DP105, Monroe, NY) was linked to the Vevo 770, and a custom LABVIEW program (National Instruments, Austin, TX) permitted acquisition of radio-frequency (rf) data corresponding to the B-mode display of the Vevo 770.

The agents were reconstituted from a dry state, according to manufacturer's instructions, using milli-Q water at a final concentration of 50 mg/ml. The samples were pumped from a reservoir into the flow phantom at a flow rate of 1 μl/s and diluted until single bubble flow could be observed on the Vevo 770 B-mode display. While an agent passed through the transducer focal zone, rf data frames were digitized at rates of up to 2 frames/s. The rf frames were acquired at a sampling rate of 500 MHz and consisted of 384 scan lines with 5000 points/line. The pulse repetition frequency was 8 kHz and the data frames spanned 8 mm, which corresponded to $\approx$21 μm separation between scan lines. Data were acquired for several Vevo 770 settings between 1% and 100% power and excitation pulses consisting of 1, 5, 10, and 15 cycles. Acoustic pressure levels for the power settings were measured with a calibrated hy-

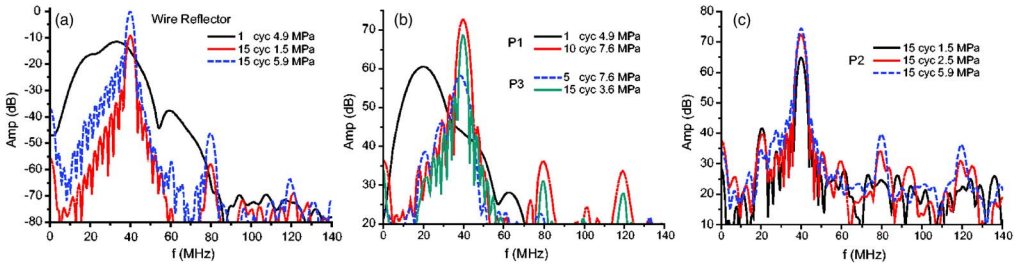


Fig. 1. (Color online) (a) Wire reflector spectra for 1 cycle and 15 cycle tone bursts. (b) Spectra for agents P1 and P3 reveal harmonic components while only (c) P2 reveals any subharmonic activity.

drophone (Precision Acoustics Ltd., Dorset, UK) and negative peak pressures will be utilized to characterize the exposure settings. In addition, pulse/echo reference spectra were acquired using a 25-μm-diam wire situated at the transducer geometric focus.

2.3 Data analysis

The rf data were analyzed manually to examine the backscatter from single bubbles passing through the transducer focal zone. A Hamming window was placed over an instance of a single contrast agent, and the windowed rf data were padded to 1024 points. The spectrum of the signal was then calculated. The spectra we report were not normalized to remove the system frequency response. We chose the rf line in the "middle" of the bubble for our analysis. Because the rf data were acquired with a real-time imaging system, each bubble received multiple exposures to the ultrasound burst. In fact, bubbles could be observed moving through the image frame over multiple rf-frame acquisitions. These exposure conditions represent a more realistic examination of how a contrast agent withstands a long-term exposure during a HFU imaging procedure. In these studies, we generally did not observe obvious contrast agent destruction, which indicates the agents are fairly robust at high frequency on a time scale of several seconds.

In addition to calculating the spectrum of single-bubble backscatter, we performed a time-frequency analysis of the data using spectrograms. For our studies, spectrograms are an ideal tool to evaluate how bubbles respond to short-duration (e.g., 1 or 5 cycles) or long-duration excitation signals (e.g., 10 or 15 cycles) and to evaluate the onset and nature of nonlinear behavior (e.g., subharmonic or harmonic components). To obtain the spectrogram, the short-time Fourier transform was computed over 128 samples (i.e., =0.26 μs) and then zero padded to 8192 samples. To achieve adequate temporal resolution, consecutive Fourier transforms overlapped by 120 points.

3. Results

3.1 Experiment

The experimental results for the three agents along with the wire phantom reference spectra are summarized in Fig. 1. Figure 1(a) shows reference spectra obtained with the wire reflector for three 40 MHz tone-burst excitation conditions. The case with 1 cycle and 4.9 MPa essentially shows the system impulse response with a broadband spectrum centered at 33 MHz and additional harmonic activity. When the tone burst was increased to 15 cycles, the center frequency was seen to shift to 40 MHz (the quoted transducer frequency) and harmonics from nonlinear propagation appeared at 80 and 120 MHz. The asymmetric sidelobes seen in the spectra relate to the convolution of the frequency response of the transducer (centered at 33 MHz) and the 40 MHz tone burst.

The backscatter spectra for representative cases of P1 and P3 are shown in Fig. 1(b). The case for P1 with 1 cycle and 4.9 MPa shows a response similar to that seen in Fig. 1(a), except that the center frequency shifted to a lower value (20 MHz). The remaining spectra for P1 and P3 are for excitations with 5–15 cycles. No subharmonic responses were observed at

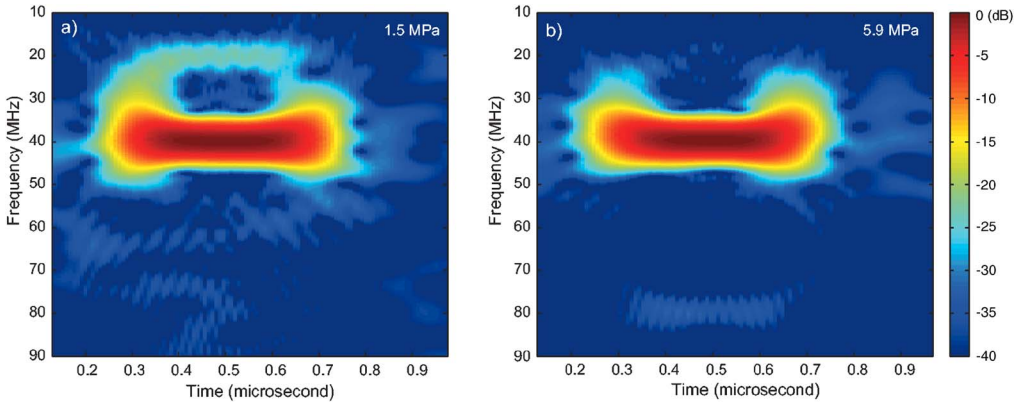


Fig. 2. (Color online) Selected spectrograms for the P2 agent, 15 cycle excitation at 40 MHz with (a) 1.5 MPa and (b) 5.9 MPa negative peak pressure excitations.

any of the Vevo 770 excitation conditions but higher-order harmonics appeared with longer tone bursts and higher excitation pressures. These observations are consistent with what has been reported in the literature for lipid-shelled agents.[4] However, based on the reference spectra, a large portion of the higher-order harmonics appear to result from nonlinear propagation rather than a nonlinear bubble response.

Figure 1(c) shows spectra for P2 using a 15 cycle excitation at three different drive pressures. The 1.5 MPa setting is interesting because a clear subharmonic was present, but there were no significant higher harmonics. When the pressure was raised to 2.5 MPa, the subharmonic was maintained and the second and third harmonic also appeared. Finally, when the pressure was raised to 5.9 MPa, the subharmonic was suppressed but the higher harmonics were still visible. This behavior is consistent with what is generally observed for a subharmonic response.[13] However, our observations are somewhat surprising considering the rigid polymer shell of the agent and the suboptimal driving conditions for a resonance response. One explanation may be that the POINT agent ruptured and that we were actually seeing the backscatter from a free bubble or an unconstrained bubble fragment still linked to the fractured shell.

Figure 2 displays spectrograms of the P2 agent excited with a 40 MHz, 15 cycle pulse at two exposure levels: 1.5 and 5.9 MPa [black and blue curves in Fig. 1(c)]. Figure 2(a), obtained at the lower pressure, displays a significant spectral component near the 40 MHz drive frequency throughout the bubble response. However, the spectrogram also reveals a significant nonlinear spectral component near 20 MHz, 25 dB below the 40 MHz component. Figure 2(b), obtained at the higher exposure pressure, also shows energy at 40 MHz, but with no visible subharmonic component. However, there was some energy near 80 MHz at 35 dB below the fundamental. As noted above, a portion of the higher harmonic content is probably due to nonlinear propagation. However, we believe that a nonlinear response from the agents also contributes.

3.2 Theoretical

Figure 3(a) shows the $R(t)$ curve of a 1.1-μm-diam agent forced with a Gaussian weighted, 40 MHz, 15 cycle, cosine pulse with a peak pressure of 5.9 MPa. An initial delay time of 0.1 μs was used for the pulse. At first, the agent, which was forced well above its resonance frequency, appears to oscillate at the forcing frequency of the transducer. As the pulse begins to decay, the onset of the subharmonic becomes visible in the $R(t)$ curve [Fig. 3(a)]. Due to the density difference between the shell and fluid, additional nonlinearities influence the dominating inertial response of the agent, especially as it is forced above its resonance. The spectrum pressure curve is shown in Fig. 3(b). The subharmonic is visible at a similar level to the experimental results and higher harmonic components appear, although at higher levels than observed experi-

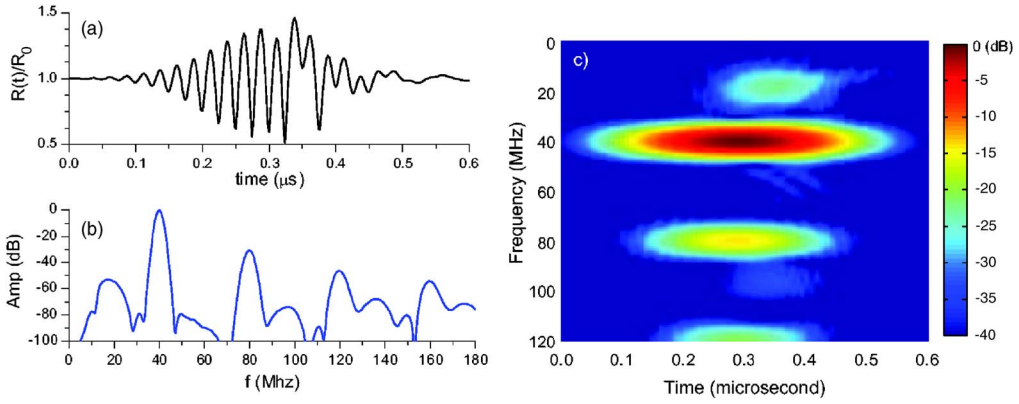


Fig. 3. (Color online) Theoretical (a) $R(t)/R_0$ curve, (b) spectrum of the pressure calculated from the $R(t)$ curve, and (c) spectrogram of pressure curve for P2 excited by a 40 MHz, 15 cycle, 5.9 MPa excitation.

mentally. The spectrogram of the $p(t)$ curve [Fig. 3(c)] shows the time evolution of the harmonic responses. The subharmonic initiates towards the middle of excitation at a later time than the higher-order harmonics, an observation that provides information on the level of dissipation in the system.

We investigated a range of parameter space for the shell properties by varying the thickness and elasticity to estimate the values used in this study. Future experimental measurements of the shell properties of polymer agents will facilitate a more direct comparison of experiments to theoretical calculations. We note that the theoretical subharmonic response occurs at somewhat higher pressure amplitude than observed experimentally; however, the simulations reveal this response is quite sensitive to small variations in wave form, pressure amplitude, shell thickness, and elastic modulus. Also, an earlier response in time of the subharmonic in the experimental data compared with the theory suggests that a transient process such as the nonspherical buckling of the shell or escape of gas might contribute to its onset.

4. Discussion and conclusions

Three polymer-shelled agents from POINT Biomedical were examined to determine their response to a 40 MHz, tone-burst excitation. The three agents showed indications of higher harmonics, but a large portion of the response could be attributed to nonlinear propagation rather than a total nonlinear agent response. Only one agent, P2, was observed to have a subharmonic response for the tested exposure conditions. We should point out that the results reported here were repeatable, but did show variability. The bubbles in each batch of agents had a size distribution and they do not all respond in an identical fashion.

A theoretical analysis of the P2 agent's radial response revealed some similarity to the experimental results. These preliminary results suggest that polymer-shelled contrast agents are viable candidates for high-frequency imaging using the fundamental frequency. However, for an optimal nonlinear response, the polymer shell material properties need to be carefully selected. More comprehensive experiments, measurement of material properties, and a complementary theory are needed to fully understand and optimize polymer-shelled contrast agents for HFU applications.

From the theory of free gas bubbles, we might expect smaller agents to be more attractive for high-frequency nonlinear imaging. However, our results for the three polymer-shelled agents indicate that this is not necessarily the case. One hypothesis for a larger agent giving a better nonlinear response is that the agent must first rupture for it to become acoustically active. If this were true, smaller agents would be more resistant to acoustic rupture because of their

stiffer shell and, therefore, more likely to scatter only the fundamental. Our experimental results also indicate that even if the agents were to have ruptured, the fragments or gas bubbles remain acoustically active for several seconds.

Acknowledgments

The authors wish to thank POINT Biomedical and Bob Short for making available the polymer-shelled agents used in these studies and Visual Sonics for providing access to the engineering mode of the Vevo 770. This work was supported in part by NIH Grant No. HL078665 (D.H.T.).

References and links

[1]K. Morgan, P. Dayton, A. Klibanov, G. Brandenburger, S. Kaul, K. Wei, and K. Ferrara, "Properties of contrast agents insonified at frequencies above 10 MHz," *IEEE Ultrason. Symp.* 1127–1130 (1996).

[2]C. X. Deng, F. L. Lizzi, R. H. Silverman, R. Ursea, and D. J. Coleman, "Imaging and spectrum analysis of contrast agents in the in vivo rabbit eye using very-high-frequency ultrasound," Ultrasound Med. Biol. **24**, 383–394 (1998).

[3]C. M. Moran, R. J. Watson, K. A. Fox, and W. N. McDicken, "In vitro acoustic characterization of four intravenous ultrasonic contrast agents at 30 MHz,"Ultrasound Med. Biol. **28**, 785–791 (2002).

[4]D. E. Goertz, E. Cherin, A. Needles, R. Karshafian, A. S. Brown, P. N. Burns, and F. S. Foster, "High frequency nonlinear b-scan imaging of microbubble contrast agents," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **52**, 65–79 (2005).

[5]J. S. Allen, D. E. Kruse, and K. W. Ferrara, "Shell waves and acoustic scattering from ultrasound contrast agents," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **48**, 409–418 (2001).

[6]A. Bouakaz, M. Versluis, and N. de Jong, "High-speed optical observations of contrast agent destruction," Ultrasound Med. Biol. **31**, 391–399 (2005).

[7]S. H. Bloch, R. E. Short, K. W. Ferrara, and E. R. Wisner, "The effect of size on the acoustic response of polymer-shelled contrast agents," Ultrasound Med. Biol. **31**, 439–444 (2005).

[8]W. S. Chen, T. J. Matula, A. A. Brayman, and L. A. Crum, "A comparison of the fragmentation thresholds and inertial cavitation doses of different ultrasound contrast agents," J. Acoust. Soc. Am. **113**, 643–651 (2003).

[9]H. Leong-Poi, J. Song, S. J. Rim, J. Christiansen, S. Kaul, and J. R. Lindner, "Influence of microbubble shell properties on ultrasound signal: Implications for low-power perfusion imaging," J. Am. Soc. Echocardiogr. **15**, 1269–1276 (2002).

[10]C. C. Church, "The effects of an elastic solid surface layer on the radial pulsations of gas bubbles," J. Acoust. Soc. Am. **97**, 1510–1521 (1995).

[11]J. S. Allen and M. M. Rashid, "Dynamics of a hyperelastic gas-filled spherical shell in a viscous fluid," J. Appl. Mech. **71**, 195–200 (2004).

[12]T. G. Leighton, *The Acoustic Bubble* (Academic, London, 1994).

[13]W. T. Shi, F. Forsberg, J. S. Raichlen, L. Needleman, and B. B. Goldberg, "Pressure dependence of subharmonic signals from contrast microbubbles," Ultrasound Med. Biol. **25**, 275–283 (1999).

ACOUSTICAL NEWS—USA

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

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

Announcement of the 2007 Election

In accordance with the provisions of the bylaws, the following Nominating Committee was appointed to prepare a slate for the election to take place on 25 May 2007: William A. Kuperman, Chair; James W. Beauchamp, N. Ross Chapman, Joseph W. Dickey, Andrea Megela-Simmons, Lily M. Wang.

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

CHARLES E. SCHMID
Executive Director

The Nominating Committee has submitted the following slate:

For President-Elect

Mark F. Hamilton

Donna L. Neff

For Vice President-Elect

Judy R. Dubno

Victor W. Sparrow

 0001-4966/2007/121(1)/1/2/$23.00

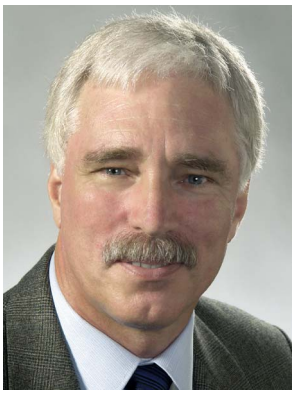
Fredericka Bell-Berti

Thomas J. Matula

Clark S. Penrod

Carl J. Rosenberg

Scott D. Sommerfeldt



Michael R. Stinson

USA Meetings Calendar

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

2007

4–8 June	153rd Meeting of the Acoustical Society of America, Salt Lake City, Utah [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org; WWW: http://asa.aip.org].
22–24 Oct.	NoiseCon 2007, Reno, NV [Website is www.inceusa.org/nc07
27 Nov.–2 Dec.	154th Meeting of the Acoustical Society of America, New Orleans, LA (note Tuesday through Saturday) [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org; WWW: http://asa.aip.org].

2008

29 June–4 July	Joint Meeting of the Acoustical Society of America, European Acoustics Association, and the Acoustical Society of France, Paris, France) [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org; WWW: http://asa.aip.org].
28 July–1 Aug.	9th International Congress on Noise as a Public Health Problem (Quintennial meeting of ICBEN, the International Commission on Biological Effects of Noise). Foxwoods Resort, Mashantucket, CT [Jerry V. Tobias, ICBEN 9, Post Office Box 1609, Groton CT 06340-1609, Tel. 860-572-0680; E-mail icben2008@att.net. WWW: www.icben.org.

ACOUSTICAL STANDARDS NEWS

Susan B. Blaeser, Standards Manager
ASA Standards Secretariat, Acoustical Society of America, 35 Pinelawn Rd., Suite 114E, Melville, NY 11747 [Tel.: (631) 390-0215; Fax: (631) 390- 0217; e-mail: asastds@aip.org]

George S. K. Wong
Acoustical Standards, Institute for National Measurement Standards, National Research Council, Ottawa, Ontario K1A 0R6, Canada [Tel.: (613) 993-6159; Fax: (613) 990-8765; e-mail: george.wong@nrc.ca]

American National Standards (ANSI Standards) developed by Accredited Standards Committees S1, S2, S3, and S12 in the areas of acoustics, mechanical vibration and shock, bioacoustics, and noise, respectively, are published by the Acoustical Society of America (ASA). In addition to these standards, ASA publishes Catalogs of Acoustical Standards, both National and International. To receive copies of the latest Standards Catalogs, please, contact Susan B. Blaeser.

Comments are welcomed on all material in Acoustical Standards News.

This Acoustical Standards News section in JASA, as well as the National and International Catalogs of Acoustical Standards, and other information on the Standards Program of the Acoustical Society of America, are available via the ASA home page http://asa.aip.org.

New Standards Website

ASA Standards has launched its new website which you can access from the ASA homepage http://asa.aip.org and click the "Standards Info" button. The site contains general information about each S Committee and U.S. TAG as well as specific information about meetings, working groups, new work items, other news and forms. We hope the make this site useful to you so please let us know what you would like to see there.

Standards Meetings Calendar—National

•4–8 June 2007

Meetings of the ASA Committee on Standards (ASACOS), ASACOS Steering, ASC S1 Acoustics, ASC S3 Bioacoustics, and ASC S12 Noise, Standards Plenary Group and the U.S. TAGs for ISO/TC 43 Acoustics, ISO/TC 43/SC 1 Noise, and IEC/TC 29 Electroacoustics will be held during the 153rd ASA Meeting, Salt Lake City, Utah.

•18–22 June 2007

The meetings of ASC S2 Mechanical Vibration and Shock and the U.S. TAGs to ISO/TC 108 and its five subcommittees will be held in San Antonio, TX in conjunction with the meeting of the Vibration Institute.

International Standards Committee Meetings

•16–20 April 2007

Meeting of ***ISO/TC 108/SC 5, Condition monitoring and diagnostics of machines,*** Prague, Czech Republic

SC 5 and all of its working groups will be joined by the following additional working groups:

TC 108/WG 31 Balancing
TC 108/SC 2/WG 1 Vibration of machines
TC 108/SC 2/WG 7 Vibration of machines with active magnetic bearings
TC 108/SC 2/WG 10 Basic techniques for vibration diagnostics.

•23–27 April 2007

Meetings of ***ISO/TC 108, TC/108 SC 2, TC 108/SC 3 and TC 108/SC 6***, joined by their working groups (except those listed above), will meet in Berlin, Germany.

Details about these meetings may be obtained from the Secretariat.

Approval of Reaccreditations

We are pleased to announce that Accredited Standards Committees, ASC S1—Acoustics; ASC S2—Mechanical, Vibration and Shock; ASC S3—Bioacoustics; and ASC S12—Noise, have been reaccredited by ANSI's Executive Standards Council, effective September 28, 2006. The most significant change in the new operating procedures is the ability of these committees to form Subcommittees which will be allowed to function as consensus bodies.

Accredited Standards Committee on Acoustics, S1

(J. P. Seiler, Chair; G. S. K. Wong, Vice Chair)

Scope: 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, tolerances and comfort.

S1 Working Groups

S1/Advisory—Advisory Planning Committee to S1 (G. S. K. Wong);
S1/WG1—Standard Microphones and their Calibration (V. Nedzelnitsky);
S1/WG4—Measurement of Sound Pressure Levels in Air (M. A. Nobile, Chair; E. Dunens, Vice Chair);
S1/WG5—Band Filter Sets (A. H. Marsh);
S1/WG17—Sound Level Meters and Integrating Sound Level Meters (B. M. Brooks);
S1/WG19—Insertion Loss of Windscreens (A.J. Campanella);
S1/WG20—Ground Impedance (K. Attenborough, Chair; J. M. Sabatier, Vice Chair);
S1/WG22—Bubble Detection and Cavitation Monitoring (Vacant);
S1/WG25—Specification for Acoustical Calibrators (P. Battenberg);
S1/WG26—High Frequency Calibration of the Pressure Sensitivity of Microphones (A. J. Zuckerwar);
S1/WG27—Acoustical Terminology (J. S. Vipperman);
S1/WG28—Passive Acoustic Monitoring for Marine Mammal Mitigation for Seismic Surveys (A. M. Thode).

S1 Inactive Working Groups:

S1/WG15—Noise Canceling Microphones (R. McKinley, Chair)
S1/WG16—FFT Acoustical Analyzers (R. J. Peppin, Chair)

S1/WG21—Electromagnetic Susceptibility (EMS) of Acoustical Instruments (J. P. Seiler, Chair)
S1/WG24—Design Response of Weighting Networks for Acoustical Measurements (G. S. K. Wong, Chair)

S1 STANDARDS ON ACOUSTICS

ANSI S1.1-1994 (R 2004) American National Standard Acoustical Terminology
ANSI S1.4-1983 (R 2006) American National Standard Specification for Sound Level Meters
ANSI S1.4A-1985 (R 2006) Amendment to ANSI S1.4-1983
ANSI S1.6-1984 (R 2006) American National Standard Preferred Frequencies, Frequency Levels, and Band Numbers for Acoustical Measurements
ANSI S1.8-1989 (R 2006) American National Standard Reference Quantities for Acoustical Levels
ANSI S1.9-1996 (R 2006) American National Standard Instruments for the Measurement of Sound Intensity
ANSI S1.11-2004 American National Standard Specification for Octave-Band and Fractional-Octave-Band Analog and Digital Filters
ANSI S1.13-2005 American National Standard Measurement of Sound Pressure Levels in Air
ANSI S1.14-1998 (R 2003) American National Standard Recommendations for Specifying and Testing the Susceptibility of Acoustical Instruments to Radiated Radio-Frequency Electromagnetic Fields, 25 MHz to 1 GHz
ANSI S1.15-1997/Part 1 (R 2006) American National Standard Measurement Microphones, Part 1: Specifications for Laboratory Standard Microphones
ANSI S1.15-2005/Part 2 American National Standard Measurement Microphones, Part 2: Primary Method for Pressure Calibration of Laboratory Standard Microphones by the Reciprocity Technique
ANSI S1.16-2000 (R 2005) American National Standard Method for Measuring the Performance of Noise Discriminating and Noise Canceling Microphones
ANSI S1.17/1-2000/Part 1 American National Standard Microphone Windscreens- Part 1: Measurements and Specification of Insertion Loss in Still or Slightly Moving Air
ANSI S1.18-1999 (R 2004) American National Standard Template Method for Ground Impedance
ANSI S1.20-1988 (R 2003) American National Standard Procedures for Calibration of Underwater Electroacoustic Transducers
ANSI S1.22-1992 (R 2002) American National Standard Scales and Sizes for Frequency Characteristics and Polar Diagrams in Acoustics
ANSI S1.24 TR-2002 ANSI Technical Report Bubble Detection and Cavitation Monitoring
ANSI S1.25-1991 (R 2002) American National Standard Specification for Personal Noise Dosimeters
ANSI S1.26-1995 (R 2004) American National Standard Method for Calculation of the Absorption of Sound by the Atmosphere
ANSI S1.40-2006 American National Standard Specifications and Verification Procedures for Sound Calibrators
ANSI S1.42-2001 (R 2006) American National Standard Design Response of Weighting Networks for Acoustical Measurements
ANSI S1.43-1997 (R 2002) American National Standard Specifications for Integrating-Averaging Sound Level Meters

Accredited Standards Committee on Mechanical Vibration and Shock, S2

(R. L. Eshleman, Chair; A. T. Herfat, Vice Chair)

Scope: Standards, specifications, methods of measurement and test, and terminology in the field of mechanical vibration and shock, and condition monitoring and diagnostics of machines, including the effects of exposure to mechanical vibration and shock on humans, including those aspects which pertain to biological safety, tolerance and comfort.

S2 Working Groups

S2/WG 1—S2 Advisory Planning Committee (R. L. Eshleman, Chair; A. T. Herfat, Vice Chair);
S2/WG2—Terminology and Nomenclature in the Field of Mechanical Vibration and Shock and Condition Monitoring and Diagnostics of Machines (D. J. Evans);
S2/WG3—Signal Processing Methods (T. S. Edwards);
S2/WG4—Characterization of the Dynamic Mechanical Properties of Viscoelastic Polymers (W. Madigosky, Chair; J. Niemic, Vice Chair);
S2/WG5—Use and Calibration of Vibration and Shock Measuring Instruments (D. J. Evans, Chair; B. E. Douglas, Vice Chair);
S2/WG6—Vibration and Shock Actuators (G. B. Booth);
S2/WG7—Acquisition of Mechanical Vibration and Shock Measurement Data (B. E. Douglas);
S2/WG8—Analysis Methods of Structural Dynamics (B. E. Douglas);
S2/WG9—Training and Accreditation (R. L. Eshleman, Chair; D. Corelli, Vice Chair);
S2/WG10—Measurement and Evaluation of Machinery for Acceptance and Condition (R. L. Eshleman, Chair; H. C. Pusey, Vice Chair);
S2/WG10/Panel 1—Balancing (R. L. Eshleman);
S2/WG10/Panel 2—Operational Monitoring and Condition Evaluation - R. Bankert);
S2/WG10/Panel 3—Machinery Testing (R. L. Eshleman);
S2/WG10/Panel 4—Prognosis (R. L. Eshleman);
S2/WG10/Panel 5—Data Processing, Communication, and Presentation (K. Bever);
S2/WG11—Measurement and Evaluation of Mechanical Vibration of Vehicles (A. F. Kilcullen);
S2/WG12—Measurement and Evaluation of Structures and Structural Systems for Assessment and Condition Monitoring (B. E. Douglas, Chair);
S2/WG13—Shock Test Requirements for Shelf-Mounted and Other Commercial Electronic Systems (J. Baker, Chair);
S2/WG39—Human Exposure to Mechanical Vibration and Shock—Parallel to ISO/TC 108/SC 4 (D. D. Reynolds, Chair; R. Dong, Vice Chair).

S2 Inactive Working Groups:

S2/WG54—Atmospheric Blast Effects (J. W. Reed, Chair; J. H. Keefer, Vice Chair).

S2 STANDARDS ON MECHANICAL VIBRATION AND SHOCK

ANSI S2.1-2000 / ISO 2041:1990 American National Standard Vibration and Shock—Vocabulary (Nationally Adopted International Standard)
ANSI S2.2-1959 (R 2006) American National Standard Methods for the Calibration of Shock and Vibration Pickups
ANSI S2.4-1976 (R 2004) American National Standard Method for Specifying the Characteristics of Auxiliary Analog Equipment for Shock and Vibration Measurements
ANSI S2.7-1982 (R 2004) American National Standard Balancing Terminology
ANSI S2.8-1972 (R 2006) American National Standard Guide for Describing the Characteristics of Resilient Mountings
ANSI S2.9-1976 (R 2006) American National Standard Nomenclature for Specifying Damping Properties of Materials
ANSI S2.16-1997 (R 2006) American National Standard Vibratory Noise Measurements and Acceptance Criteria of Shipboard Equipment
ANSI S2.17-1980 (R 2004) American National Standard Techniques of Machinery Vibration Measurement
ANSI S2.19-1999 (R 2004) American National Standard Mechanical Vibration - Balance Quality Requirements of Rigid Rotors, Part 1: Determination of Permissible Residual Unbalance, Including Marine Applications
ANSI S2.20-1983 (R 2006) American National Standard Estimating Airblast Characteristics for Single Point Explosions in Air, with a Guide to Evaluation of Atmospheric Propagation and Effects
ANSI S2.21-1998 (R 2002) American National Standard Method for Preparation of a Standard Material for Dynamic Mechanical Measurements

ANSI S2.22-1998 (R 2002) American National Standard Resonance Method for Measuring the Dynamic Mechanical Properties of Viscoelastic Materials

ANSI S2.23-1998 (R 2002) American National Standard Single Cantilever Beam Method for Measuring the Dynamic Mechanical Properties of Viscoelastic Materials

ANSI S2.24-2001 (R 2006) American National Standard Graphical Presentation of the Complex Modulus of Viscoelastic Materials

ANSI S2.25-2004 American National Standard Guide for the Measurement, Reporting, and Evaluation of Hull and Superstructure Vibration in Ships

ANSI S2.26-2001 (R 2006) American National Standard Vibration Testing Requirements and Acceptance Criteria for Shipboard Equipment

ANSI S2.27-2002 American National Standard Guidelines for the Measurement and Evaluation of Vibration of Ship Propulsion Machinery

ANSI S2.28-2003 American National Standard Guidelines for the Measurement and Evaluation of Vibration of Shipboard Machinery

ANSI S2.29-2003 American National Standard Guidelines for the Measurement and Evaluation of Vibration of marine Shafts on Shipboard Machinery

ANSI S2.31-1979 (R 2004) American National Standard Method for the Experimental Determination of Mechanical Mobility, Part 1: Basic Definitions and Transducers

ANSI S2.32-1982 (R 2004) American National Standard Methods for the Experimental Determination of Mechanical Mobility, Part 2: Measurements Using Single-Point Translational Excitation

ANSI S2.34-1984 (R 2005) American National Standard Guide to the Experimental Determination of Rotational Mobility Properties and the Complete Mobility Matrix

ANSI S2.42-1982 (R 2004) American National Standard Procedures for Balancing Flexible Rotors

ANSI S2.43-1984 (R 2005) American National Standard Criteria for Evaluating Flexible Rotor Balance

ANSI S2.46-1989 (R 2005) American National Standard Characteristics to be Specified for Seismic Transducers

ANSI S2.48-1993 (R 2006) American National Standard Servo-Hydraulic Test Equipment for Generating Vibration - Methods of Describing Characteristics

ANSI S2.60-1987 (R 2005) American National Standard Balancing Machines - Enclosures and Other Safety Measures

ANSI S2.61-1989 (R 2005) American National Standard Guide to the Mechanical Mounting of Accelerometers

ANSI S2.70-2006 American National Standard Guide for the Measurement and Evaluation of Human Exposure to Vibration Transmitted to the Hand *(revision of ANSI S3.34-1986)*

ANSI S2.71-1983 (R 2006) American National Standard Guide to the Evaluation of Human Exposure to Vibration in Buildings *(reaffirmation and redesignation of ANSI S3.29-1983).*

Accredited Standards Committee on Bioacoustics, S3

(C. A. Champlin, Chair; R. F. Burkard, Vice Chair)

Scope: Standards, specifications, methods of measurement and test, and terminology in the fields of psychological and physiological acoustics, including aspects of general acoustics, which pertain to biological safety, tolerance and comfort.

S3 Working Groups

S3/Advisory—Advisory Planning Committee to S3 (C. A. Champlin, Chair; R. F. Burkard, Vice Chair);

S3/WG35—Audiometers (R. L. Grason);

S3/WG36—Speech Intelligibility (R. S. Schlauch);

S3/WG37—Coupler Calibration of Earphones (B. Kruger);

S3/WG39—Human Exposure to Mechanical Vibration and Shock (D. D. Reynolds, Chair; R. Dong, Vice Chair);

S3/WG43—Method for Calibration of Bone Conduction Vibrator (J. D. Durrant);

S3/WG48—Hearing Aids (D. A. Preves);

S3/WG51—Auditory Magnitudes (R. P. Hellman);

S3/WG56—Criteria for Background Noise for Audiometric Testing (J. Franks);

S3/WG59—Measurement of Speech Levels (L. A. Wilber and M. C. Killion, Co-Chairs);

S3/WG60—Measurement of Acoustic Impedance and Admittance of the Ear (Vacant);

S3/WG62—Impulse Noise with Respect to Hearing Hazard (J. H. Patterson, Chair; R. Hamernik, Vice Chair);

S3/WG67—Manikins (M. D. Burkhard);

S3/WG72—Measurement of Auditory Evoked Potentials (R. F. Burkard);

S3/WG76—Computerized Audiometry (A. J. Miltich);

S3/WG78—Thresholds (W. A. Yost);

S3/WG79—Methods for Calculation of the Speech Intelligibility Index (C. V. Pavlovic);

S3/WG81—Hearing Assistance Technologies (L. Thibodeau and L. A. Wilber, Co-Chairs);

S3/WG82—Basic Vestibular Function Test Battery (C. Wall, III);

S3/WG83—Sound Field Audiometry (T. R. Letowski);

S3/WG84—Otoacoustic Emission (G. R. Long);

S3/WG86—Audiometric Data Structures (W. A. Cole and B. Kruger, Co-Chairs);

S3/WG88—Standard Audible Emergency Evacuation and Other Signals (I. Mande);

S3/WG89—Spatial Audiometry in Real and Virtual Environments (J. Besing);

S3/WG90—Animal Bioacoustics (A. E. Bowles);

S3/WG91—Text-to-Speech Synthesis Systems (A. K. Syrdal and C. Bickley, Co-Chairs).

S3/WG92—Effects of Sound on Fish and Turtles (R. R. Fay and A. N. Popper Co-Chairs).

S3 Liaison Group:

S3/L-1 S3 U. S. TAG Liaison to IEC/TC 87 Ultrasonics (W. L. Nyborg).

S3 Inactive Working Groups:

S3/WG71 Artificial Mouths (R. McKinley);

S3/WG80 Probe-tube Measurements of Hearing Aid Performance (W. A. Cole);

S3/WG85 Allocation of Noise-Induced Hearing Loss (R. A. Dobie, Chair).

S3 STANDARDS ON BIOACOUSTICS

ANSI S3.1-1999 (R 2003) American National Standard Maximum Permissible Ambient Noise Levels for Audiometric Test Rooms

ANSI S3.2-1989 (R 1999) American National Standard Method for Measuring the Intelligibility of Speech over Communication Systems

ANSI S3.4-2005 American National Standard Procedure for the Computation of Loudness of Steady Sound

ANSI S3.5-1997 (R 2002) American National Standard Methods for Calculation of the Speech Intelligibility Index

ANSI S3.6-2004 American National Standard Specification for Audiometers

ANSI S3.7-1995 (R 2003) American National Standard Method for Coupler Calibration of Earphones

ANSI S3.13-1987 (R 2002) American National Standard Mechanical Coupler for Measurement of Bone Vibrators

ANSI S3.18-2002 / ISO 2631-1:1997 American National Standard Mechanical vibration and shock – Evaluation of human exposure to whole-body vibration—Part 1: General requirements (Nationally Adopted International Standard)

ANSI S3.18-2003/ISO 2631-4: 2001 American National Standard Mechanical vibration and shock—Evaluation of human exposure to whole body vibration—Part 4: Guidelines for the evaluation of the effects of vibration and rotational motion on passenger and crew comfort in fixed-guideway transport systems (Nationally Adopted International Standard)

ANSI S3.20-1995 (R 2003) American National Standard Bioacoustical Terminology
ANSI S3.21-2004 American National Standard Methods for Manual Pure-Tone Threshold Audiometry
ANSI S3.22-2003 American National Standard Specification of Hearing Aid Characteristics
ANSI S3.25-1989 (R 2003) American National Standard for an Occluded Ear Simulator
ANSI S3.35-2004 American National Standard Method of Measurement of Performance Characteristics of Hearing Aids under Simulated Real-Ear Working Conditions
ANSI S3.36-1985 (R 2001) American National Standard Specification for a Manikin for Simulated *in situ* Airborne Acoustic Measurements
ANSI S3.37-1987 (R 2002) American National Standard Preferred Earhook Nozzle Thread for Postauricular Hearing Aids
ANSI S3.39-1987 (R 2002) American National Standard Specifications for Instruments to Measure Aural Acoustic Impedance and Admittance (Aural Acoustic Immittance)
ANSI S3.40-2002/ISO 10819:1996 American National Standard Mechanical vibration and shock—Hand-arm vibration—Method for the measurement and evaluation of the vibration transmissibility of gloves at the palm of the hand (Nationally Adopted International Standard)
ANSI S3.41-1990 (R 2001) American National Standard Audible Emergency Evacuation Signal
ANSI S3.42-1992 (R 2002) American National Standard Testing Hearing Aids with a Broad-Band Noise Signal
ANSI S3.44-1996 (R 2006) American National Standard Determination of Occupational Noise Exposure and Estimation of Noise-Induced Hearing Impairment
ANSI S3.45-1999 American National Standard Procedure for Testing Basic Vestibular Function
ANSI S3.46-1997 (R 2002) American National Standard Methods of Measurement of Real-Ear Performance Characteristics of Hearing Aids

Accredited Standards Committee on Noise, S12

(R. D. Hellweg, Chair; W. J. Murphy, Vice Chair)

Scope: 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.

S12 Working Groups

S12/Advisory—Advisory Planning Committee to S12 (R. D. Hellweg, Chair);
S12/WG3—Measurement of Noise from Information Technology and Telecommunications Equipment (K. X. C. Man, Chair);
S12/WG11—Hearing Protector Attenuation and Performance (E. H. Berger, Chair);
S12/WG13—Method for the Selection of Hearing Protectors that Optimize the Ability to Communicate (D. Byrne, Chair);
12/WG14—Measurement of the Noise Attenuation of Active and/or Passive Level Dependent Hearing Protective Devices (J. Kalb, Chair; W. J. Murphy, Vice Chair);
S12/WG15—Measurement and Evaluation of Outdoor Community Noise (P. D. Schomer);
S12/WG18—Criteria for Room Noise (R. J. Peppin);
S12/WG23—Determination of Sound Power (J. Schmitt and B. M. Brooks, Co-Chairs);
S12/WG31—Predicting Sound Pressure Levels Outdoors (R. J. Peppin, Chair; L. Pater, Vice Chair);
S12/WG32—Revision of ANSI S12.7—1986 Methods for Measurement of Impulse Noise (A. H. Marsh);
S12/WG33—Revision of ANSI S5.1-1971 Test Code for the Measurement of Sound from Pneumatic Equipment (B. M. Brooks);
S12/WG36—Development of Methods for Using Sound Quality (G.L. Ebbit and P. Davies, Co-Chairs);
S12/WG38—Noise Labeling in Products (R. D. Hellweg and J. Pope, Co-Chairs);
S12/WG40—Measurement of the Noise Aboard Ships (S. Antonides, Chair; S. Fisher, Vice Chair);
S12/WG41—Model Community Noise Ordinances (L. Finegold, Chair; B. M. Brooks, Vice Chair);
S12/WG44—Speech Privacy (G.C. Tocci, Chair; D. Sykes, Vice Chair);
S12/WG45—Measurement of Occupational Noise Exposure from Telephone Equipment (A.K. Woo, Chair);
S12/WG46—Acoustical Performance Criteria for Relocatable Classrooms (T. Hardiman and P. D. Schomer, Co-Chairs);
S12/WG47—Underwater Noise Measurements of Ships (M. Bahtiarian, Chair).

S12 Liaison Groups:

S12/L-1 IEEE 85 Committee for TAG Liaison - Noise Emitted by Rotating Electrical Machines (Parallel to ISO/TC 43/SC 1/WG 13) (R. G. Bartheld);
S12/L-2 Measurement of Noise from Pneumatic Compressors Tools and Machines (Parallel to ISO/TC 43/SC 1/WG 9) (Vacant);
S12/L-3 SAE Committee for TAG Liaison on Measurement and Evaluation of Motor Vehicle Noise (parallel to ISO/TC 43/SC 1/WG 8) (R. F. Schumacher);
S12/L-4 SAE Committee A-21 for TAG Liaison on Measurement and Evaluation of Aircraft Noise (J. D. Brooks);
S12/L-5 ASTM E-33 on Environmental Acoustics (to include activities of ASTM E33.06 on Building Acoustics, parallel to ISO/TC 43/SC 2 and ASTM E33.09 on Community Noise) (K. P. Roy);
S12/L-6 SAE Construction-Agricultural Sound Level Committee (I. Douell);
S12/L-7 SAE Specialized Vehicle and Equipment Sound Level Committee (T. M. Disch);
S12/L-8 ASTM PTC 36 Measurement of Industrial Sound (B. M. Brooks, Vice Chair).

S12 Inactive Working Groups:

S12/WG9 Annoyance Response to Impulsive Noise (L. C. Sutherland, Chair);
S12/WG19 Measurement of Occupational Noise Exposure (J. P. Barry and R. Goodwin, Co-Chairs);
S12/WG27 Outdoor Measurement of Sound Pressure Level (G. A. Daigle, Chair);
S12/WG29 Field Measurement of the Sound Output of Audible Public-Warning Devices (Sirens) (P. Graham, Chair);
S12/WG 37 Measuring Sleep Disturbance Due to Noise (K. S. Pearsons, Chair);

S12 STANDARDS ON NOISE

ANSI S12.1-1983 (R 2006) American National Standard Guidelines for the Preparation of Standard Procedures to Determine the Noise Emission from Sources
ANSI S12.2-1995 (R 1999) American National Standard Criteria for Evaluating Room Noise
ANSI S12.3-1985 (R 2006) American National Standard Statistical Methods for Determining and Verifying Stated Noise Emission Values of Machinery and Equipment
ANSI S12.5-1990 (R 1997) American National Standard Requirements for the Performance and Calibration of Reference Sound Sources
ANSI S12.6-1997 (R 2002) American National Standard Methods for Measuring the Real-Ear Attenuation of Hearing Protectors
ANSI S12.7-1986 (R 2006) American National Standard Methods for Measurements of Impulse Noise
ANSI S12.8-1998 (R 2003) American National Standard Methods for Determining the Insertion Loss of Outdoor Noise Barriers

ANSI S12.9-1988/Part 1 (R 2003) American National Standard Quantities and Procedures for Description and Measurement of Environmental Sound, Part 1

ANSI S12.9-1992/Part 2 (R 2003) American National Standard Quantities and Procedures for Description and Measurement of Environmental Sound, Part 2: Measurement of Long-Term, Wide-Area Sound

ANSI S12.9-1993/Part 3 (R 2003) American National Standard Quantities and Procedures for Description and Measurement of Environmental Sound, Part 3: Short-Term Measurements with an Observer Present

ANSI S12.9-2005/Part 4 American National Standard Quantities and Procedures for Description and Measurement of Environmental Sound, Part 4: Noise Assessment and Prediction of Long-Term Community Response

ANSI S12.9-1998/Part 5 (R 2003) American National Standard Quantities and Procedures for Description and Measurement of Environmental Sound, Part 5: Sound Level Descriptors for Determination of Compatible Land Use

ANSI S12.9-2000/Part 6 (R 2005) American National Standard Quantities and Procedures for Description and Measurement of Environmental Sound Part 6: Methods for Estimation of Awakenings Associated with Aircraft Noise Events Heard in Homes

ANSI S12.10-2002 / ISO 7779:1999 American National Standard Acoustics – Measurement of airborne noise emitted by information technology and telecommunications equipment (Nationally Adopted International Standard)

ANSI S12.11-2003/Part 1 / ISO 10302: 1996 (MOD) American National Standard Acoustics – Measurement of noise and vibration of small air-moving devices - Part 1: Airborne noise emission (Modified Nationally Adopted International Standard)

ANSI S12.11-2003/Part 2 American National Standard Acoustics – Measurement of Noise and Vibration of Small Air-Moving Devices - Part 2: Structure-Borne Vibration

ANSI S12.12-1992 (R2002) American National Standard Engineering Method for the Determination of Sound Power Levels of Noise Sources Using Sound Intensity

ANSI S12.13 TR-2002 ANSI Technical Report Evaluating the Effectiveness of Hearing Conservation Programs through Audiometric Data Base Analysis

ANSI S12.14-1992 (R 2002) American National Standard Methods for the Field Measurement of the Sound Output of Audible Public Warning Devices Installed at Fixed Locations Outdoors

ANSI S12.15-1992 (R 2002) American National Standard For Acoustics*B* Portable Electric Power Tools, Stationary and Fixed Electric Power Tools, and Gardening Appliances - Measurement of Sound Emitted

ANSI S12.16-1992 (R 2002) American National Standard Guidelines for the Specification of Noise of New Machinery

ANSI S12.17-1996 (R 2006) American National Standard Impulse Sound Propagation for Environmental Noise Assessment

ANSI S12.18-1994 (R 2004) American National Standard Procedures for Outdoor Measurement of Sound Pressure Level

ANSI S12.19-1996 (R 2006) American National Standard Measurement of Occupational Noise Exposure

ANSI S12.23-1989 (R 2006) American National Standard Method for the Designation of Sound Power Emitted by Machinery and Equipment

ANSI S12.30-1990 (R 2002) American National Standard Guidelines for the Use of Sound Power Standards and for the Preparation of Noise Test Codes

ANSI S12.42-1995 (R 2004) American National Standard Microphone-in-Real-Ear and Acoustic Test Fixture Methods for the Measurement of Insertion Loss of Circumaural Hearing Protection Devices

ANSI S12.43-1997 (R 2002) American National Standard Methods for Measurement of Sound Emitted by Machinery and Equipment at Workstations and other Specified Positions

ANSI S12.44-1997 (R 2002) American National Standard Methods for Calculation of Sound Emitted by Machinery and Equipment at Workstations and other Specified Positions from Sound Power Level

ANSI S12.50-2002 / ISO 3740:2000 American National Standard Acoustics – Determination of sound power levels of noise sources – Guidelines for the use of basic standards (Nationally Adopted International Standard)

ANSI S12.51-2002 / ISO 3741:1999 American National Standard Acoustics – Determination of sound power levels of noise sources using sound pressure – Precision method for reverberation rooms (Nationally Adopted International Standard)

ANSI S12.53/1-1999 (R 2004) / ISO 3743-1:1994 American National Standard Acoustics - Determination of sound power levels of noise sources - Engineering methods for small, movable sources in reverberant fields- Part 1: Comparison method for hard-walled test rooms (Nationally Adopted International Standard)

ANSI S12.53/2-1999 (R 2004) / ISO 3743-2:1994 American National Standard Acoustics - Determination of sound power levels of noise sources using sound pressure- Engineering methods for small, movable sources in reverberant fields- Part 2: Methods for special reverberation test rooms (Nationally Adopted International Standard)

ANSI S12.54-1999 (R 2004) / ISO 3744:1994 American National Standard Acoustics - Determination of sound power levels of noise sources using sound pressure–Engineering method in an essentially free field over a reflecting plane (Nationally Adopted International Standard)

ANSI S12.55-2006 / ISO 3745:2003 American National Standard Acoustics - Determination of sound power levels of noise sources using sound pressure—Precision methods for anechoic and hemi-anechoic rooms (Nationally Adopted International Standard) *(This standard replaces ANSI S12.35-1990)*

ANSI S12.56-1999 (R 2004) / ISO 3746:1995 American National Standard Acoustics - Determination of sound power levels of noise sources using sound pressure - Survey method using an enveloping measurement surface over a reflecting plane (Nationally Adopted International Standard)

ANSI S12.57-2002 / ISO 3747:2000 American National Standard Acoustics—Determination of sound power levels of noise sources using sound pressure—Comparison method *in situ* (Nationally Adopted International Standard)

ANSI S12.60-2002 American National Standard Acoustical Performance Criteria, Design Requirements, and Guidelines for Schools

ANSI S12.65-2006 American National Standard for Rating Noise with Respect to Speech Interference.

ASA Committee on Standards (ASACOS)

ASACOS (P. D. Schomer, Chair and ASA Standards Director)

U. S. Technical Advisory Groups (TAGS) for International Standards Committees

ISO/TC 43 Acoustics, **ISO/TC 43 /SC 1** Noise (P. D. Schomer, U.S. TAG Chair)

ISO/TC 108 Mechanical Vibration and Shock (D. J. Evans, U.S. TAG Chair)

ISO/TC 108/SC2 Measurement and Evaluation of Mechanical Vibration and Shock as Applied to Machines, Vehicles and Structures (A. F. Kilcullen, and R. F. Taddeo U.S. TAG Co-Chairs)

ISO/TC 108/SC3 Use and Calibration of Vibration and Shock Measuring Instruments (D. J. Evans, U.S. TAG Chair)

ISO/TC 108/SC4 Human Exposure to Mechanical Vibration and Shock (D. D. Reynolds, U.S. TAG Chair)

ISO/TC 108/SC5 Condition Monitoring and Diagnostic Machines (D. J. Vendittis, U.S. TAG Chair; R. Taddeo, U.S. TAG Vice Chair)

ISO/TC 108/SC6 Vibration and Shock Generating Systems (G. Booth, U.S. TAG Chair)

IEC/TC 29 Electroacoustics (V. Nedzelnitsky, U.S. Technical Advisor)

Standards News from the United States

(Partially derived from *ANSI Reporter*, and *ANSI Standards Action*, with appreciation)

American National Standards Call for Comment on Proposals Listed

This section solicits comments on proposed new American National Standards and on proposals to revise, reaffirm, or withdraw approval of existing standards. The dates listed in parenthesis are for information only.

REAFFIRMATIONS

ASME (American Society of Mechanical Engineers)

BSR/ASME MFC-5M-1985 (R200x), Measurement of Liquid Flow in Closed Conduits using Transit-Time Ultrasonic Flowmeters (reaffirmation of ANSI/ASME MFC-5M-1985 (R2001)) This Standard applies only to ultrasonic flowmeters that base their operation on the measurement of transit times of acoustic signals. Further, this Standard concerns only the application of such meters when used to measure the volumetric flow rate of a liquid exhibiting homogeneous acoustic properties and flowing in a completely filled closed conduit. (November 7, 2006)

IEEE (Institute of Electrical and Electronics Engineers)

BSR/IEEE 656-1993 (R200x), Standard for the Measurement of Audible Noise from Overhead Transmission Lines (reaffirmation of ANSI/IEEE 656-1993 (R2000))

Uniform procedures are established for manual and automatic measurement of audible noise from overhead transmission lines. Their purpose is to allow valid evaluation and comparison of the audible noise performance of various overhead lines. (December 19, 2006)

New National Adoptions

ASA (ASC S12) (Acoustical Society of America)

BSR S12.5-200x ISO 6926:1999, Acoustics - Requirements for the Performance and Calibration of Reference Sound Sources Used for the Determination of Sound Power Levels (identical national adoption and revision of ANSI S12.5-1990 (R1997))

Reference sound sources are used extensively in "comparison methods" for determining the noise emissions of physically stationary sound sources. A reference sound source, of known sound power output, is used to establish the numerical relationship between the sound power level of a source, in a given location in a given acoustical environment and the space- and time-averaged sound pressure level at a set of microphone positions. (This is an identical national adoption of ISO 6926:1999.) (November 13, 2006)

REVISIONS

ASA (ASC S3) (Acoustical Society of America)

BSR S3.45-200x, Procedures for Testing Basic Vestibular Function (revision of ANSI S3.45-1999)

Defines procedures for performing and reporting a battery of tests for the evaluation of human vestibular function. Six different tests are specified. Stimuli are presented to evoke eye movement by a subject whose response is determined by measurement of electrical signals generated by the eye movements or by means of video imaging. Specifies test procedures, measurements, data analysis, and data reporting requirements. These tests, including the data analysis and reporting procedures, are called the Basic Vestibular Function Test Battery. Test interpretation is not addressed. (November 13, 2006)

Supplements

TIA (Telecommunications Industry Association)

BSR/TIA 470.110-C-1-200x, Telecommunications - Telephone Terminal Equipment - Handset Acoustics Performance Requirements—Addendum 1 (supplement to ANSI/TIA 470-110-C-2004)

This addendum provides changes to TIA 470.110-C; Telecommunications - Telephone Terminal Equipment - Handset Acoustic Performance Requirements. (November 20, 2006)

American National Standards Final Action

The following American National Standards have received final approval from the ANSI Board of Standards Review.

REVISIONS

EIA (Electronic Industries Alliance)

ANSI/EIA/CEA364-28E-2006, Vibration Test Procedure for Electrical Connectors and Sockets (revision of ANSI/EIA 364-28D-1999): 8/28/2006

ASA (ASC S1) (Acoustical Society of America)

ANSI S1.40-2006, Specification and Verification Procedures for Sound Calibrators (revision of ANSI S1.40-1984 (R2001)): 9/20/2006

REAFFIRMATIONS

IEEE (Institute of Electrical and Electronics Engineers)

ANSI/IEEE C37.082-1982 (R2006), Standard Methods for the Measurement of Sound Pressure Levels of AC Power Circuit Breakers (reaffirmation of ANSI/IEEE C37.082-1982 (R2000)): 10/2/2006

Projects Withdrawn from Consideration

An accredited standards developer may abandon the processing of a proposed new or revised American National Standard or portion thereof if it has followed its accredited procedures. The following projects have been withdrawn accordingly:

ASA (ASC S2) (Acoustical Society of America)

As a matter of housekeeping, the ASA Secretariat withdrew PINS (project initiation notice statements) that had been submitted in the 1990s related to several projects that are no longer under consideration in S2 and for which WGs are no longer active. Work on these topics is ongoing in ISO/TC 108/SC 5 and the U.S. input will progress through the U.S. TAG.

BSR S2.XX (S2/WG 95)-199x, Electrical Techniques for the Purposes of Condition Monitoring and Diagnostics of Machines (new standard)
BSR S2.XX (S2/WG 94)-199x, Life Usage Monitoring (new standard)
BSR S2.XX (S2/WG 93)-199x, Condition Monitoring and Diagnostics of Power Transformers (new standard)
BSR S2.XX (S2/WG 92)-199x, Training and Acceleration in the Field of Condition Monitoring and Diagnostics of Machines (new standard)

ASA (ASC S3) (Acoustical Society of America)

As a matter of housekeeping, the ASA Secretariat withdrew the following PINS (project initiation notice statements) that had been submitted in the 1990s because this standard was developed in the ISO committee instead and has been published as ISO 2631-5:2004:

BSR S3.49-200x, Human Response to Repetitive Mechanical Shock (new standard)

Standards News from Abroad

(Partially derived from *ANSI Reporter* and *ANSI Standards Action*, with appreciation)

International Organization for Standardization (ISO)

Newly Published ISO and IEC Standards

Listed here are new and revised standards recently approved and promulgated by ISO—the International Organization for Standardization

ISO Standards

ACOUSTICS (TC 43)

ISO 4869-2/Cor1:2006, Acoustics - Hearing protectors - Part 2: Estimation of effective A-weighted sound pressure levels when hearing protectors are worn—Corrigendum.

MECHANICAL VIBRATION AND SHOCK (TC 108)

ISO 6070/Cor1:2006, Auxiliary tables for vibration generators - Methods of describing equipment characteristics - Corrigendum.

ISO 14839-3:2006, Mechanical vibration - Vibration of rotating machinery equipped with active magnetic bearings—Part 3: Evaluation of stability margin.

EARTH-MOVING MACHINERY (TC 127)

ISO/TR 25398:2006, Earth-moving machinery—Guidelines for assessment of exposure to whole-body vibration of ride-on machines—Use of harmonized data measured by international institutes, organizations and manufacturers.

INDUSTRIAL FANS (TC 117)

ISO 13347-2/Cor1:2006, Industrial fans - Determination of fan sound power levels under standardized laboratory conditions—Part 2: Reverberant room method—Corrigendum.

ISO Technical Specifications

ACOUSTICS (TC 43)

ISO/TS 4869-5:2006, Acoustics—Hearing protectors—Part 5: Method for estimation of noise reduction using fitting by inexperienced test subjects.

IEC Standards

ELECTROACOUSTICS (TC 29)

IEC 60318-5 Ed. 1.0 b:2006, Electroacoustics—Simulators of human head and ear—Part 5: 2 cm3 coupler for the measurement of hearing aids and earphones coupled to the ear by means of ear inserts.

ISO Draft Standard

BUILDING CONSTRUCTION MACHINERY AND EQUIPMENT (TC 195)

ISO/DIS 19433, Building construction machinery and equipment—Pedestrian-controlled vibratory plates—Terminology and commercial specifications - 11/24/2006.

ISO/DIS 19452, Building construction machinery and equipment—Pedestrian-controlled vibratory (percussion) rammers—Terminology and commercial specifications—11/24/2006.

BOOK REVIEWS

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

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

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

Digital Speech Transmission: Enhancement, Coding, and Error Concealment

Peter Vary and Rainer Martin

Wiley, 2006. 608 pp. Price $120 (hardcover).
ISBN: 0471560189

This book deals with the broad field of digital speech transmission mainly applied to digital mobile radio. It presents the underlying theory, principles, and recent practice of digital speech signal processing. The main topics included in this rather comprehensive text are modeling of speech production, speech coding, error concealment, bandwidth extension of speech signals, noise reduction, and acoustic echo control. These are the key technical areas that have had a major impact on successful development of modern speech communication systems. Hence, this book addresses an important technical topic that has had a major impact on communication technology.

The book has 13 chapters and two appendices. It is divided into three parts. The first part (Chapters 2–5) deals with the fundamentals of digital signal processing: models of speech production, the anatomy of hearing, spectral transformations, filter banks for spectral analysis and synthesis, and a review of stochastic processes and estimation theory. Models of speech production, as presented in Chapter 2, are the essential elements for development of efficient speech coding and decoding schemes. Similarly, auditory perception models are used later to guide development of noise reduction and speech bandwidth extension techniques. The review of fundamentals of digital signal processing and stochastic processes makes this book accessible for readers who are interested in speech processing but do not have a strong background in signal processing.

The second part of the book (Chapters 6–8) is the focal point of the book. It focuses on the basic theory and practice of speech coding. This section starts with a thorough discussion of the fundamental elements of speech coding: linear prediction. Linear prediction is a basic technique for estimation of the coefficients of autoregressive (AR) filter coefficients. As derived in Chapter 2, the AR filter closely approximates the speech production model and is a key component of all speech compression algorithms. The chapter on linear prediction is followed by a chapter on various scalar and vector quantization techniques. Quantization is another key element of efficient representation of analogue speech signals in the digital domain. Scalar quantization techniques include: uniform, nonuniform, optimal, and adaptive quantization. In vector quantization, the input vector is placed into a closest codebook cell. Since the codebook is known at the receiver, high transmission efficiency is achieved by transmitting only the codebook index. This is an example of how sophisticated quantization technique can be used for efficient speech transmission.

Chapter 8, entitled *Speech Coding,* is the heart of this book. The idea of speech coding is to transmit acceptable quality speech at significantly reduced bit rates (bits/sample). The speech signal is sampled at the rate of 8 kHz and quantized at 8 bits/sample, resulting in the bit rate of 64 kbit/s. For economic reasons, this bit rate is excessive for wireless communication. Cordless telephones and cellular networks restrict the average bit rate transmission to 0.5–2 bits/sample. Wideband speech transmission requires even greater reduction in the bits per sample, hence more effective speech coding algorithms are required. Chapter 8 develops a comprehensive description of the most important speech coding algorithms, such as linear predictive coding *vocoder, differential pulse code modulation*, and *code excited linear prediction* coding. Other advanced coding techniques are also discussed in this chapter. Speech coding algorithms can be subdivided into wave form, parametric or hybrid coding algorithms. In *wave form coding,* quantized outputs of the analysis filter, such as linear prediction error, are transmitted. Linear predictive error filter coefficients are adjusted adaptively according to correlation properties of the input speech signal. At the receiver, the speech signal is reconstructed by applying the residual signal to the synthesis filter. This technique reduces the dynamic range of the transmitted signal. *Parametric coders (vocoders),* transmit a set of model parameters and control parameters rather than coded wave forms. Speech is reconstructed at the receiver by exciting a time-varying synthesis filter using either white noise or a periodic sequence of pitch pulses in order to synthesize unvoiced and voiced sounds. This configuration can be interpreted as a speech production model, where the time-varying synthesis filter approximates the vocal track model. In the *hybrid coding* approach, the excitation signal is generated similar to waveform coding, and the parameters of a synthesis filter are transmitted as side information. It is pointed out that an important common feature of all three themes is that the time-varying synthesis filter approximates the vocal track transfer function of speech generation. All speech coding algorithms try to achieve acceptable compromises between high speech quality, low bit rate, low complexity, and limited signal delay. Appendix A specifies these parameters for a number of standard speech coding algorithms.

Chapter 9, entitled *Error Concealment and Soft Decision Source Coding,* is the first chapter of the third part of the book that addresses various techniques for speech enhancement, such as error concealment, single channel noise reduction, acoustic echo cancellation, multi-channel noise reduction, beamforming, and bandwidth extension. Chapter 9 discusses some standard error concealment techniques as used in the Global System for Mobile communication and the *Universal Mobile Telecommunication System* mobile systems. These systems use an approach called *hard decision decoding*. Another generic approach is *soft decision decoding* that can be applied to any parametric source decoder because it uses source redundancy, channel reliability information, and parameter estimation. Muting techniques, which significantly improve subjective speech quality in degraded channels, are discussed. Finally an iterative source-channel decoding, that is derived from the *turbo principle*, is discussed too briefly. Turbo codes achieve near optimum error correction coding and decoding. The application of turbo principles to near error free digital transmission is a major breakthrough in almost error free digital communication and deserves more attention. However, the bibliography references a key paper on this subject.

Chapter 10 addresses the topic of bandwidth extension of telephone speech signals, a technique that is used to improve the mediocre audio quality caused by the limited bandwidth of the old analog telephone systems. Bandwidth extension is accomplished by artificially adding some spectral components to the narrowband signal. These components are generated using either information extracted from decoded narrowband transmission or from transmitted side information. Bandwidth extension is a new, interesting research topic.

Chapter 11 is concerned with algorithms for reducing additive noise. In the first part of the chapter, the basic principles of Wiener filters, spectral subtraction technique, and noise power spectral density estimation techniques are presented. The second part of the chapter is devoted to the use of techniques which use nonlinear optimal noise estimators. Maximum likelihood, maximum *a posteriori,* and minimum mean square estimators for estimation of complex DFT coefficients and spectral amplitudes are derived. The chapter concludes with a discussion of spatial correlation properties of sound fields, two-microphone noise cancellation, and implementation of adaptive noise cancellers. Chapter 12 is devoted to multi-channel noise re-

 0001-4966/2007/121(1)/10/2/$23.00

duction. This topic is closely related to the topics typically discussed in the optimum and adaptive array processing literature. Here the reader should use some caution, because array processing literature uses far-field and plain-wave propagation assumptions, which may not be satisfied for acoustic noise sources in typical enclosed acoustic environments.

Acoustic echo control, the topic of the Chapter 13, is crucial for successful implementation of hands-free car phones, multimedia systems that have speech input, and teleconference facilities. In these systems the echo control problem is caused by acoustic coupling between loudspeaker and microphone. In addition to the desired near-field speaker, the microphone picks up undesired noise and troublesome signals from the far-end speaker. Most wireless digital communication networks have relatively long signal delays that, without echo cancellation, cause annoying echoes at the far-end speaker. Due to long two-way transmission delays, echoes are not masked by the far-end speaker's own voice. Thus, echo control is crucial for high quality hands-free communication. Echo cancellation is a classical problem in adaptive signal processing, dating back to analog telephone days, where the echo was generated by imperfect hybrid circuits that coupled two-line local area lines to the four-wire long distance transmission lines. Several early adaptive echo cancellation algorithms addressed this problem. This chapter focuses on new and important applications, evaluation criteria, a number of well-known adaptive algorithms, and other echo control techniques. The algorithms that are derived and studied in detail are: *least mean squares* (LMS), *nonlinear least mean squares*, *recursive least-squares*, and *affine projection* algorithms. Due to its simplicity, the LMS algorithm is the workhorse of adaptive algorithms. Frequency domain adaptive filtering is also examined. Echo cancellation may not provide sufficient echo suppression; therefore additional echo control measures are necessary. These measures include combining an echo canceller with a center clipper, voice controlled switching, and adaptive post filtering. In summary, the basic principles, algorithms, and implementation of echo cancellation and other echo control systems are thoroughly presented in this chapter.

The book concludes with two appendices. Appendix A presents some of the most common coders-decoders (codec) with their features, properties, and implementation details. This is an extremely valuable addition to the book because it provides a strong connection between basic principles and theory, as it is presented in preceding chapters, and actual practice and international standards as presented in Appendix A. Appendix B summarizes information on speech quality assessment. Speech quality measure and methods are important for evaluation of speech transmission systems.

In conclusion, this book is a thorough review of recent development and practice in the broad field of digital speech transmission. It is accessible to readers with some background in digital signal processing and speech processing. The first part of the book (Chapters 3–5) provides additional background on spectral analysis, digital filter banks, stochastic signals, and estimation. A nice feature is that each chapter is followed by a comprehensive and up-to-date bibliography on the topics covered, enabling the interested reader to dig deeper into topics that interest him/her. The book is highly recommended for readers who are interested in the principles and state of the art of digital speech transmission.

LEON H. SIBUL
Applied Research Laboratory
The Pennsylvania State University
University Park, Pennsylvania 16804
and Technical University of Tallinn
Tallinn, Estonia

REVIEWS OF ACOUSTICAL PATENTS

Lloyd Rice
11222 Flatiron Drive, Lafayette, Colorado 80026

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

Reviewers for this issue:

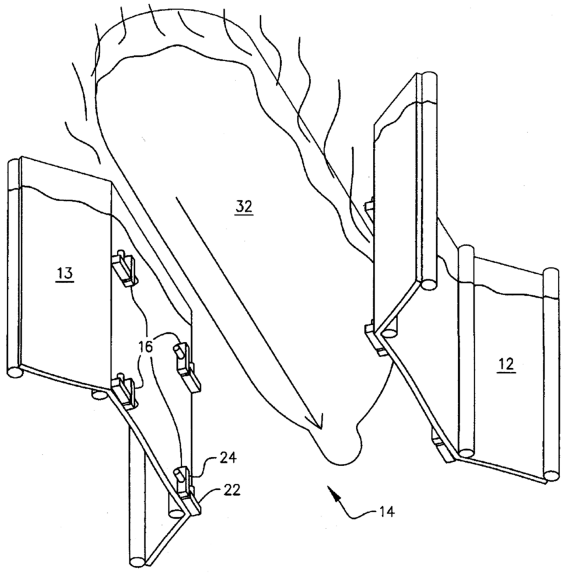
GEORGE L. AUGSPURGER, *Perception, Incorporated, Box 39536, Los Angeles, California 90039*
ANGELO CAMPANELLA, *3201 Ridgewood Drive, Hilliard, Ohio 43026-2453*
ALIREZA DIBAZAR, *Department of BioMed Engineering, University of Southern California, Los Angeles, California 90089*
JOHN M. EARGLE, *JME Consulting Corporation, 7034 Macapa Drive, Los Angeles, California 90068*
SEAN A. FULOP, *California State University, Fresno, 5245 N. Backer Avenue M/S PB92, Fresno, California 93740-8001*
JEROME A. HELFFRICH, *Southwest Research Institute, San Antonio, Texas 78228*
DAVID PREVES, *Starkey Laboratories, 6600 Washington Ave. S., Eden Prarie, Minnesota 55344*
CARL J. ROSENBERG, *Acentech Incorporated, 33 Moulton Street, Cambridge, Massachusetts 02138*
NEIL A. SHAW, *Menlo Scientific Acoustics, Inc., Post Office Box 1610, Topanga, California 90290*
WILLIAM THOMPSON, JR., *Pennsylvania State University, University Park, Pennsylvania 16802*
ERIC E. UNGAR, *Acentech, Incorporated, 33 Moulton Street, Cambridge, Massachusetts 02138*
ROBERT C. WAAG, *University of Rochester, Department of Electrical and Computer Engineering, Rochester, New York 14627*

7,072,244

43.30.Xm UNDERWATER EXTERIOR SHIP HULL IMAGING SYSTEM EMPLOYING A REMOTE MICROPROCESSOR CONTROLLED ACOUSTIC TRANSDUCER ARRAY

Mark S. Rogers, assignor to Hull Underwater Imaging Systems, Incorporated
4 July 2006 (Class 367/88); filed 4 June 2004

A multibeam acoustic transducer array for producing three-dimensional (3D) high-resolution images of a moving ship's hull **32** is described. The transducers **16** consist of orthogonal L-shaped pairs positioned on opposite sides of a shipping channel **14**. The arrays can either be



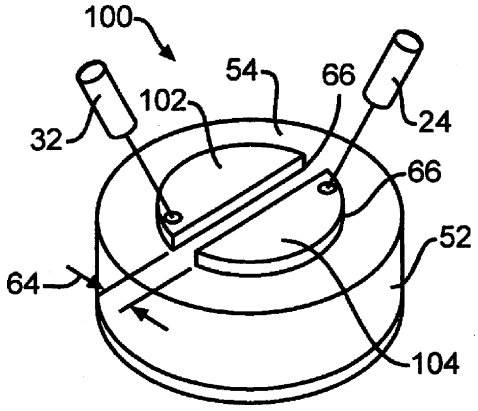
mounted in a fixed configuration on the fenders **12** and **13** of a defined shipping channel or be suspended from a mobile support vessel. One of the transducers in each of the pairs produces a broad horizontal beam but a narrow vertical beam while its orthogonal companion does just the opposite. The echoes from the ship's hull are transmitted to a central processing unit to create visual images.—WT

7,075,216

43.35.Zc LATERAL FIELD EXCITED ACOUSTIC WAVE SENSOR

John F. Vetelino, assignor to University of Maine System Board of Trustees
11 July 2006 (Class 310/338); filed 31 March 2004

This patent discloses the use of transverse shear-mode resonators composed of a piezoelectric substrate material to sense the presence of surface-adsorbed materials. The author attempts to differentiate this device from standard thickness shear-mode resonators by citing the lateral electric field arrangement from the slot electrodes and the alignment of this slot with particular crystal axes of the substrate. It is difficult to see what is really



novel here, as this is simply a special case of a family of devices that have been thoroughly investigated. The author gives several examples of sensing applications for the invention, such as dielectric constant measurements in liquids and gases, chemical sensors, and conductivity measurements of liquid solutions.—JAH

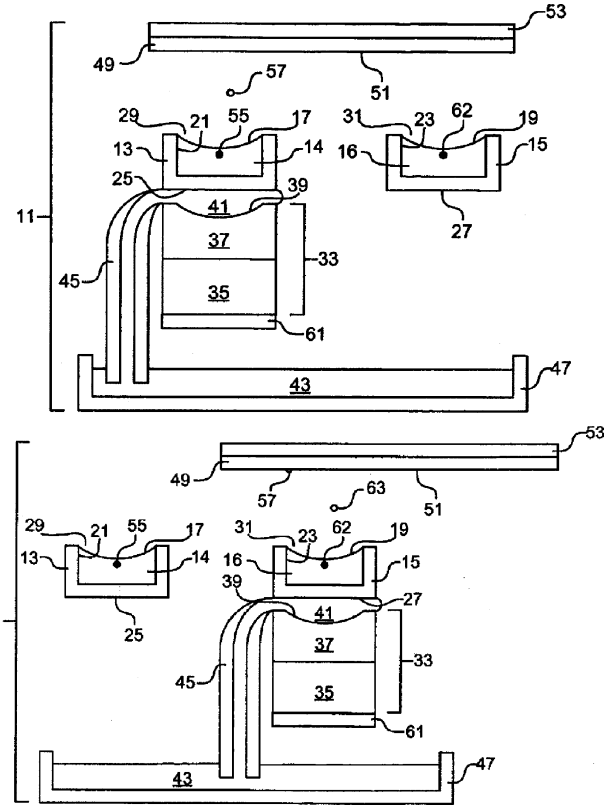
 0001-4966/2007/121(1)/12/19/$23.00

7,070,260

43.35.Zc DROPLET DISPENSATION FROM A RESERVOIR WITH REDUCTION IN UNCONTROLLED ELECTROSTATIC CHARGE

Mitchell W. Mutz *et al.*, assignors to Labcyte, Incorporated
4 July 2006 (Class 347/55); filed 9 January 2003

To dispense liquids **14** (or **16**) containing biological material onto the surface **51** of material **49** held by substrate **53**, transducer **37** sends ultrasound through coupling **41** to bowl **13** (or **15**) into the liquid to be focused at **55** (or **62**), causing droplet **57** to be ejected toward the target surface. Multiple fluids can be deposited this way by shuttling the receiver carriage, including target surface **49-53** and transducer carriage **45-41-61**, to bowls **13**, **15**, etc. For a continuous process, the electric field caused by the elec



trostatic charge built up on fluids **14** and **16** will alter the droplet trajectory and deposition site. Included in the 87 claims are many means to remove that charge, including a conductive liquid **14**, a conductive coupling medium **41**, ultraviolet (UV) radiation, a conducting contact or conductive fluid in intermittent contact with the bowl, means ''to add or remove electrons from the reservoir,'' means to ground the reservoir, etc. (there are several more such means claimed).—AJC

7,072,479

43.38.Ar CAPACITOR MICROPHONE

Shioto Okita, assignor to Kabushiki Kaisha Audio-Technica
4 July 2006 (Class 381/174); filed in Japan 11 January 2002

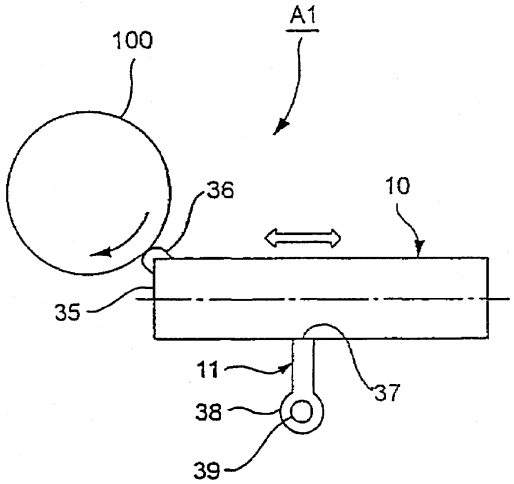
The patent describes a construction method that effectively reduces the contact area of the spacing ring that separates the diaphragm and electret backplate. This increases the effective area of the diaphragm, allowing a slight increase in microphone sensitivity.—JME

7,078,847

43.38.Ar PIEZOELECTRIC ACTUATOR, TIMEPIECE, AND PORTABLE DEVICE

Osamu Miyazawa *et al.*, assignors to Seiko Epson Corporation
18 July 2006 (Class 310/323.02); filed in Japan 21 December 1998

This patent discloses the details of a piezoelectrically driven, linear-to-rotary mechanical linkage that appears to be a watch escapement. The authors describe how the action of piezoelectric bar **10** can be set in longitudinal oscillation to cause projection **36** to rub intermittently against rotor **100**, and thereby spin it. The patent teaches how the selection of the mounting point **39** and the resonant frequencies of the bar are important for getting



the most efficient conversion of linear to rotary energy and gives several examples of how the arrangement of parts can be varied in order to achieve the same end. Although no design equations are given, one gets the feeling that this is not a field where mathematical analysis is useful outside of selecting resonant frequencies. The patent provides an interesting gallery of geometries for implementing a watch escapement or similar linear-to-rotary motion conversion.—JAH

7,084,552

43.38.Ar ANISOTROPIC ACOUSTIC IMPEDANCE MATCHING MATERIAL

Mahesh C. Bhardwaj, assignor to The Ultran Group, Incorporated
1 August 2006 (Class 310/327); filed 15 January 2004

This patent discloses the use of two materials of different acoustic impedances in a matrix for the purpose of matching the impedance of piezoelectric driver materials to the materials being driven. According to the patent, the layer must be one-quarter wavelength thick, and the material in the matrix must consist of fibers or rods oriented along the direction of motion of the transducer. The authors make no prescriptions for the ratio of impedances of said materials, or how to best arrange the fibers or columns—apparently all is well as soon as one mixes and aligns the material to make it anisotropic. It is not clear how the ideas discussed in this patent improve upon the developments of the past 30 years in 1–3 matrix composites.—JAH

7,084,554

43.38.Ar BIMORPH MEMS DEVICES

Baomin Xu and William S. Wong, assignors to Palo Alto Research Center, Incorporated
1 August 2006 (Class 310/332); filed 20 December 2004

This patent teaches a systematic approach to creating bimorph actuators for microelectronics mechanical systems (MEMS) applications using laser-assisted liftoff techniques. The goal seems to be to obtain smooth, thick films of piezoelectric transducers (PZT) on silicon and other substrates of interest to MEMS designers and to do it in a reproducible manner. The inventors have come up with an approach that allows lamination of PZT surfaces that might normally be unpredictably rough with sufficient mechanical integrity to make low-loss bimorph transducers. Although there is nothing fundamentally new about this approach, the organization and description of the steps involved may be of interest to others in the piezoelectric-MEMS field.—JAH

6,984,902

43.38.Ct HIGH EFFICIENCY VIBRATION ENERGY HARVESTER

Jiankang Huang *et al.*, assignors to Ferro Solutions, Incorporated
10 January 2006 (Class 310/26); filed 29 January 2004

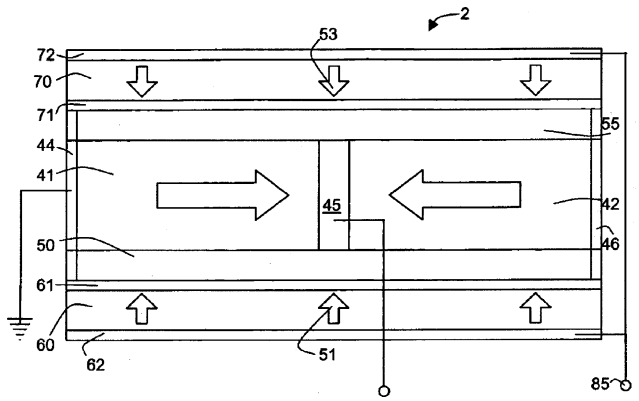
The device described in this patent would use magnetostrictive or other such materials to generate electrical power from vibrations, such as the vibration of an air duct. The electrical power could then be used locally, such as to power a sensor that would measure the temperature of the air in the duct. The patent considers in some detail the relative merits of alternate methods of generating power, such as piezoelectricity, magnetic coils, variable capacitance devices, etc. Various geometries are considered by which the magnetostrictive effects may be optimized.—DLR

7,075,217

43.38.Fx LAMINATED PIEZOELECTRIC TRANSFORMER

Alfredo Vazquez Carazo, assignor to Face International Corporation
11 July 2006 (Class 310/359); filed 9 April 2003

This patent discloses a novel configuration of piezoelectric materials for use as a voltage transformer. The authors review a large number of existing types, both linear (stacked) and transverse. They then proceed to show how their new design overcomes the limitations of the previous designs, mainly in the form of high-frequency behavior. The new design is



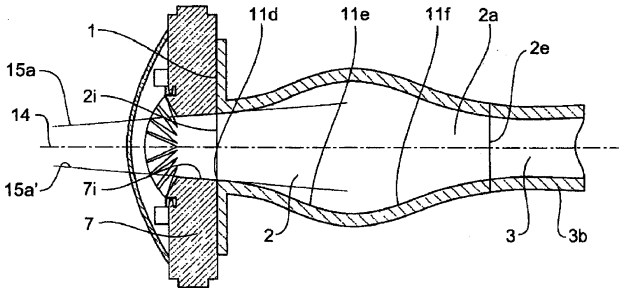
apparently like two Rosen-type transformers put back to back, so that the mechanical impedances to the left and right of the driving portion are equal, thus balancing the device response to first order and reducing the tendency of the device to bend. The patent is well written, brief, and has a concise review of the pros and cons of the existing designs. The technical depth is not very high, but sufficient to let the reader decide where the problems with these devices lie.—JAH

7,044,265

43.38.Ja CONSTANT DIRECTIVITY ACOUSTIC HORN

David John Murphy, assignor to Krix Loudspeakers Pty, Limited
16 May 2006 (Class 181/192); filed in Australia 17 September 2002

The transition from the exit **2i** of a compression driver **7** to the throat of a horn **3** is the subject of this patent, despite the title. The transition from the circular exit of the driver to the rectangular cross section **2e** of the horn



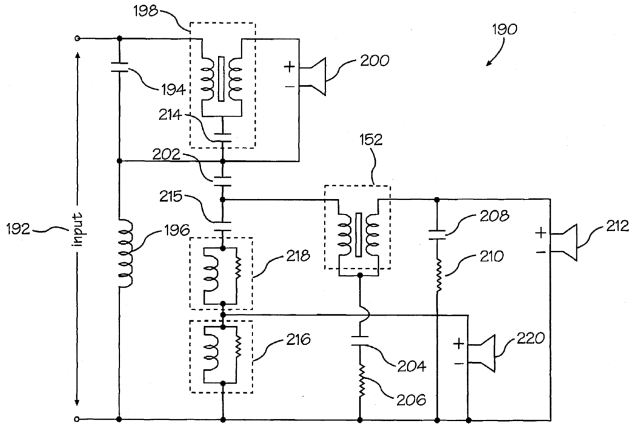
may have area discontinuities that can cause distortion. The patent presents a solution to this problem by varying the cross section of the transition so that the distance in one direction may increase while decreasing in the other.—NAS

7,085,389

43.38.Ja INFINITE SLOPE LOUDSPEAKER CROSSOVER FILTER

Richard T. Modafferi, assignor to Modafferi Acoustical Systems
1 August 2006 (Class 381/99); filed 30 September 2003

In an earlier United States Patent 4,771,466, the inventor argued the case for very steep filter slopes in multiway loudspeaker systems and disclosed practical methods for achieving initial slopes of 120 dB per octave or more in passive crossover networks. Coupled inductors helped realize the high-Q notches required. This patent describes several new improvements exemplified in the three-way crossover circuit shown. The network topology



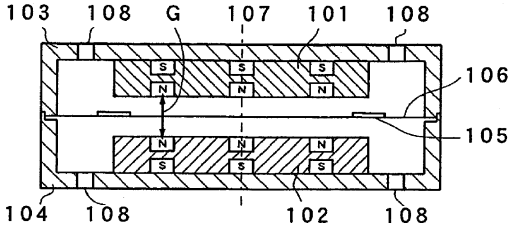
has been changed from parallel to series to smooth out the input impedance, and highpass slopes are now less steep than lowpass slopes to allow greater overlap in the crossover region. According to the patent, these changes have proved to be an audible asset while still maintaining lowered wave interference and reduced distortion. Well, right you are if you think you are.—GLA

7,085,394

43.38.Ja ELECTRO-ACOUSTIC TRANSDUCER AND ELECTRONIC DEVICE

Sawako Usuki and Shuji Saiki, assignors to Matsushita Electric Industrial Company, Limited
1 August 2006 (Class 381/412); filed in Japan 9 October 2001

A number of loudspeaker designs use a kind of printed-circuit voice coil on a planar diaphragm. The assembly is suspended between opposing magnets that produce a squished-out field more or less parallel to the dia-



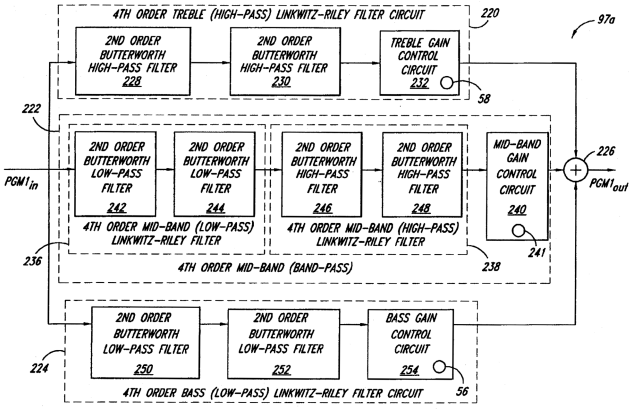
phragm. In this variant, magnets **101** and **102** take the form of circular disks and voice coil **105** is a flat spiral. Sound exits through multiple vents **108**. The device is intended for use in cellular phones, laptop computers, and the like.—GLA

7,043,032

43.38.Lc TONE-CONTROL CIRCUIT AND METHOD FOR CONDITIONING RESPECTIVE FREQUENCY BANDS OF AN AUDIO SIGNAL

Philip R. Jeffs and Dennis A. Bohn, assignors to Rane Corporation
9 May 2006 (Class 381/101); filed 15 June 1999

The title "turntablist" refers to someone who in the past may have been called a disc jockey (DJ). The patent also lets us know that a turntablist may use a control called a "hamster." With this introduction, the patent describes fourth-order tone control circuits for DJ mixing equipment that



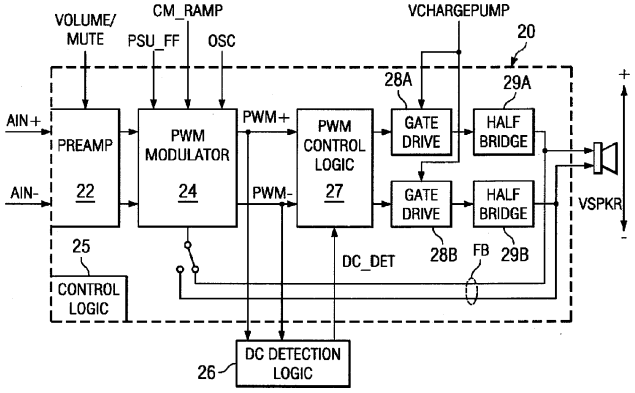
may in fact use cascaded second-order filters. The patent provides a cogent description of DJ mixing and processing equipment as well as the operation of the same. The assignee is a major manufacturer of DJ mixing and processing equipment.—NAS

7,078,964

43.38.Lc DETECTION OF DC OUTPUT LEVELS FROM A CLASS D AMPLIFIER

Lars Risbo and James Teng, assignors to Texas Instruments, Incorporated
18 July 2006 (Class 330/10); filed 12 October 2004

This is a well-written patent that includes useful background information about the operation of Class AD audio power amplifiers. The invention is concerned with the problem of unwanted dc offset at the output terminals. A simple highpass filter will either attenuate some low-frequency audio signals or fail to block slowly fluctuating offset. An improved dc detector is



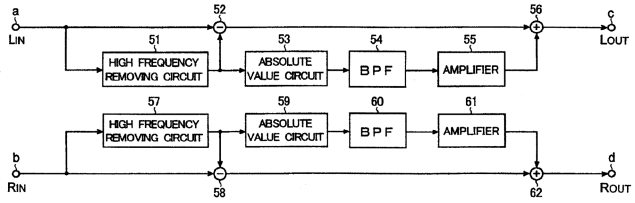
described which operates as a sigma-delta modulator in combination with a digital lowpass sine filter operating at a much slower clock rate. "In this manner, the cut-off frequency of the sine filter can be extremely low, so that dc output signal components can be detected even in the presence of low-frequency components in the audio output during detection."—GLA

7,079,660

43.38.Lc BASS COMPENSATION DEVICE AND A SOUND SYSTEM USING THE DEVICE

Shigeji Ohama and Hiroyuki Saito, assignors to Rohm Company, Limited
18 July 2006 (Class 381/98); filed in Japan 16 April 2001

Practical applications of synthetic bass are older than the art of audio reproduction and can be found in many pipe organs. Most implementations, whether acoustical or electronic, attempt to fool the ear into hearing a nonexistent fundamental tone. The miniaturization of audio playback devices has prompted the development of more recent electronic variants. Some of these filter out unwanted low-frequency components and then perform fre-



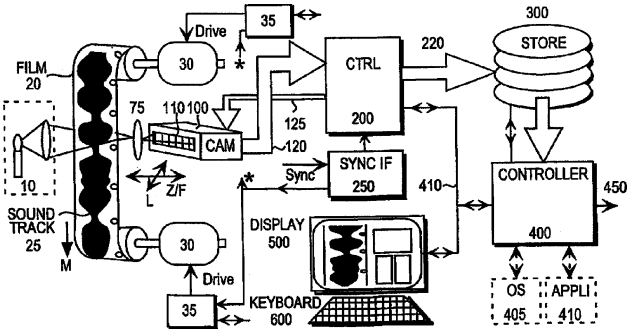
quency doubling—a process that requires a second lowpass filter. However, very-low-frequency signals may be outside the safe frequency range, even when doubled. What to do? Replacing the lowpass filter with a bandpass filter **54**, **60** solves the problem, and the patent claims cover three possible variants of this idea.—GLA

7,053,978

43.38.Md CORRECTION OF OPTICAL FILM SOUNDTRACK DEFICIENCIES

Jaime Arturo Valenzuela and Vincent Richard Williams, assignors to Technicolor, Incorporated
30 May 2006 (Class 352/27); filed 30 August 2002

Optical sound tracks on motion picture film strips have been around for more than 75 years. The patent provides a capsule description of analog



optical soundtrack recording and playback, the problems inherent with the process, and a means for improving the playback of optical soundtracks. The description of the invention is clearly written.—NAS

7,075,390

43.38.Rh SURFACE ACOUSTIC WAVE DEVICE

Akihiro Bungo *et al.*, assignors to Mitsubishi Materials Corporation
11 July 2006 (Class 333/193); filed in Japan 11 November 1999

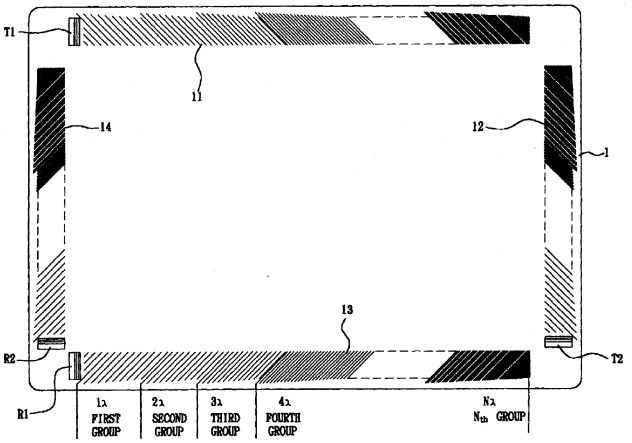
A rf bandpass filter is claimed that has minimal attenuation in its passband using a langasite (Lanthanum Gallium Silicate) piezoelectric element cut with specified Euler angles and with specified aluminum electrode dimensions.—AJC

7,075,525

43.38.Rh METHOD FOR DESIGNING REFLECTION STRIPES OF AN ACOUSTIC TOUCH SCREEN

Eric Cheng *et al.*, assignors to Onetouch Technologies Company, Limited
11 July 2006 (Class 345/177); filed in Taiwan 11 March 2003

A touch screen using surface acoustic waves (SAW), where more constant scattered SAW signal strength leads to improved position resolution, is



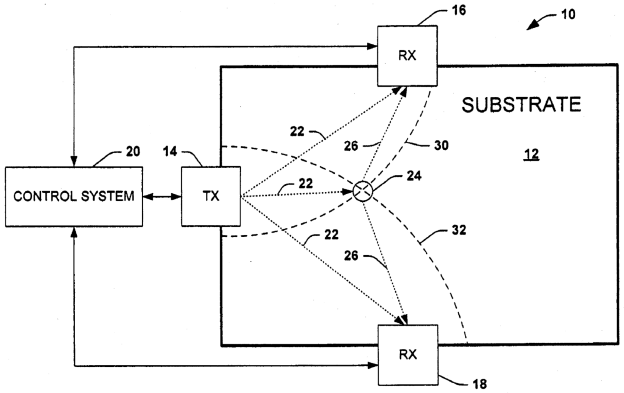
achieved by reducing the signal degradation for touch positions furthest (e.g., ***N*th group**) from the transducers **T1** and **R1** and from **T2** and **R2**. This is achieved by tapering the length of the surface wave reflectors **11**, **12**, **13**, **14** in every group.—AJC

7,079,118

43.38.Rh TOUCH SCREEN USING ECHO-LOCATION

David J. Benard *et al.*, assignors to Rockwell Automation Technologies, Incorporated
18 July 2006 (Class 345/173); filed 23 February 2004

Touch screen system **10** is claimed where the touch sensitive surface covering substrate **12** comprises a glass micro sheet 30–300 μm thick separated by a fine gap created by residual dust particles or by applying a powder before assembly. Local pressure at **24** by a finger or a stylus will locally deform the microsheet to contact substrate **12**. The propagation of surface acoustic waves **22** emitted by ultrasound transmitter **14** into the



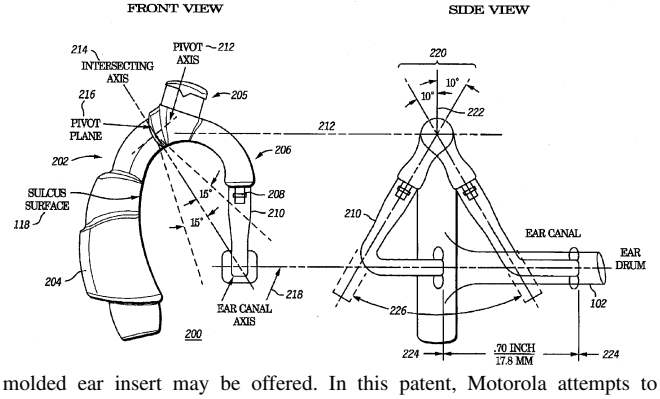
surface of the substrate adjacent to the microsheet is perturbed at contact **24**. This in turn causes scattering of the surface wave **22** arriving there. Signals caused by scattered surface waves **26** arriving at receiving transducers **16** and **18** are processed by controller **20** to deduce the position of contact **24** according to the time of flight to **16** and **18** as well as the amplitude of the envelope of the ultrasound SAW pulses.—AJC

7,082,207

43.38.Si ADJUSTABLE BEHIND-THE-EAR COMMUNICATION DEVICE

Gary M. Rapps, assignor to Motorola, Incorporated
25 July 2006 (Class 381/381); filed 14 December 2001

A behind-the-ear communications set combines transducers and electronics in a single package. To provide proper operation yet fit comfortably over long periods of time, a choice of several ear inserts or even a custom-



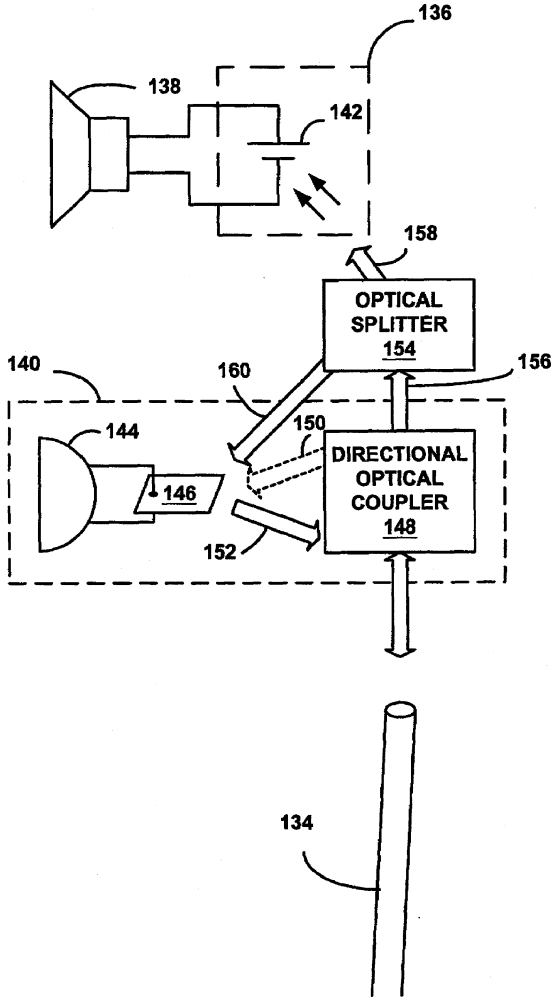
molded ear insert may be offered. In this patent, Motorola attempts to achieve an acceptable all-in-one approach by adding an adjustable pivot to the sound delivery tube.—GLA

7,072,475

43.38.Si OPTICALLY COUPLED HEADSET AND MICROPHONE

Frank A. DeNap *et al.*, assignors to Sprint Spectrum L.P.
4 July 2006 (Class 381/74); filed 27 June 2001

A combination microphone-receiver headset can take many forms, but the great majority of designs rely on electrical wiring to connect the headset with associated electronic equipment. This patent points out that the connecting cable can pick up unwanted electrical interference and, in some cases, radiate harmful interference. An alternative optical link involving a splitter, a directional optical coupler, etc., is described at some length.—GLA

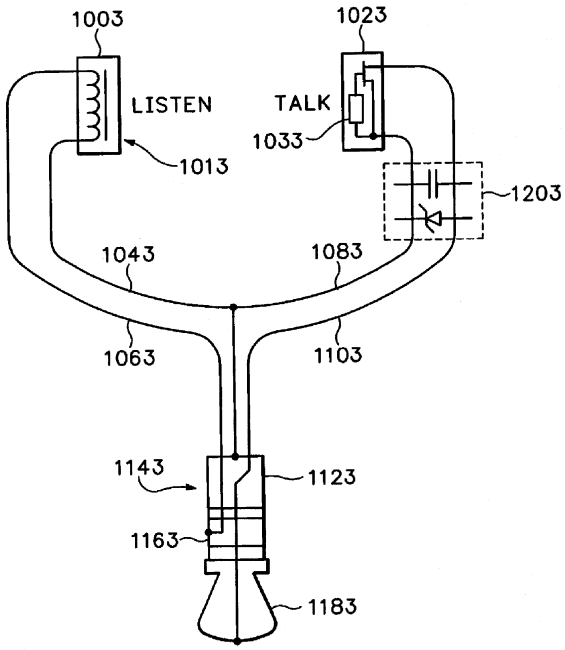


7,072,476

43.38.Si AUDIO HEADSET

Donald R. White and Masahisa Masuda, assignors to Matech, Incorporated
4 July 2006 (Class 381/74); filed 13 January 2003

This is a good example of a never-ending patent. The original application was filed in 1997 and has appeared as United States Patents 5,907,538 and 6,370,245, along with numerous continuation-in-parts and provisional applications. The basic feature common to all of these is the use of a microphone in the ear to pick up speech. According to the patent, such a microphone has the advantage of being "isolated from background noise and vibrations due to bone conduction." Among other embodiments, this latest patent envisions a dual-mode transducer that can function as a microphone for two-way communication and as a receiver for stereo playback.—GLA

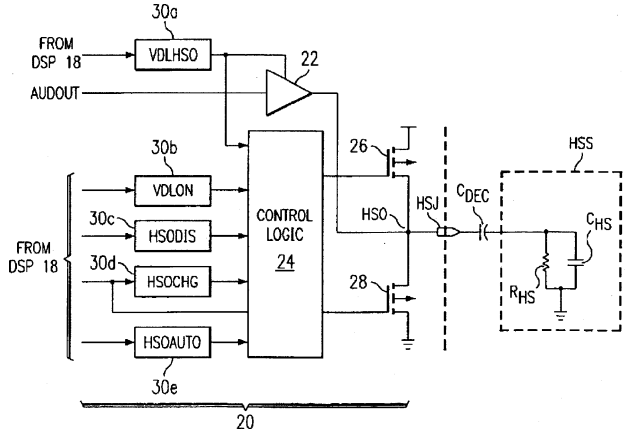


7,084,701

43.38.Si PREVENTING POWER-ON AUDIO OUTPUT NOISE IN A WIRELESS TELEPHONE HANDSET

Damien Mendoza *et al.*, assignors to Texas Instrument, Incorporated
1 August 2006 (Class 330/51); filed in 10 May 2002

Solid-state audio amplifiers, especially single-ended amplifiers, have a tendency to produce a loud pop when turned on. The patent at hand uses several paragraphs to discuss this phenomenon and explain why all previous pop suppression methods are somehow lacking. Three more pages of text and 33 patent claims then describe improved circuitry employing muting transistors **26** and **28** whose gates are driven by suitable control logic.—GLA

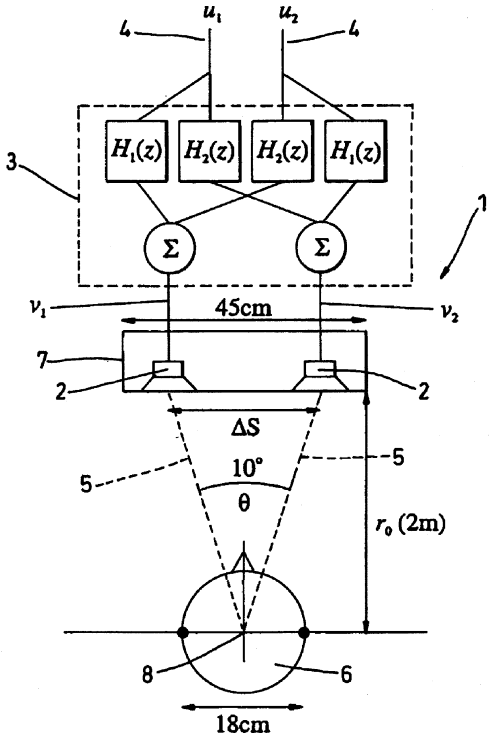


7,072,474

43.38.Vk SOUND RECORDING AND REPRODUCTION SYSTEMS

Philip Arthur Nelson *et al.*, assignors to Adaptive Audio, Limited
4 July 2006 (Class 381/17); filed in United Kingdom 16 February 1996

The history of this patent goes back more than ten years to an original filing in Great Britain in 1996, and includes earlier United States Patent 6,760,447 [reviewed in J. Acoust. Soc. Am. **116**(6), 3253 (2004)]. The inventors' "stereo dipole" is one of several patented schemes to provide full surround-sound reproduction from a pair of closely spaced loudspeakers. However, the associated crosstalk cancellation circuitry differs from other methods. Rather than a fixed delay, the phase difference between the two



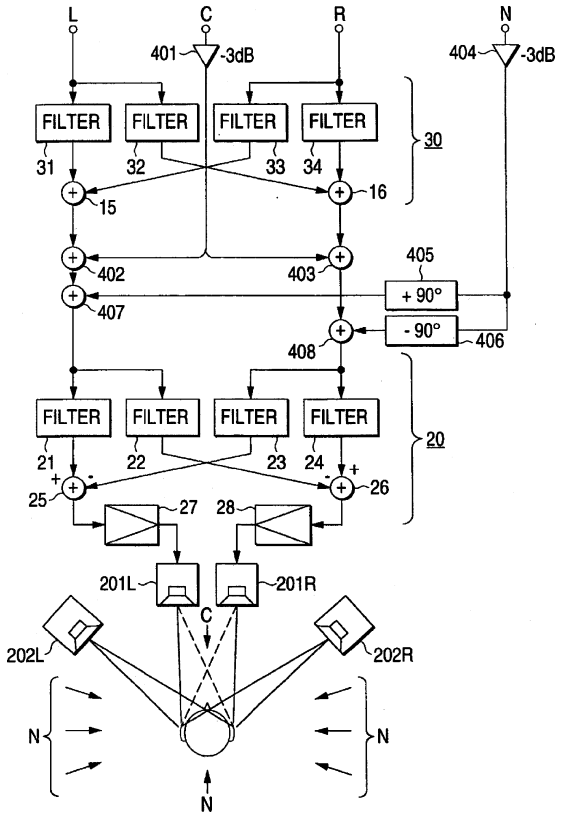
speakers "varies with frequency from low frequencies where the vibrations are substantially out of phase to high frequencies where the vibrations are in phase." (Some of the required processing can be included in the recording itself.) The patent includes a thorough discussion of the requirements for such virtual stereo synthesis, along with a great many graphs and diagrams. However, a number of considerations, such as directional patterns and the possible influence of early room reflections, are largely ignored.—GLA

7,076,068

43.38.Vk THREE-DIMENSIONAL SOUND REPRODUCING APPARATUS AND A THREE-DIMENSIONAL SOUND REPRODUCTION METHOD

Hiromi Sotome *et al.*, assignors to Yamaha Corporation
11 July 2006 (Class 381/17); filed in Japan 21 June 1996

This patent covers a great deal of information that has been discussed in these pages over the past few years regarding the playback of encoded surround sound over a single pair of frontal loudspeakers. The spreading of front-left and front-right program outside the loudspeaker base is, by now, ancient art. This patent proposes an additional "nonlocalization" signal that is sensed as arriving generally from the sides and rear of the listener. While such effects can be generated in a carefully controlled listening environ-



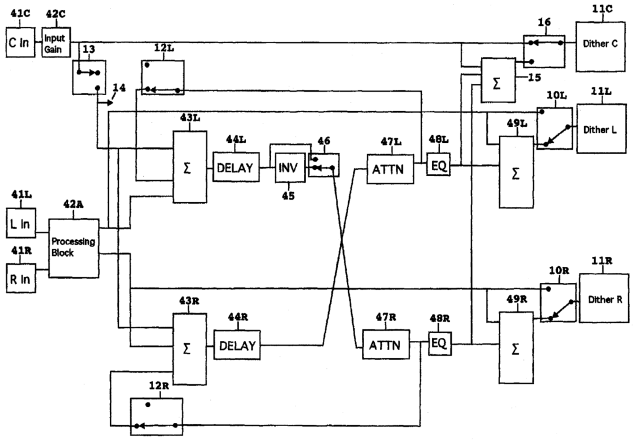
ment, they are normally sensitive to even slight head movements. It is doubtful if all of this works as neatly as suggested.—JME

7,076,071

43.38.Vk PROCESS FOR ENHANCING THE EXISTING AMBIENCE, IMAGING, DEPTH, CLARITY, AND SPACIOUSNESS OF SOUND RECORDINGS

Robert A. Katz, Longwood, Florida
11 July 2006 (Class 381/307); filed 8 June 2001

The patent describes a fairly complex procedure in which stereo signals are delayed and adjusted in level in a repeating sequence as follows: "The process includes generating a first audio signal and generating a second audio signal; delaying and attenuating said second audio signal to form a third audio signal; summing said third audio signal with said first audio to



form a fourth audio signal, etc." Upwards of twenty-five iterations may be necessary in this chain. For each signal generation step, the delay and attenuation values are carefully determined to lie within audible fusion limits defined by Haas [Blauert, *Spatial Hearing* (1999), Sec. 3.1.2] in order to

achieve sequential masking of the delays as such. The result of this is a subjective increase in perceived ambience with little, if any, spectral alteration in the overall signal. Additionally, the process ''unmasks 60–90 milliseconds or more of the early reverberation in the sound recording, thus enhancing the character of the sound recording which comes from the recording hall.''—JME

7,082,371

43.40.At FUNDAMENTAL MISTUNING MODEL FOR DETERMINING SYSTEM PROPERTIES AND PREDICTING VIBRATORY RESPONSE OF BLADED DISKS

Jerry H. Griffin and Drew M. Feiner, assignors to Carnegie Mellon University
25 July 2006 (Class 702/56); filed 30 April 2004

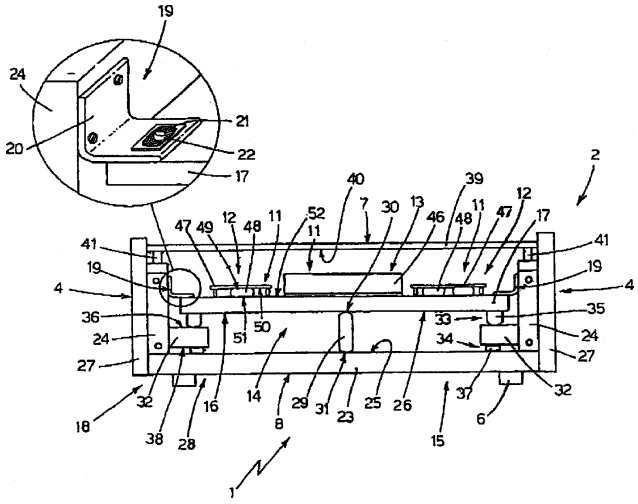
Mistuning here refers to the natural frequencies of individual blades of a bladed disk (as in turbines, compressors, fans) deviating from the ideal situation where all blades have the same natural frequency. Such mistuning may lead to some blades experiencing increased vibrations and failing early due to high-cycle fatigue. The present patent describes several measurement and computational approaches toward determining mistuning, including use of a computationally efficient ''fundamental mistuning model.''—EEU

7,043,029

43.40.Tm DEVICE FOR USE IN AN AUDIO SIGNAL PROCESSING SYSTEM

Sergio Luca Zini and Ruggero Ruggeri, both of Bologna, Italy
9 May 2006 (Class 381/71.7); filed in Italy 22 December 1998

Vibrations of cables, circuit boards, and other components in an audio system are, according to this patent, deleterious to the final quality of the original musical message. The patent describes a scheme wherein materials, such as ''harmonic essence'' woods such as cypress, fir, and beech, and ''hard essence'' woods such as ebony and maple, can be used to construct a transmission circuit **14** for ''controlled distribution and for discharge of



undesired vibrations'' in an audio processing circuit, which appears to be a CD or DVD player, although it may be any electronic or mechano-electronic device. Although the patent draws on methods more appropriate for luthiers and similar crafts, how these techniques can be useful in mass market consumer electronics items is not clear to this reviewer. A method is also disclosed for controlled distribution of undesired vibration in a signal cable. To which one can say ''metaphysics.''—NAS

7,080,564

43.40.Tm VIBRATORY TRANSDUCER

Alfred Rieder and Wolfgang Drahm, assignors to Endress and Hauser Flowtec AG
25 July 2006 (Class 73/861.357); filed in Germany 8 May 2002

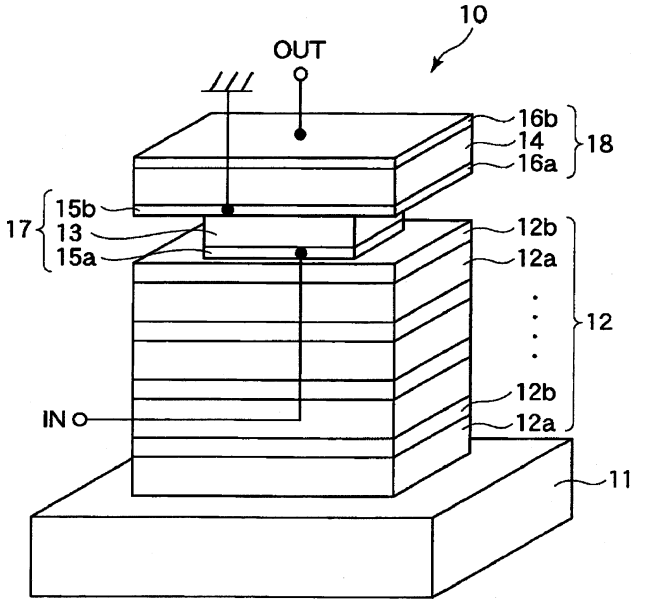
This patent describes a vibratory viscosimeter in which a length of straight pipe is vibrated in torsion about its longitudinal axis at its resonant frequency and the viscosity of the fluid flowing in the pipe is deduced from the energy required to keep the pipe vibrating at a given amplitude. In order to increase the accuracy of the device, it is desired to reduce the energy transmitted from the viscosimeter pipe to the pipes to which it is connected. This is accomplished by attaching torsional vibration absorbers, tuned to the aforementioned resonant frequency, to the ends of the viscosimeter pipe.—EEU

7,075,214

43.40.Sk PIEZOELECTRIC RESONATOR AND ELECTRONIC COMPONENT PROVIDED THEREWITH

Kenji Inoue *et al.*, assignors to TDK Corporation
11 July 2006 (Class 310/328); filed in Japan 10 March 2004

This patent concerns the construction of bandpass filters comprising film bulk acoustic resonators (FBARs), and the stagger tuning of such an arrangement to make a wider bandpass response. The authors propose that one can create a series connection of such filters by simply stacking them on top of each other. The weak coupling required for stagger tuning is obtained by changing the cross section of the resonators, and the stagger tuning itself



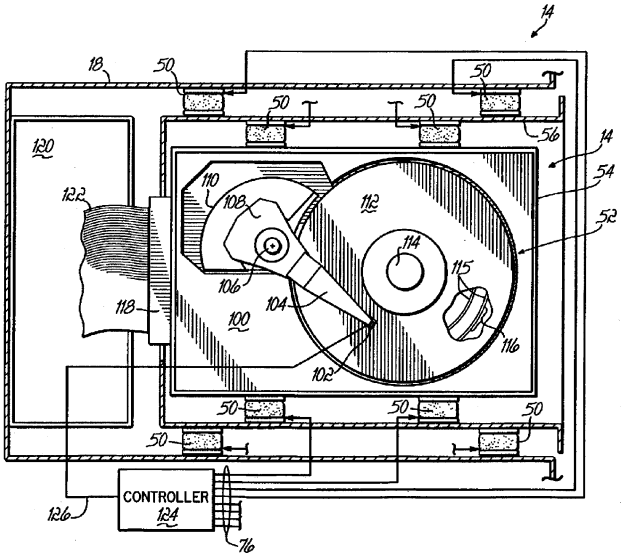
is obtained by changing the thickness of successive layers slightly. This is a standard technique adapted to microelectronics mechanical systems (MEMS) fabrication methods in a not particularly elegant fashion. There are a few examples given of materials and dimensions necessary to get this working in the 2 to 3 GHz frequency range.—JAH

7,072,274

43.40.Vn ADAPTIVE MOUNTING SYSTEM FOR DATA STORAGE DEVICES

Gordon James Smith and Sri M. Sri-Jayantha, assignors to Lenovo (Singapore) Pte, Limited
4 July 2006 (Class 369/247.1); filed 31 May 2002

In data storage device **14**, adaptive vibration mounts **50** coupled to structural member **56** and the enclosure **18** are configured to damp the transfer of vibration to the data transfer mechanism **100–108**. The active element **50** responds to electric stimulus. Its materials may include an electrostatically stricted polymer (ESSP) to produce countering forces, a magnetorheological (MR) material, and an electrorheological (ER) material, the



latter two capable of producing damping. Controller **124** monitors the position error signal (PES) **126** from read/write head **102** and develops signals **76** to drive actuators **50** to produce optimal electromotive forces that optimize the stiffness and damping parameters of those adaptive mounts, thereby creating a vibration-free state for data-handling module **53**, **54**, **100–116**.—AJC

7,073,384

43.40.Yq METHOD AND APPARATUS FOR REMOTE MEASUREMENT OF VIBRATION AND PROPERTIES OF OBJECTS

Dimitri Donskoy *et al.*, assignors to Stevens Institute of Technology
11 July 2006 (Class 73/657); filed 23 August 2000

Unlike a laser-doppler vibrometer, the device described in this patent employs amplitude- or phase-modulated, incoherent electromagnetic radiation (light or microwaves) emitted toward a vibrating object. The reflected and/or scattered signal, which is modulated by the vibration, is received by a remote sensor and demodulated to extract and analyze the vibration waveform. This device does not require the expensive precision optical elements of a laser vibrometer and is less sensitive to vibration of the transmitting and receiving systems. In order to enhance performance of the vibrometer (which, as in laser vibrometers, depends on relative motion between the measurement system and the observed object), accelerometers are mounted on the transmitters and receivers and the measured vibrations of these elements are taken into account via an innovative algorithm. Since microwave radiation can penetrate into objects, use of this radiation permits measurement of vibrations of internal components.—EEU

7,084,553

43.40.Yq VIBRATING DEBRIS REMOVER

Damian R. Ludwiczak, Medina, Ohio
1 August 2006 (Class 310/328); filed 24 September 2004

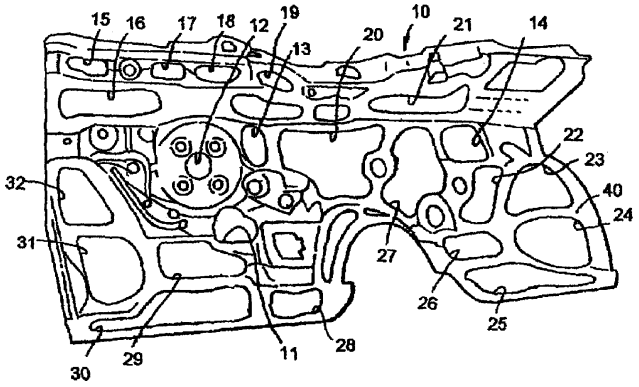
In order to remove debris from a structure, such as ice from the window of a vehicle, a device consisting of a vibration generator and a coupler is attached to the edge of the structure. The coupler is intended to enhance the transmitted vibrations by providing impedance matching.—EEU

7,070,848

43.50.Gf VEHICLE ACOUSTIC BARRIER

Michael T. Campbell, assignor to Cascade Engineering, Incorporated
4 July 2006 (Class 428/137); filed 21 October 2002

A 30% reduction in the weight of vehicle-firewall acoustical barrier-absorber mat **10** is obtained by spacing that mat from the sheet-metal body panel with an absorber mat, by using a dense material **40** for the barrier, and



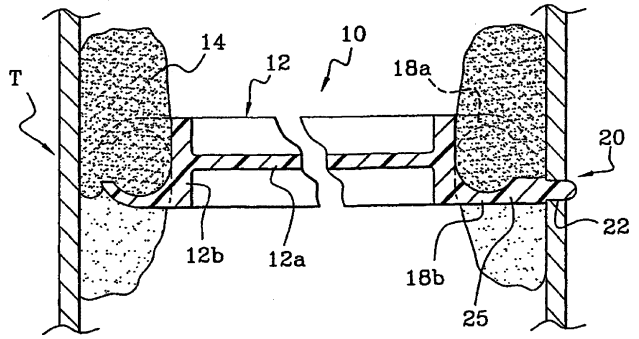
then cutting holes **13–32** in that barrier where the sound level is not great. (The steering-wheel column passes through at **12**). The noise reduction values for sound penetrating the panel and mat are essentially the same as for the original heavier mat.—AJC

7,077,461

43.50.Gf SOUND INSULATING DEVICE DESIGNED TO BE MOUNTED IN A TUBULAR PART, IN PARTICULAR A MOTOR VEHICLE BODY PART

Florence Ratet, assignor to Hutchinson
18 July 2006 (Class 296/187.02); filed in France 2 July 2001

To reduce the labor required to install sound barrier **10** inside tube **T** that enters a vehicle interior, a process is claimed where a cord of thermally expandable material is wrapped on the periphery of plug **10** whose body **12**



includes sound barrier **12a**, cord retaining fingers **18a** and **18b**, and positioning tab **25-20** that, in turn, seats in hole **22**. Application of heat during

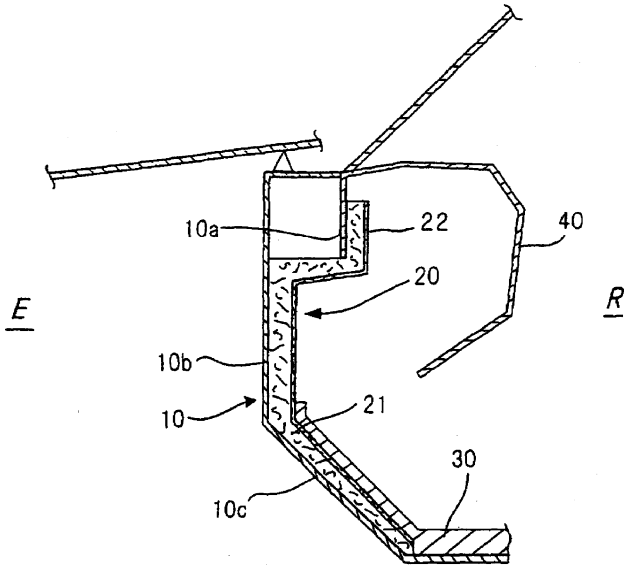
manufacturing causes the cord material **14** to expand to seal the gap between **T** and **10**. The thermo-expansion cord material may be a polyamide.—AJC

7,080,712

43.50.Gf SOUNDPROOF MATERIAL FOR VEHICLE AND METHOD OF MANUFACTURING THE MATERIAL

Tomohiro Tsuiki *et al.*, assignors to Kasai Kogyo Company, Limited
25 July 2006 (Class 181/204); filed in Japan 23 January 2001

Sound transmission loss (TL) improvement is claimed whereby sound absorbing material **21** is placed on the interior side surface of the vehicle-body panel **10** under dash board **40**, while incorporating enhanced sound absorption due to the focusing property of the concave dashboard backside.



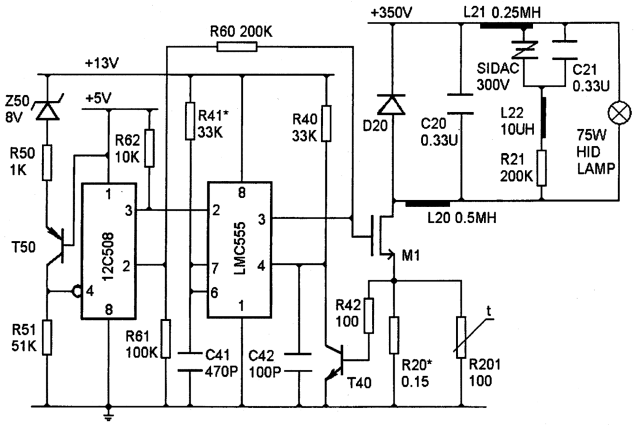
The authors barely distinguish between TL and sound absorption, though they do represent the TL improvement above 1 kHz when skin **22** is added. Several material arrangements **20-21-22** and the methods of assembling them and fastening them by gluing are claimed.—AJC

7,084,578

43.50.Gf SYSTEM AND METHOD FOR PREVENTING ACOUSTIC ARC RESONANCE IN A HID LAMP

Mingfu Gong, Cupertino, California
1 August 2006 (Class 315/194); filed 10 May 2004

High intensity discharge (HID) lamps suffer acoustic arc resonance (AAR) when starting and when operating, which produces audible noise. The frequency, amplitude, and the conditions for which AAR occurs will vary with ambient temperature and with lamp age. Magnetic HID ballasts are not susceptible to AAR, but are inefficient. The frequency at which efficient HID electronic ballasts feed pulses of current to the HID lamp for starting and for its running, from 1 kHz to over 100 kHz, includes the AAR frequency range. Claims include the following facts: Pulse deduction (omitting one or more pulses) provides over-current protection during starting and when an over current is detected. A bi-directional thyrister (SIDAC) is used as a starter switch. AAR is much less likely when the pulse frequency and pulse width are varied randomly. A pseudo random sequence is claimed that optimally avoids AAR. Pulse phase modulation (PPM) can also avoid AAR. Not more than ten contiguous pulses are allowed with the same phase. The frequency of the current pulses may be from a few kHz to a few hundred



kHz. To avoid low-frequency AAR, the period of any constant phase pulse sequence should be not any longer than 5 ms.—AJC

7,076,922

43.55.Ti INTERCONNECTING SOUND ATTENUATING ENCLOSURE

Jacob L. Parres, Camarillo, California
18 July 2006 (Class 52/79.5); filed 7 October 2003

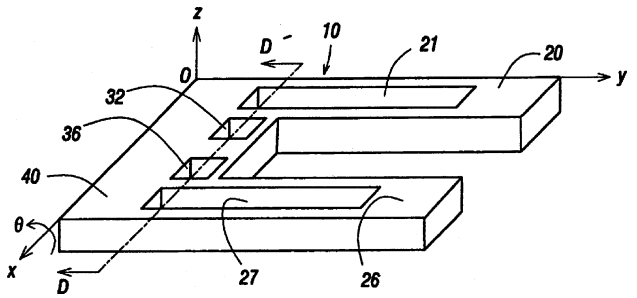
This patent describes a kit of parts, including pipe fittings and panels, that assembles into a self-contained sound isolation container for use as a music practice room or the like. It is demountable and transportable. The structural frame is formed with pipes and slip-in fittings.—CJR

7,071,794

43.58.Hp QUARTZ CRYSTAL RESONATOR, UNIT HAVING RESONATOR, OSCILLATOR HAVING UNIT, ELECTRONIC APPARATUS HAVING OSCILLATOR, AND METHOD FOR MANUFACTURING ELECTRONIC APPARATUS

Hirofumi Kawashima, assignor to Piedek Technical Laboratory
4 July 2006 (Class 333/187); filed in Japan 6 March 2002

This patent discloses the use of grooved tuning fork tines to increase the coupling constant for tuning fork resonators as used in high-precision timing devices. The authors argue that the use of grooved areas **21** and **27** of



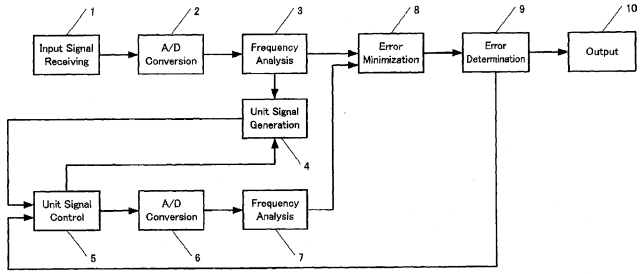
the tuning fork tines will make the oscillator more stable due to the suppression of the second harmonic when the grooves are made so as to satisfy certain dimensional constraints. Since this is relatively easy to fabricate, this is a useful incremental improvement for the art.—JAH

7,076,433

43.60.Bf APPARATUS AND PROGRAM FOR SEPARATING A DESIRED SOUND FROM A MIXED INPUT SOUND

Masashi Ito and Hiroshi Tsujino, assignors to Honda Giken Kogyo Kabushiki Kaisha
11 July 2006 (Class 704/500); filed in Japan 24 January 2001

One of the long-standing goals in general signal analysis has been the reliable separation of one signal component from an ensemble of similar components. The patent describes a number of techniques, one of which is shown in the figure. After a given segment of the signal has been windowed and analyzed in the frequency domain **3** to produce its power spectrum, it is



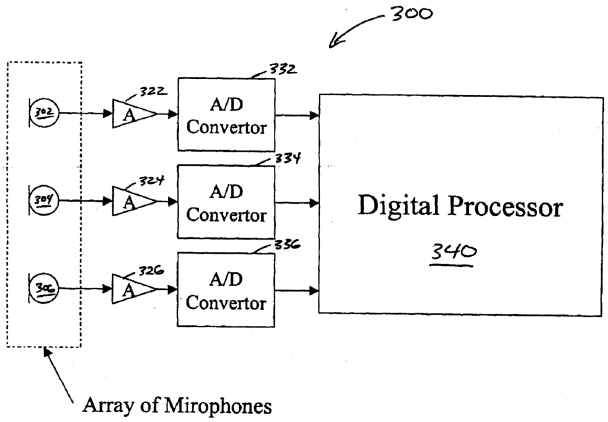
then fed to a unit signal generation block **4**. In a recursive procedure, successive peaks in the power spectrum are analyzed and tracked, ostensibly leading to an estimate of what may actually constitute the desired signal. Such elements as ''harmonic structure, pitch consistency, intonation, onset/offset information, and/or sound direction'' may be used in this process.—JME

7,084,801

43.60.Cg APPARATUS AND METHOD FOR ESTIMATING THE DIRECTION OF ARRIVAL OF A SOURCE SIGNAL USING A MICROPHONE ARRAY

Radu Victor Balan and Justinian Rosca, assignors to Siemens Corporate Research, Incorporated
1 August 2006 (Class 341/155); filed 5 June 2003

The patent describes in broad terms an adaptive microphone array for



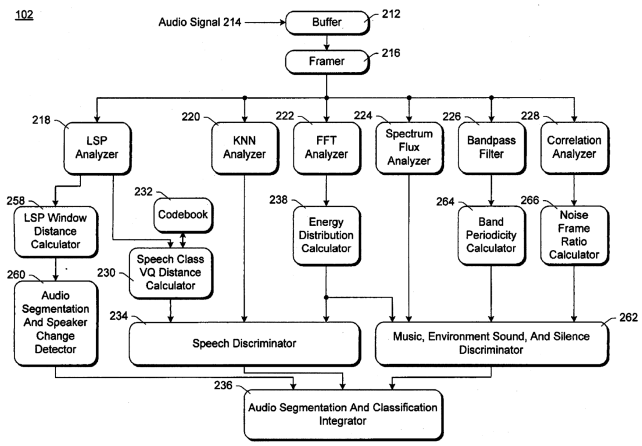
determining the direction of a sound source. Applications include optimized hands-free pickup of speech in noisy environments.—JME

7,080,008

43.60.Bf AUDIO SEGMENTATION AND CLASSIFICATION USING THRESHOLD VALUES

Hao Jiang and Hong-Jiang Zhang, assignors to Microsoft Corporation
18 July 2006 (Class 704/214); filed 11 May 2004

In computer-based audio information retrieval, it is useful to be able to identify the nature of the audio signal; such determinations as ''silence, environmental sound, speech, nonspeech, or music'' can be useful in parsing a large collection of audio elements. The patent describes a statistical



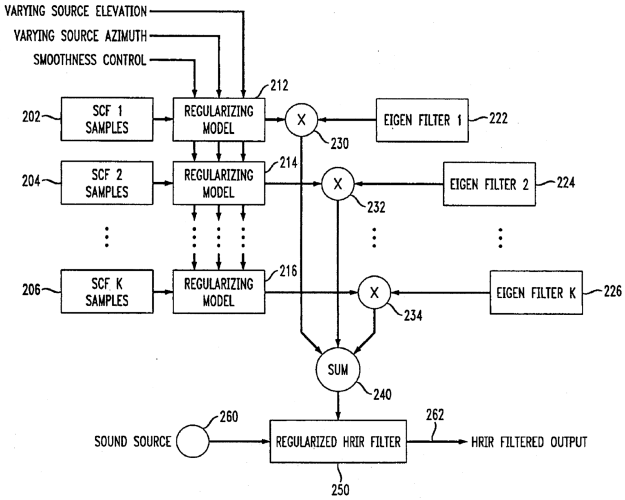
method for doing this based on the following elements: ''Line-spectrum pairs, noise frame ratio, periodicity of particular bands, spectrum flux features, and energy distribution in one or more bands.'' The figure shows a signal flow diagram of how this can be done.—JME

7,085,393

43.60.Dh METHOD AND APPARATUS FOR REGULARIZING MEASURED HRTF FOR SMOOTH 3D DIGITAL AUDIO

Jiashu Chen, assignor to Agere Systems, Incorporated
1 August 2006 (Class 381/310); filed 13 November 1998

HRTFs axe implemented in a variety of applications where simulated sounds can be panned in various directions about the listener. head-related transfer function (HRTF) data is normally collected at discrete bearing



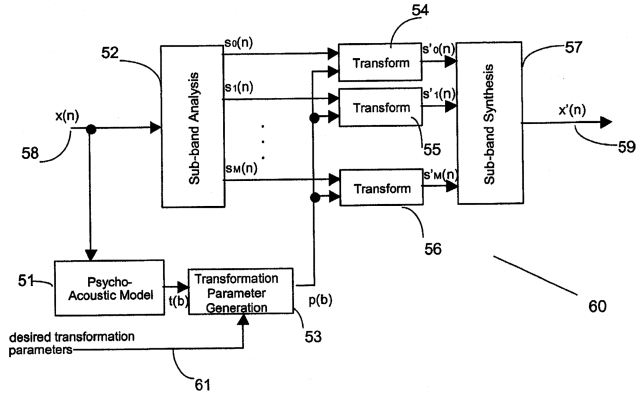
angles about the listener and smooth panning is difficult to achieve when moving from one position to another. The patent discusses procedures for smoothly interpolating between adjacent positions without producing artifacts.—JME

7,050,965

43.60.Ek PERCEPTUAL NORMALIZATION OF DIGITAL AUDIO SIGNALS

Alex A. Lopez-Estrada, assignor to Intel Corporation
23 May 2006 (Class 704/200.1); filed 3 June 2002

Critical bands and loudness masking have been used to code audio signals for quite awhile. The same technique can be used to "normalize" audio signals when converting digital signals from one sample rate to another to minimize artifacts. A method is described that uses critical bands, the absolute threshold of hearing, a psycho-acoustic model, and wavelet analysis to minimize high computational complexity. This patent could also be classified under PACS 43.60.Hj.—NAS

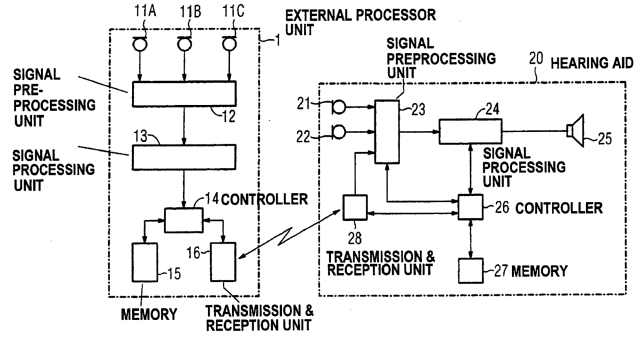


7,072,480

43.66.Ts HEARING AID SYSTEM WITH A HEARING AID AND AN EXTERNAL PROCESSOR UNIT

Uwe Rass, assignor to Siemens Audiologische Technik GmbH
4 July 2006 (Class 381/314); filed in Germany 24 June 2002

The signal processing on the output signal of a microphone in an external processor unit is adapted to a specific hearing environment, e.g., inside an automobile. The processed signal is sent wirelessly to the hearing aid. The external processor unit can remain in a particular hearing environment permanently and may be configured with separate signal processors, each of which is dedicated to a specific listening environment.—DAP

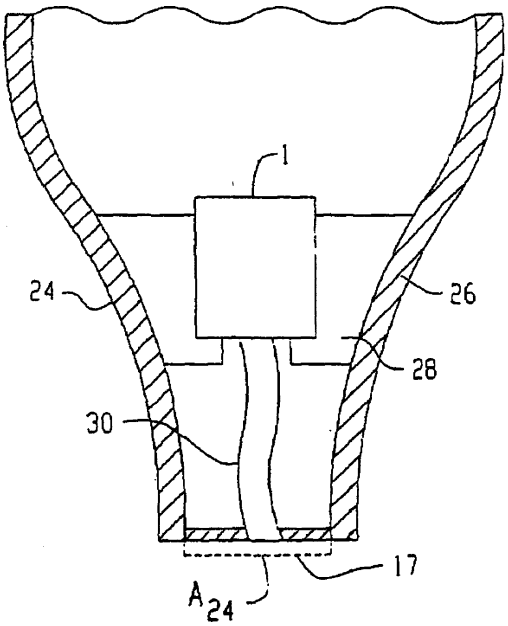


7,074,296

43.66.Ts IN-EAR HEARING AID AND METHOD FOR ITS MANUFACTURE

Andi Vonlanthen, assignor to Phonak AG
11 July 2006 (Class 156/293); filed 2 August 2004

An acoustically transparent elastomeric membrane is placed on the output aperture of a custom hearing aid to prevent earwax from entering and occluding the hearing aid. A support plate with a blind hole has the membrane as a base and is bonded to the hearing aid housing.—DAP

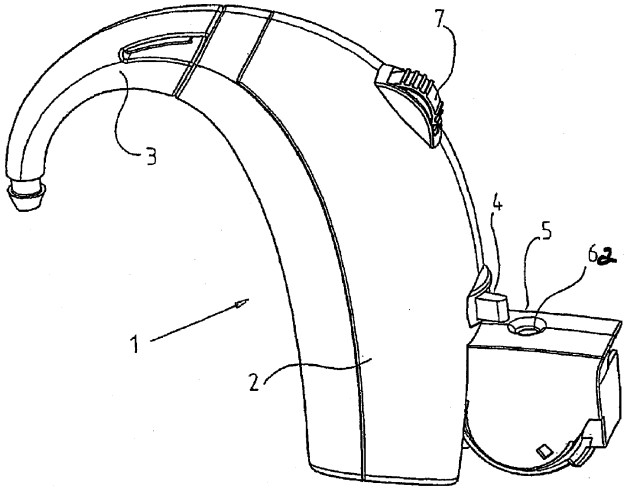


7,076,075

43.66.Ts HEARING AID WITH PIVOTABLE BATTERY DRAWER HAVING OPENING IN END WALL FOR BATTERY REMOVAL

Michael F. Jørgensen, assignor to Oticon A/S
11 July 2006 (Class 381/323); filed in Denmark 20 December 2000

To prevent children from removing the hearing aid battery and damaging the battery drawer, the battery drawer has two side features that prevent gripping the planar sides of the battery. The battery may be removed by inserting an object through an aperture in the battery drawer and pushing the battery out.—DAP

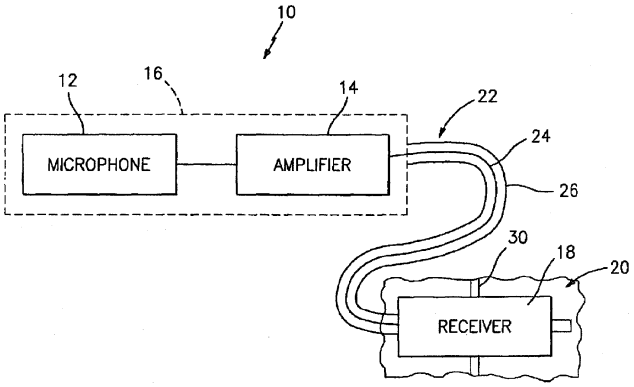


7,076,076

43.66.Ts HEARING AID SYSTEM

Natan Bauman, assignor to Vivatone Hearing Systems, LLC
11 July 2006 (Class 381/328); filed 10 September 2002

A hearing aid receiver is packaged in a housing that has several flexible arms made out of plastic or bendable wire. The tips of the flexible arms



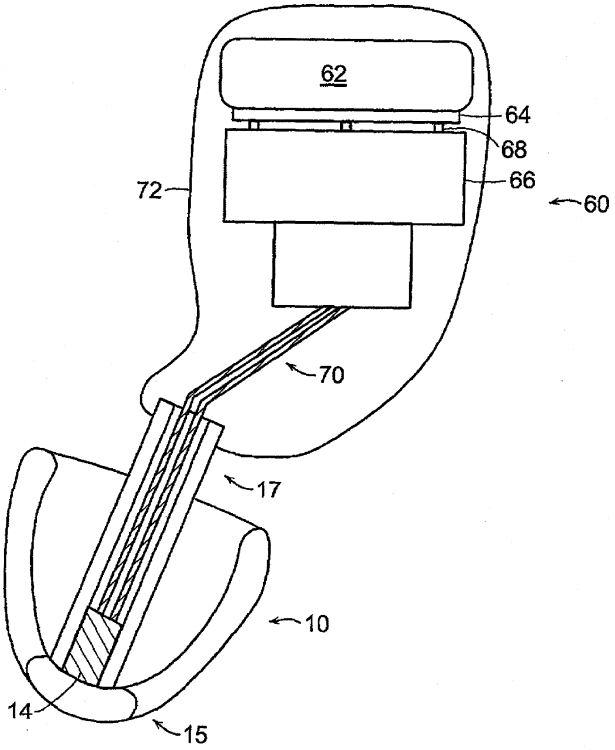
contact the ear canal wall and act as springs to keep the receiver housing in place.—DAP

7,082,206

43.66.Ts FLEXIBLE HEARING AID TIP WITH AN INTEGRAL RECEIVER

Derek D. Mahoney *et al.*, assignors to Sarnoff Corporation
25 July 2006 (Class 381/328); filed 13 March 2001

The hearing aid receiver is mounted within the bore of a flexible mushroom-shaped, compliant tip that is attached to a separate hearing aid



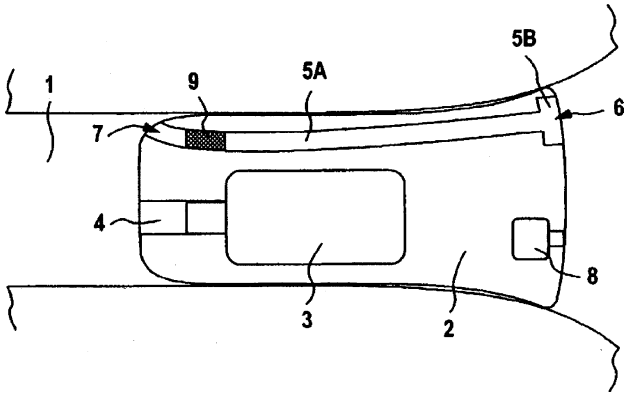
housing. The tip may be formed from rubber materials having different durometers to ensure stability for easy insertion, while retaining softness for a good acoustic seal and comfort.—DAP

7,079,662

43.66.Ts HEARING AID DEVICE WEARABLE IN THE EAR OR HEARING AID DEVICE HAVING AN OTOPLASTIC WEARABLE IN THE EAR

Torsten Niederdränk, assignor to Siemens Audiologische Technik GmbH
18 July 2006 (Class 381/328); filed in Germany 6 February 2002

To prevent acoustic feedback oscillation, high-frequency acoustic feedback through the hearing aid vent channel is suppressed to prevent it from getting back into the hearing aid microphone. The vent channel con-



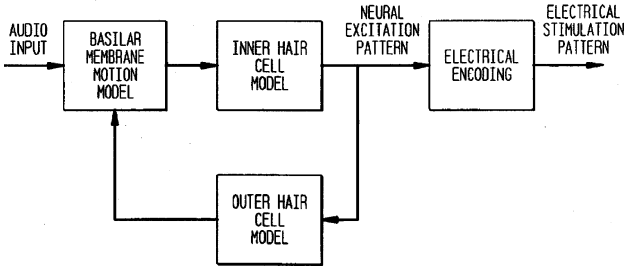
sists of a combination of a proximal long and narrow channel, which may contain an acoustic damper, and a very short distal portion with a wider cross section.—DAP

7,082,332

43.66.Ts SOUND PROCESSOR FOR A COCHLEAR IMPLANT

Peter John Blamey *et al.*, assignors to Cochlear, Limited
25 July 2006 (Class 607/57); filed in Australia 19 June 2000

Multiple-channel-per-electrode processing is used to produce a neural



excitation pattern which mimics the spatio-temporal pattern made by the traveling wave on a basilar membrane of a normal-hearing cochlea.—DAP

7,082,205

43.66.Ts METHOD FOR *IN SITU* MEASURING AND CORRECTING OR ADJUSTING THE OUTPUT SIGNAL OF A HEARING AID WITH A MODEL PROCESSOR AND HEARING AID EMPLOYING SUCH A METHOD

Soren Erik Westermann, assignor to Widex A/S
25 July 2006 (Class 381/312); filed 9 November 1998

An electroacoustic model of the hearing aid and ear simulates the sound signal that would be produced in an ear canal. The model is adapted with an error signal signifying the difference between the actual signal mea-

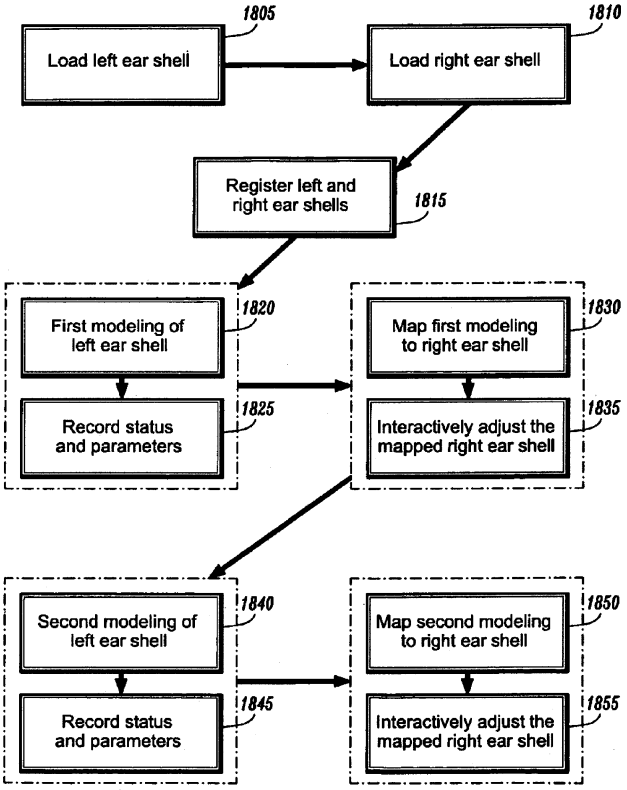
sured in the ear canal with a probe microphone and the simulated sound signal.—DAP

7,084,870

43.66.Ts INTERACTIVE BINAURAL SHELL MODELING FOR HEARING AIDS

Tong Fang *et al.*, assignors to Siemens Corporate Research, Incorporated
1 August 2006 (Class 345/420); filed 18 December 2003

This patent utilizes the symmetry between the physical shapes of a person's left and right ears in formulating computerized models for custom



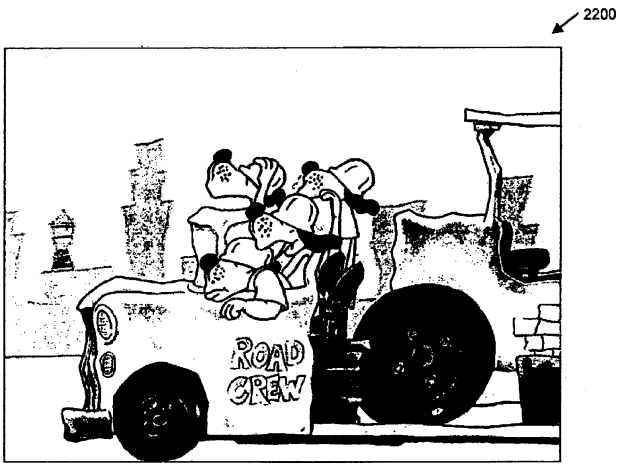
hearing aid ear shells. Data obtained from modeling the ear impression of the first ear shell is mapped for adjustment of the second ear shell model to eliminate inconsistencies and streamline the process.—DAP

6,986,663

43.70.Ep METHOD AND APPARATUS FOR AUTOMATED TRAINING OF LANGUAGE LEARNING SKILLS

Elizabeth H. Budra *et al.*, assignors to Scientific Learning Corporation
17 January 2006 (Class 434/169); filed 26 February 2004

The patent describes a comprehensive package of computer programs designed to teach language skills, including semantics, syntax, phonological and morphological categories, spelling, and sentence and paragraph comprehension, among others. Several of these branches of linguistics are defined



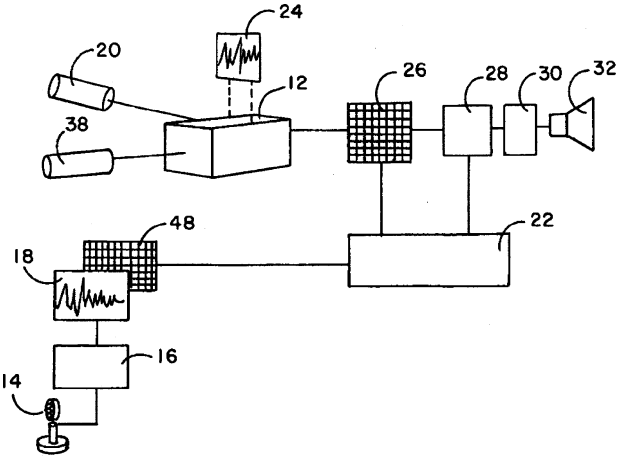
here and discussed in some detail and the game descriptions are fairly explicit as to how each of the language aspects is addressed during game play. As is seen in the many drawings of computer screens during game play, the system is clearly targeted to a fairly young audience.—DLR

6,978,240

43.70.Kv SPEECH TRANSLATION SYSTEM UTILIZING HOLOGRAPHIC DATA STORAGE MEDIUM

Gregory R. Brotz, Sheboygan, Wisconsin
20 December 2005 (Class 704/277); filed 31 January 2002

This patent entertains a rather radical notion: That the immense storage capacity and the associative recall capabilities of holographic storage could be used to translate one spoken language into another. The idea is that the raw waveforms of words or phrases in the source language would provide the holographic keys to look up the corresponding words or phrases in the target language. You then just pipe the result into your loudspeaker **32** and voilá. In fact, if your holographic memory is large enough, you can also get outputs in multiple languages simultaneously. Actually, most of the work, as



the system is described here, is done by a correlator **22**, considered to be external to the holographic process **12**. The author envisions, for example, that noise reduction would be possible. What is missing in the description is any real consideration of the range of variability that could be "corrected" by the holographic lookup process. My intuition is that you would need samples, not only of every possible word or phrase, but each of the same spoken in just about every possible manner. Maybe by 2106, our technology will be up to that requirement.—DLR

6,988,064

43.72.Ar SYSTEM AND METHOD FOR COMBINED FREQUENCY-DOMAIN AND TIME-DOMAIN PITCH EXTRACTION FOR SPEECH SIGNALS

Tenkasi V. Ramabadran and Alexander Sorin, assignors to Motorola, Incorporated
17 January 2006 (Class 704/218); filed 31 March 2003

Described here is a speech pitch extractor that uses both time-domain and frequency-domain analyzers to provide candidate pitch estimates for each frame of the speech signal. Another aspect of the system is that it be of low complexity, suitable for use in a portable, hand-held device, where portions of a speech recognition operation may be performed by a server elsewhere in the system. For the actual details of the frequency-domain pitch candidate extraction, the reader is referred to United States Patent 6,587,816. The time-domain system is essentially a down-sampled, low-pass-filtered autocorrelation system. The frequency-based analyzer is said to work best for low-pitch signals, whereas the correlator is said to perform better for high-pitch signals.—DLR

6,980,834

43.72.Ja METHOD AND APPARATUS FOR PERFORMING TEXT TO SPEECH SYNTHESIS

Vishwa N. Gupta and Paul Boucher, assignors to Nortel Networks, Limited
27 December 2005 (Class 455/563); filed 5 December 2002

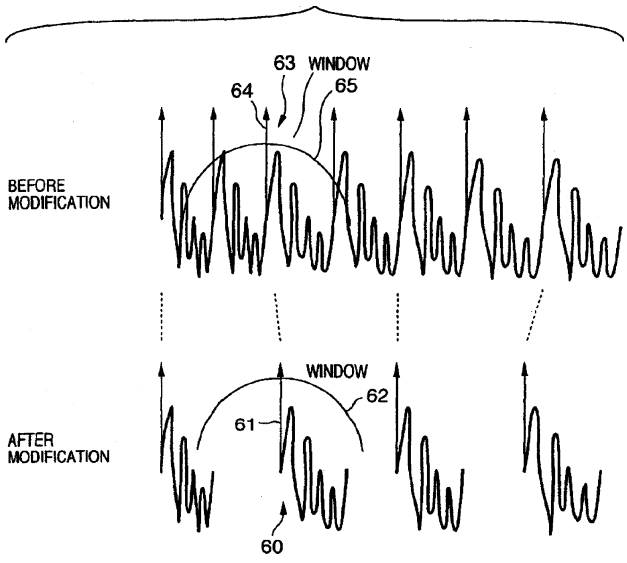
This patent is based on the well-known idea that a speech signal can be broken down into components, such as the content versus a vocal tract description. This idea is then applied in a manner almost as well-known, in the context of a central, or base, station connected to smaller terminal units, in this case, cell phones. By putting a speech synthesis process in the cell phone, the message content may be sent in the form of highly compact data bits, character codes, etc., while the vocal tract descriptions may be sent much less often. The claims cover linear predictive coding as the separation method, a technique used in speech processing since the 1970s. In other words, there is nothing new here. The speech processing details are not further discussed and exactly what gets separated out from what is equally unclear.—DLR

6,980,955

43.72.Ja SYNTHESIS UNIT SELECTION APPARATUS AND METHOD, AND STORAGE MEDIUM

Yasuo Okutani and Yasuhiro Komori, assignors to Canon Kabushiki Kaisha
27 December 2005 (Class 704/258); filed in Japan 31 March 2000

The patent describes a synthesis process by which speech waveform segments are selected from a database and modified as required in order to provide a good fit to the prior segment and to the message requirements. All of the modifications would be performed in the time domain, so there is no involvement of spectral considerations. The essence of the patent is an *n*-best matching process by which the nearest matching segment is selected.



Candidate segments are identified using a search method referred to as an "A*" technique, described only briefly. Concatenation is done using a pitch-synchronous overlap-and-add technique.—DLR

6,983,249

43.72.Ja SYSTEMS AND METHODS FOR VOICE SYNTHESIS

Hideo Sakai, assignor to International Business Machines Corporation
3 January 2006 (Class 704/258); filed in Japan 26 June 2000

This patent covers a voice-response order fulfillment system which selects characteristics of the synthesized speech according to stored customer information and preferences. With this system, the customer would specify the characteristics of the synthetic voice by choosing a particular character or personality, such as a well-known actor or television personality. In that case, a central aspect of the patented system is a provision by which appropriate royalties may be determined for the use of the famed voice characteristics.—DLR

6,988,068

43.72.Ja COMPENSATING FOR AMBIENT NOISE LEVELS IN TEXT-TO-SPEECH APPLICATIONS

Francis Fado and Peter J. Guasti, assignors to International Business Machines Corporation
17 January 2006 (Class 704/225); filed 25 March 2003

Essentially, this is nothing more than a sound-level meter connected to the output volume control of a speech synthesizer. Before a speech-output playback is started, a measurement is taken of the current ambient sound level. The reading is used to adjust the speech-output volume. Other niceties, such as controlling the speech pitch or formant structure to mimic human changes of vocal effort, are well beyond what is considered here.—DLR

6,988,069

43.72.Ja REDUCED UNIT DATABASE GENERATION BASED ON COST INFORMATION

Michael Stuart Phillips, assignor to Speechworks International, Incorporated
17 January 2006 (Class 704/258); filed 31 January 2003

This patent is concerned with methods of reducing the inventory of speech fragments needed by a low-complexity speech synthesizer of the type that works by concatenating fragments of prerecorded speech. This

synthesis method typically requires a substantial library of segmented speech fragments. The quality of the synthesized speech depends crucially on the range of suitable fragments available and how the selected fragments are modified or adjusted as they are joined with other selected fragments to make up the output speech signal. The emphasis in this patent is on a more elaborate segment modification procedure, allowing suitable-quality speech to be synthesized with a smaller inventory of recorded segments.—DLR

7,080,015

43.72.Ja SYNCHRONIZATION CONTROL APPARATUS AND METHOD, AND RECORDING MEDIUM

Keiichi Yamada *et al.*, assignors to Sony Corporation
18 July 2006 (Class 704/275); filed in Japan 28 December 1999

A technique is described by which a robot head could be made to move its ''articulators'' in sync with the output of a speech synthesizer. The method seems to be working on a phonemic basis, with various rules to shorten or lengthen certain movements in certain contexts. By this means, ''the robot imitates the utterance operations of human beings and animals more natural (sic).''—SAF

7,082,396

43.72.Ja METHODS AND APPARATUS FOR RAPID ACOUSTIC UNIT SELECTION FROM A LARGE SPEECH CORPUS

Mark C. Beutnagel *et al.*, assignors to AT&T Corporation
25 July 2006 (Class 704/258); filed 19 December 2003

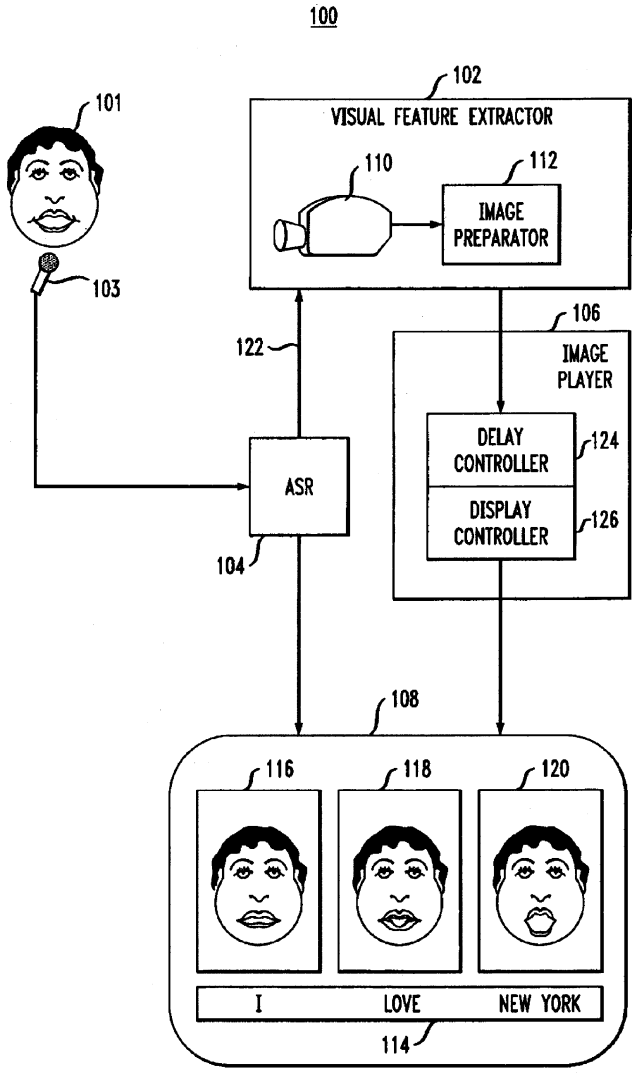
Standard sorts of speech synthesis systems concatenate small units of recorded speech from a large corpus. Each concatenation comes at a cost, which depends upon the degree to which the original recordings are compatible. This patent proposes a valuable technique for reducing the cost of this database computation for a large corpus by leveraging the fact that less than 1% of the possible sequential pairs of acoustic units generally occur in actual speech.—SAF

7,076,429

43.72.Lc METHOD AND APPARATUS FOR PRESENTING IMAGES REPRESENTATIVE OF AN UTTERANCE WITH CORRESPONDING DECODED SPEECH

Sara H. Basson *et al.*, assignors to International Business Machines Corporation
11 July 2006 (Class 704/272); filed 27 April 2001

The patent seems amusing at first, but makes sense in its intended context of enhancing speech recognition transcription aids for the hearing impaired. Normal automatic speech recognition (ASR)-generated text does not appear on a display until considerably after the utterance was actually spoken, thereby eliminating the chance to observe any mismatch between



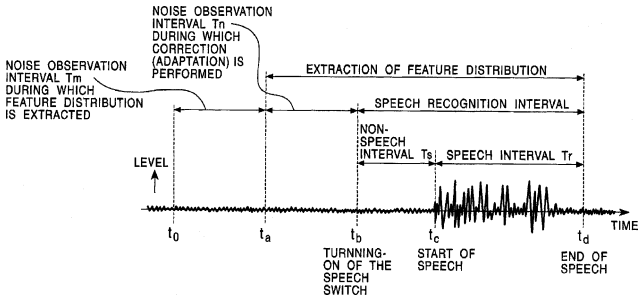
facial movements and erroneous text that could allow a hearing impaired user to realize that something went awry. The method described allows a few visual snapshots of the speaker's face to be snipped out of a video signal accompanying the speech, and displayed in relative time-alignment with the recognizer's textual output. It is not made clear how well this idea actually works.—SAF

6,985,860

43.72.Ne MODEL ADAPTATION APPARATUS, MODEL ADAPTATION METHOD, STORAGE MEDIUM, AND PATTERN RECOGNITION APPARATUS

Hironaga Nakatsuka, assignor to Sony Corporation
10 January 2006 (Class 704/244); filed in Japan 31 August 2000

In this speech recognition system, a waveform recorder unit continually samples the nonspeech environment up until the moment when speech begins. The spectral characteristics of the nonspeech sample are then used to adapt the models for analysis of the subsequently collected speech data. Three statistical methods are applied to separate speech and noise, referred



to here as ''most likelihood,'' ''complex statistic,'' and ''minimum distance-maximum separation,'' all described in some detail.—DLR

6,985.865

43.72.Ne METHOD AND SYSTEM FOR ENHANCED RESPONSE TO VOICE COMMANDS IN A VOICE COMMAND PLATFORM

Kevin Packingham and Elizabeth Roche, assignors to Sprint Spectrum L.P.
10 January 2006 (Class 704/275); filed 26 September 2001

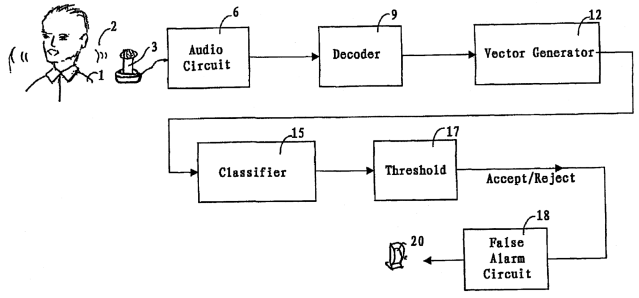
A computerized system of some sort includes a speech recognizer which can be used to enter commands for operating the system. Apparently, what the authors believe to be new here is that recognized commands may be used to change the recognition system itself, such as changing the recognition grammar, as well as changing computer system conditions. Some sorts of recognition grammars, not described here, but perhaps defined or discussed in some of the referenced publications, are referred to as global grammars and alias grammars. The intended context appears to be a sort of network browser application in which, for example, bookmarks may be selected by voice.—DLR

7,072,750

43.72.Ne METHOD AND APPARATUS FOR REJECTION OF SPEECH RECOGNITION RESULTS IN ACCORDANCE WITH CONFIDENCE LEVEL

Xiaobo Pi and Ying Jia, assignors to Intel Corporation
4 July 2006 (Class 701/36); filed 8 May 2001

This is what one might call a ''ham sandwich patent,'' meaning it finds a way to present a patentable technique for doing something that has already been done many times over. In this case, that is speech recognition of names



for mobile phone calling list access. In spite of the enthusiasm evident in the patent, the reported result of at best 53% correct rejection is hardly cause for excitement.—SAF

7,072,828

43.72.Ne APPARATUS AND METHOD FOR IMPROVED VOICE ACTIVITY DETECTION

Norman W. Petty, assignor to Avaya Technology Corporation
4 July 2006 (Class 704/210); filed 13 May 2002

This patent is related to resolving front-end clipping and long holder times in voice activation systems. A voice activation algorithm removes silence from uttered signals, which usually causes the speech utterance to sound choppy. This invention introduces queuing of the speech signal while the system is busy packetizing the speech portion of the uttered signal. The transmitter not only sends packets, it also sends a queuing signal along with the packet. Although the packetizing time is short and the signal acquired in this time period is not significant, the queuing signal helps to reduce choppiness of the signal in the receiver.—AAD

7,072,829

43.72.Ne SPEECH RECOGNITION FROM CONCURRENT VISUAL AND AUDIBLE INPUTS

Tetsujiro Kondo and Norifumi Yoshiwara, assignors to Sony Corporation
4 July 2006 (Class 704/222); filed in Japan 9 October 1998

This complex and almost incomprehensible patent details specific techniques for combining parameters computed from face images with typical speech acoustic parameters into a speech recognition procedure. The patent focuses on parameterizing the images and on combining this data with acoustics for vector codebook development.—SAF

7,072,833

43.72.Ne SPEECH PROCESSING SYSTEM

Jebu Jacob Rajan, assignor to Canon Kabushiki Kaisha
4 July 2006 (Class 704/233); filed in United Kingdom 2 June 2000

This invention introduces a computationally expensive method for detecting presence of voice in the audio signal stream. First, the input audio signal is parameterized. Then parameters of the signal are tested against prestored speech models with a pretrained probability density function to locate the speech in the signal.—AAD

7,072,834

43.72.Ne ADAPTING TO ADVERSE ACOUSTIC ENVIRONMENT IN SPEECH PROCESSING USING PLAYBACK TRAINING DATA

Guojun Zhou, assignor to Intel Corporation
4 July 2006 (Class 704/244); filed 5 April 2002

This patent mainly relates to a method for adapting an existing speech recognizer to a new environment which has different noise characteristic than the environment it was trained for. To this aim, the original training recordings are played back in the new environment and rerecorded. The new recordings are used for adaptation of the speech models.—AAD

7,072,835

43.72.Ne METHOD AND APPARATUS FOR SPEECH RECOGNITION

Tomohiro Konuma *et al.*, assignors to Matsushita Electric Industrial Company, Limited
4 July 2006 (Class 704/251); filed in Japan 23 January 2001

A method is described for reducing the usual proliferation of candidate words in a simplified speech recognition search procedure for a target word in a limited-vocabulary system. The specifics of the method are almost impossible to describe—or at least, that is one explanation for the evident problems that were experienced by the patent's authors.—SAF

7,072,836

43.72.Ne SPEECH PROCESSING APPARATUS AND METHOD EMPLOYING MATCHING AND CONFIDENCE SCORES

Yuan Shao, assignor to Canon Kabushiki Kaisha
4 July 2006 (Class 704/255); filed in United Kingdom 12 July 2000

A perennial problem with simplified, limited-vocabulary speech recognition systems is that they are prone to recognize almost anything as one of their few target items. This patent describes the computation of a confidence score for a word match within the typical Bayesian likelihood framework, which uses the *posterior* probability of a correct recognition result. Measuring match likelihood by the *posterior* makes this method nontextbook.—SAF

7,076,422

43.72.Ne MODELING AND PROCESSING FILLED PAUSES AND NOISES IN SPEECH RECOGNITION

Mei-Yuh Hwang, assignor to Microsoft Corporation
11 July 2006 (Class 704/226); filed 13 March 2003

Ergodic hidden Markov models are used to acoustically model, uh, filled pauses for a speech recognition system. The patent notes that such a model will generalize to explicitly model any number of nonspeech noises including breathing, and even silence. With such noises being modeled, the patent notes that they also then must be included in the language models. This would seem to enlarge the search space significantly, since the location of such noises cannot be grammatically predicted to any extent.—SAF

7,076,425

43.72.Ne VOICE RECOGNITION DEVICE WITH LARGER WEIGHTS ASSIGNED TO DISPLAYED WORDS OF RECOGNITION VOCABULARY

Takeshi Ono and Okihiko Nakayama, assignors to Nissam Motor Company, Limited
11 July 2006 (Class 704/252); filed in Japan 19 March 2001

This patent proposes to enable a user to assist with a speech recognition search on the fly by providing, in a display, some number of possible word matches ranked by confidence. The user could then select the displayed term(s), so that their weighting could be increased during continued search. The results of this process could presumably be used to update the models used by the system. Alternative embodiments allow the user to provide a range of possible terms on the display in advance of recognition, to effectively limit the search space. The intended application seems to be navigation systems, where a range of street names could be listed alphabetically.—SAF

7,076,428

43.72.Ne METHOD AND APPARATUS FOR SELECTIVE DISTRIBUTED SPEECH RECOGNITION

Tasos Anastasakos *et al.*, assignors to Motorola, Incorporated
11 July 2006 (Class 704/270.1); filed 30 December 2002

A system for distributed speech recognition is proposed where, in advance of recognition search, a ''grammar type'' is selected for the server-side engine. The patent wants to cover all conceivable ways of actually indicating the grammar type (e.g., ''days of the week''), since it is seemingly patenting the very notion of selecting from among multiple grammars in advance of the search.—SAF

7,080,011

43.72.Ne SPEECH LABEL ACCELERATORS AND TECHNIQUES FOR USING SAME

Yoanna Baumgartner *et al.*, assignors to International Business Machines Corporation
18 July 2006 (Class 704/231); filed 2 August 2001

This comprehensive patent discusses programming techniques increasing the efficiency of speech recognition that may be suitable for use in smaller communications devices such as personal digital assistants (PDAs).—JME

7,082,391

43.72.Ne AUTOMATIC SPEECH RECOGNITION

John W. Merrill, assignor to Intel Corporation
25 July 2006 (Class 704/231); filed 14 July 1998

In the context of a voice-enabled computer software application, spoken commands to the computer are the object of speech recognition. This patent presents a very simple and general outline of the idea of limiting the recognition search to a relevant subvocabulary dependent upon the particular menu, dialog, etc. that is being used in the application at the time of spoken input. It thereby does not entirely live up to its all-inclusive title.—SAF

7,082,393

43.72.Ne HEAD-WORN, TRIMODAL DEVICE TO INCREASE TRANSCRIPTION ACCURACY IN A VOICE RECOGNITION SYSTEM AND TO PROCESS UNVOCALIZED SPEECH

Roy J. Lahr, assignor to RAST Associates, LLC
25 July 2006 (Class 704/233); filed 27 March 2002

This patent describes a head-worn device including a microphone for speech recognition, but also a set of mouth-observing cameras and an ultrasonic generator which can inject an acoustic signal into the mouth. The main, rather surprising innovation in this patent is the idea that, by acoustically probing the shape of the vocal tract, and supplementing the information with image data from the cameras, speech recognition could be performed using ''unvocalized'' mouth movements such as would result from whispering or even simply going through the motions of the speech. Sadly, there is no indication that this complex and incredible scheme has actually been implemented or tested.—SAF

7,082,395

43.72.Ne SIGNAL INJECTION COUPLING INTO THE HUMAN VOCAL TRACT FOR ROBUST AUDIBLE AND INAUDIBLE VOICE RECOGNITION

Carol A. Tosaya and John W. Sliwa, Jr., both of Los Altos, California
25 July 2006 (Class 704/246); filed 3 October 2002

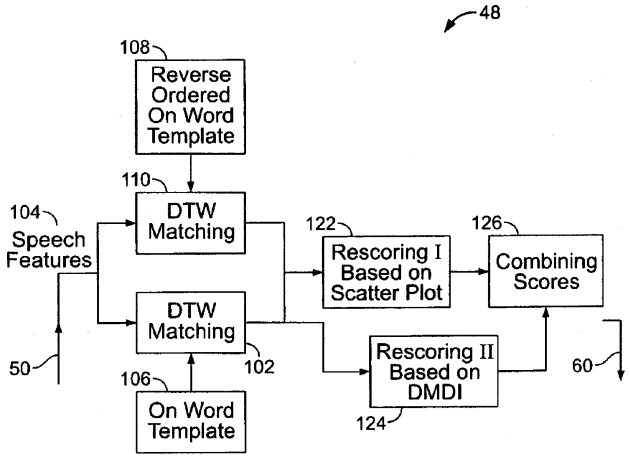
This is another patent revolving around a method of performing speech recognition without actually requiring voiced speech. The suggestion once again is to inject an acoustic probe signal into the vocal tract to extract information about the resonant frequencies, either for the purpose of supplementing the information garnered from ordinary speech, or to completely replace that information and allow the speaker to not vocalize the speech.—SAF

7,085,717

43.72.Ne SCORING AND RE-SCORING DYNAMIC TIME WARPING OF SPEECH

Veton Z. Kepuska and Harinath K. Reddy, assignors to Thinkengine Networks, Incorporated
1 August 2006 (Class 704/241); filed 21 May 2002

After time warping, two distances are measured between vectors representing an utterance and those representing a template. The two measurements are made with different vector orders achieved, for example, by reversing the orders of the utterance vectors and the template vectors. If the



second distance measures are outside of a predetermined threshold, the template is considered to represent an out-of-vocabulary utterance, even if the first distance measure implies a match.—DAP

7,070,565

43.80.Vj SOLID HYDROGEL COUPLING FOR ULTRASOUND IMAGING AND THERAPY

Shahram Vaezy *et al.*, assignors to University of Washington
4 July 2006 (Class 600/459); filed 30 May 2003

These couplings transmit ultrasound well, are shaped to accommodate the focus of a high intensity focused ultrasound (HIFU) beam, withstand high temperatures encountered in HIFU therapy, and allow application of medication in the coupling material to tissue in contact with the coupler.—RCW

7,074,187

43.80.Vj SYSTEM AND METHOD FOR IMPROVING ULTRASOUND IMAGE ACQUISITION AND REPLICATION FOR REPEATABLE MEASUREMENTS OF VASCULAR STRUCTURES

Robert H. Selzer, Los Angeles and Howard N. Hodis, South Pasadena, both of California
11 July 2006 (Class 600/440); filed 13 December 2002

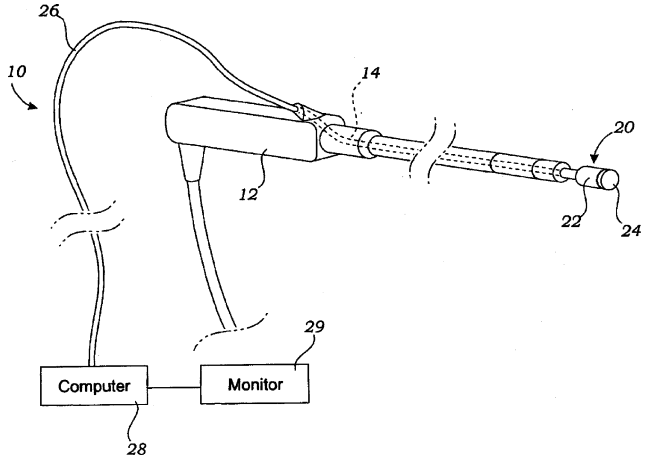
A computer-based method is used to facilitate accurate replication of images obtained over examinations at different times. In the method, a split-screen display shows an image from an earlier examination on one side of the screen and a real-time image from the current examination on the other side of the screen. The ultrasound transducer producing the real-time image is adjusted manually to obtain an image that best matches the image from the earlier examination.—RCW

7,074,189

43.80.Vj ENDOSCOPICALLY DELIVERABLE ULTRASOUND IMAGING SYSTEM AND METHOD OF USE

Valentino Montegrande, Irvine, California
11 July 2006 (Class 600/462); filed 24 November 2003

In this imaging system, an ultrasound probe is placed in the channel of



an endoscope and is connected to a computer by a flexible cord that fits through the channel.—RCW

7,077,807

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

Hans Torp *et al.*, assignors to G.E. Vingmed Ultrasound AS
18 July 2006 (Class 600/438); filed 10 November 2003

Quantitative descriptions of tissue parameters such as velocity, displacement, and strain or strain rate are calculated in a direction that is not necessarily along the direction of the ultrasound beam and the calculated descriptions are displayed as functions of time or spatial position.—RCW

Review of field studies of aircraft noise-induced sleep disturbance

David S. Michaud[a)]
Consumer & Clinical Radiation Protection Bureau, Acoustics Division, Health Canada, Ottawa, Ontario, K1A 1C1 Canada

Sanford Fidell
Fidell Associates, 23139 Erwin Street, Woodland Hills, California 91367

Karl Pearsons
Pearsons Psychoacoustics, 22689 Mulholland Drive, Woodland Hills, California 91364

Kenneth C. Campbell
School of Psychology, University of Ottawa, Ottawa, Ontario, K1N 6N5 Canada

Stephen E. Keith
Consumer & Clinical Radiation Protection Bureau, Acoustics Division, Health Canada, Ottawa, Ontario, K1A 1C1 Canada

(Received 5 May 2006; revised 23 August 2006; accepted 16 October 2006)

Aircraft noise-induced sleep disturbance (AN-ISD) is potentially among the more serious effects of aircraft noise on people. This literature review of recent field studies of AN-ISD finds that reliable generalization of findings to population-level effects is complicated by individual differences among subjects, methodological and analytic differences among studies, and predictive relationships that account for only a small fraction of the variance in the relationship between noise exposure and sleep disturbance. It is nonetheless apparent in the studied circumstances of residential exposure that sleep disturbance effects of nighttime aircraft noise intrusions are not dramatic on a per-event basis, and that linkages between outdoor aircraft noise exposure and sleep disturbance are tenuous. It is also apparent that AN-ISD occurs more often during later than earlier parts of the night; that indoor sound levels are more closely associated with sleep disturbance than outdoor measures; and that spontaneous awakenings, or awakenings attributable to nonaircraft indoor noises, occur more often than awakenings attributed to aircraft noise. Predictions of sleep disturbance due to aircraft noise should not be based on over-simplifications of the findings of the reviewed studies, and these reports should be treated with caution in developing regulatory policy for aircraft noise. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2400613]

PACS number(s): 43.50.Qp, 43.50.Lj [BSF] Pages: 32–41

I. INTRODUCTION

Aircraft noise-induced sleep disturbance (AN-ISD) is often viewed as a potential public health hazard because common experience suggests that adequate sleep is essential to overall well being. Therefore, it is reasonable to explore whether the prevalence of AN-ISD can act as the basis for regulatory policy regarding tolerable levels of community exposure to aircraft noise. This paper reviews recent studies of aircraft noise and sleep to examine how they can be used to yield inferences about the prevalence of AN-ISD.

The effects of noise on sleep are mediated by many factors, including sound level, number, duration, time of occurrence, short- and long-term intermittency and consistency of distributions of aircraft noise intrusions into sleeping quarters. Uncertainty in estimates of at-ear aircraft noise levels and the degree to which noise events exceed at-ear background noise levels in sleeping quarters, as well as individual differences such as age, sex, noise sensitivity, sensitization and habituation, health status and the salience of intruding noises also affect the ability of aircraft noise to disturb sleep (Nordic Council of Ministers, 1994). Thus, it is challenging to summarize and predict population-level sleep disturbance by aircraft noise (Finegold and Elias, 2002; Anderson and Miller, 2005; Passchier-Vermeer, 2003). As is the case with efforts to predict other population-level environmental noise effects, the limitations and reliability of the underlying sleep disturbance data and interpretive methods merit close attention.

After a short review of sleep patterns and methods used for assessing them, this article focuses on findings of field studies of AN-ISD reported between 1990 and 2003. Information is reviewed about the ability of aircraft noise to (1) interfere with the ability to fall asleep, (2) curtail sleep duration, (3) lessen the perceived quality of sleep, (4) awaken people from sleep, and (5) increase bodily movements during sleep. Physiological changes associated with aircraft noise exposure, such as changes in stress hormone levels, are re-

[a)]Author to whom correspondence should be addressed. Electronic mail: dmichaud@hc-sc.gc.ca

viewed elsewhere (e.g., Health Canada, 2001). Laboratory findings are considered only to the extent that they relate to physiological costs of accumulated sleep debt, as it is unclear how sleep disturbance information collected in places other than familiar sleeping quarters can be generalized to in-home environments (Pearsons *et al.*, 1995; Michaud *et al.*, 2005), and recent field experiments provide direct information about in-home AN-ISD in any event.

II. CHARACTERIZATION OF DISTURBANCE OF SLEEP PATTERNS

Sleep cycles, as determined from polysomnography (PSG), typically last from 90 to 110 min and occur between four and six times during a full night. People cycle through lighter and deeper sleep stages throughout the night, with considerable individual differences in patterns from night to night. Electroencephalographic (EEG) activity during sleep is classified with respect to the frequency and shape of wave forms into Stages 1, 2, 3, 4, and rapid eye movement (REM). Stage 1 is the lightest sleep. Deeper stages of sleep are associated with decreases in respiration and heart rates in Stages 2, 3, and 4. Since EEG wave forms during Stages 3 and 4 exhibit the lowest frequency, they are also referred to as slow wave sleep (SWS).

A typical night's sleep that is undisturbed by noise includes about 2 h of slow wave sleep, three quarters of which accumulate in the first half of the night. In contrast, REM sleep, which also lasts for about 2 h, occurs predominantly during the latter half of the night. Passchier-Vermeer *et al.* (2002), among others, suggested that REM sleep interruption is most likely to occur in the presence of continuous noise, while slow wave sleep interruption is more sensitive to intermittent noise intrusions, such as those produced by aircraft noise. Recovery from sleep loss is characterized by protracted periods of time spent in the stage of sleep that has been curtailed. Slow wave sleep is generally restored before other stages (De Gennaro *et al.*, 2005).

Awakening is more likely to occur from the REM stage of the ultradian non-REM/REM sleep cycle. People probably do not fully adapt to accumulated sleep debt (Dinges *et al.*, 1997; van Dongen *et al.*, 2003). While people are capable of some degree of accommodation to a sleep-depriving schedule, a modicum of evidence suggests that cognitive functioning as measured by judgment, reaction time, and other tasks may remain impaired (Pilcher *et al.*, 1996; Williamson *et al.*, 2000; Harrison *et al.*, 2000) despite apparent adaptation.

While noise exposure can increase the likelihood of a shift from a deeper to a lighter sleep stage, spontaneous awakenings or shifts in body position often occur in the transition from SWS to Stage 1 or Stage 2 sleep, even in the absence of noise intrusions. Also, REM sleep may also terminate with an abrupt arousal or awakening. Arousals from sleep may be due to normal physiological processes, and can also serve functional purposes in calling attention to imminent danger. Bodily movements are a necessary means of relieving pressure points.

Although PSG remains the most sensitive means for assessing changes in sleep states, it has disadvantages of cost, intrusiveness to sleepers, and complexity of interpretation that limit its usefulness in field settings. Behaviorally confirmed awakening (BCA) as well as actimetry, both indirect behavioral measures of sleep, have become useful alternatives to PSG as field methods for assessing several aspects of sleep. Rylander *et al.* (1972) and Horonjeff *et al.* (1982) demonstrated that subjects in familiar sleeping quarters can reliably push a button upon awakening for periods of many nights, and that their responses can be usefully related to nearby field measurements of nighttime noise levels, whether produced by aircraft or other noise sources.

An actimeter is a wristwatch-like device that detects and stores acceleration or motion in excess of user-defined thresholds within successive temporal "epochs." Since people are more active during waking than sleep, actimetric data can be analyzed to infer values of sleep parameters from patterns of activity and inactivity. As a convenient and cost-effective alternative to PSG, actimetry can provide reasonable measures of total sleep time, total wake time, and numbers of awakenings. Actimetry can also provide a useful approximation of "arousal" (originally a term that implied a shift from deeper sleep stages into lighter or Stage I sleep) that has been shown to be consistent with EEG-measured awakenings (Ollerhead *et al.*, 1992).

Sleep actimetry cannot, however, yield information about time spent in different sleep stages, amounts of non-REM and REM sleep, or fragmentation due to brief arousals. Further, a high correlation between total sleep time defined by PSG criteria and total sleep time defined by actimetry can occur if actimetry overestimates total sleep time, as by misclassifying time awake without movement as sleep. Ancoli-Israel *et al.* (2003) suggested that actimetry is more likely to detect sleep than to detect wake states, leading to high sensitivity but low specificity and accuracy. Thus, correlations between PSG- and actimetrically defined sleep may not fully justify reliance on the simpler technique.

III. FINDINGS FROM RECENT FIELD STUDIES

Most recent field studies of AN-ISD have relied on actimetric, BCAs techniques and questionnaires to assess sleep disturbance due to nighttime aircraft noise exposure. Only one study (Horne *et al.*, 1994) has attempted to cross validate actimetric with EEG information. All recent field studies included male and female adult subjects exposed to a range of nighttime aircraft noise levels. Not all of these studies were able to definitively attribute nighttime noise intrusions in sleeping quarters to aircraft. They are reviewed nonetheless, because reasonable analysis assumptions were made to support the inference that the sources of the noise events were very likely to have been aircraft. Aircraft noise levels were measured both indoors and outdoors in some studies, while noise levels in sleeping quarters were estimated in others by assuming typical structural noise reductions. In nearly all cases in which both indoor and outdoor sound measurements were made, outdoor sound levels failed to predict sleep disturbance. Table I summarizes the basic methods of each field study.

TABLE I. Comparison of methodological aspects of five recent field studies of aircraft noise-induced sleep disturbance.

Reference	Sample size (Number of Subjects/ Subject-Nights)	Definition of Aircraft noise Events	Measurement Location(s)	Indications of Sleep Disturbance
Passchier-Vermeer *et al.* (2002)	418/4598	Time of occurrence of overflight as defined by airport noise monitoring system	Indoors and outdoors	Motility in 15 s epochs; questionnaire responses; reaction times
Ollerhead *et al.* (1992)	400/5742	Simultaneous outdoor A-weighted sound levels at several monitoring locations, checked against airport logs	Outdoors	Motility in 30 s epochs; limited EEG; sleep logs
Fidell *et al.* (1995a)	85/1887	A-weighted outdoor and indoor sound levels in excess of site-specific thresholds for more than 2 s	Indoors and outdoors	Behavioral awakening; questionnaire responses
Fidell *et al.* (1995b)	77/2717	A-weighted outdoor sound levels in excess of site-specific thresholds (60 or 70 dB at different airports) for more than 2 s; A-weighted indoor sound levels in excess of site-specific thresholds (50 or 60 dB at different airports) for more than 2 s	Indoors and outdoors	Behavioral awakening; two forms of motility in 30 s epochs; questionnaire responses
Fidell *et al.* (2000)	22/686	Indoor A-weighted sound level in excess of 50 dB for 10 s; outdoor A-weighted sound level in excess of 60 dB for 10 s	Indoors and outdoors	Behavioral awakening; motility in 30 s epochs; questionnaire responses

A. Ollerhead *et al.* (1992)

Ollerhead *et al.* (1992) and Hume *et al.* (2003) describe a large-scale study of awakenings and actimetrically monitored movements at two sites around each of Heathrow, Gatwick, Stansted and Manchester airports. The subjects at the eight sites included 211 female and 189 male residents between the ages of 20 and 70 years, sleeping in their own homes. Subjects completed sleep diaries and wore actimeters for 15 nights. The sleep of a subsample of subjects was also monitored with in-home EEG instrumentation (Hume *et al.*, 2003). Outdoor noise measurements were taken of aircraft noise event data to link with actimetric data recorded in successive 30 s epochs. Responses collected by actimetry from the 50 subjects at each site were pooled and averaged for comparison with aircraft noise events for that site. The actimeter clocks remained synchronous within ±5 s. An aircraft noise event in the Ollerhead *et al.* study was defined as the occurrence of an outdoor sound in excess of a 60 dB threshold. The number of aircraft noise events during the night varied across the eight sites from 1 to 20.

The 400 subjects awakened from sleep 6457 times, of which 351 (5.4%) awakenings could be attributed to aircraft noise events. Awakenings attributable to aircraft noise events were far less common than those ascribed to toilet visits, tending to children, and other non-noise specific reasons. Sleep became more disturbed in general as the night progressed, but not necessarily because of exposure to aircraft noise events.

The main finding of Ollerhead *et al.* (1992) was that very few of the test subjects were at risk of substantial sleep loss due to aircraft noise. Ollerhead *et al.* (1992) noted that sleep was largely unaffected by aircraft noise events at outdoor L_{max} values lower than about 80 dB (SEL ~90 dB). Ollerhead *et al.* (1992) showed that above 90 dB SEL, the awakening rate due to an aircraft noise event was somewhere between 1 in 60 and 1 in 100. Ollerhead *et al.* (1992) attributed the infrequency of AN-ISD to the familiarity and adaptation of neighborhood residents to the noise source.

Although large variations in numbers of aircraft noise events were observed across the eight study locations, variability in actimetric responses was relatively small. Ollerhead *et al.* (1992) further noted that sensitivity to aircraft noise was lower during the earlier part of the sleep period than during the later part of the sleep period.

Ollerhead *et al.* (1992) also noted that individuals who classified themselves as most sensitive to noise were 2.5 times more likely to be awakened by an aircraft noise event than individuals who classified themselves as the least sensitive to noise.

B. Fidell *et al.* (1995a, 1995b, 2000)

Fidell *et al.* (1995a) reported a field study of 1 month duration in which simultaneous measurements were made of aircraft noise and sleep disturbance in the homes of 27 individuals living near the main runway of a military airfield, and of 35 subjects living near Los Angeles International Airport (LAX). An additional 23 subjects living in neighborhoods without appreciable aircraft noise exposure served as controls. Among the 85 subjects, who ranged in age from 19 to 79 years of age, 38 were men and 47 were women. Sub-

jects were instructed to press a bedside button upon awakening for any reason whatsoever. No actimetric or EEG measurements were made.

Noise measurements were taken outdoors in the vicinity of residents living near the two airport locations and within the bedrooms of each test participant. Average A-weighted sound levels were recorded every 2 s between 2200 and 0800 h. Each subject also completed an evening and morning questionnaire intended to assess subjective tiredness during the day, overall sleep quality, and recalled number of awakenings during the night, as well as their estimated time to fall asleep.

Fidell *et al.* (1995a) were able to attribute about 16% of the awakenings to noise events. For those awakenings that could be attributed to aircraft noise intrusions in sleeping quarters, each 1 dB increase in SEL increased the likelihood of awakening by only 0.17%. Like Ollerhead *et al.* (1992), Fidell *et al.* (1995a) observed that the likelihood of awakening due to a noise event (although not necessarily an aircraft noise event) increased throughout the sleep period. Fidell *et al.* (1995a) found an increase of a factor of 1.06 in the likelihood of awakening for each 15 min since retiring. Subjective reports of evening tiredness were related to awakenings by noise events.

Ambient noise levels in the bedrooms were inversely related to awakenings, such that the odds of awakening were reduced by a factor of 0.05 for each 1 dB increase in ambient noise levels. This finding resembles the observation of Passchier-Vermeer *et al.* (2002) (described below) that indoor average sound levels are inversely related to the probability of motility attributable to individual aircraft noise events. For an increase of one spontaneous awakening, Fidell *et al.* (1995a) showed that the probability of awakening due to a noise event was reduced by a factor of 0.26.

The mean indoor SEL that awakened subjects was 81 dB, while the mean indoor SEL that failed to awaken subjects was 74 dB. No change in the awakenings was observed when nighttime aircraft noise exposure at the military airfield was reduced by 6 dB (from L_{eq} 54 to 48 dB) over weekends. Fidell *et al.* (1995a) found that indoor SEL was the only reliable predictor of sleep disturbance within 2 and 5 min of a noise event. Although greater SEL values were associated with a greater likelihood of awakening to aircraft noise, the slope of the relationship was quite shallow: an increase of 10 dB in SEL was associated with only a 1.7% increase in awakenings. Cumulative noise exposure over the entire night did not predict sleep disturbance, and hence the study did not support adoption of L_{night} as a useful predictor of sleep disturbance.

Fidell *et al.* (1995b) reported another field study in which motility and BCA were used as indices of sleep disturbance. The venue for the second study was Denver, Colorado, where an opportunity was available to observe both the effect of reductions in aircraft noise on sleep among people living near Stapleton International Airport (DEN), which was scheduled to close, and the effect of increases in aircraft noise on sleep among people living near the new Denver International Airport (DIA), which was about to open. As in the prior study by these authors, simultaneous noise measurements were made both outdoors and in sleeping quarters. In total, 2717 subject nights of observations were made. Subjects ranged in age from young adults to the elderly, and were evenly distributed by gender.

In addition to subjective reports from evening and morning questionnaires, both actimetric and behavioral awakening measurements of sleep disturbance were made in 30 s epochs during three nighttime periods: 0100–0130, 0300–0330, and 0500–0530. The percentage of noise-induced behavioral awakenings increased 0.25% per 1 dB increase in indoor SEL. It was also found that for each increase of 1 dB in ambient L_{eq} levels, the actimetric and behavioral awakening responses due to noise events decreased by 2%–6%. Noise events were more likely to awaken men than women.

A statistically significant negative trend in behavioral awakenings was noted following the start of aircraft operations at DIA, despite a large increase in the number of indoor noise events. Prior to the opening of DIA, an average of 1.71 behavioral awakenings was observed per night. Following the opening of DIA, the average awakenings decreased to 1.13 per night. Average numbers of behavioral awakenings per night before (1.8) and after (1.64) the closing of DEN were not reliably different from one another. The similar number of awakenings per night may have been related to the failure of indoor noise levels to change appreciably, despite a decrease in outdoor levels from a nighttime L_{eq} of 58–46 dB.

The percentage of 30 s epochs containing actimetrically detected bodily movements ranged from 17% for noise events between 65 and 69 dB L_{max} indoor to 31% for events between 70 and 74 dB L_{max} indoor. Considerable variability was observed between indoor L_{max} values and motility. Motility was greater than 17% for noise events when L_{max} was below 65 dB, but less than 31% for noise events when L_{max} was higher than 74 dB.

Subsets of the subjects around DEN prior to its closure wore two types of actimeters. It was found that an actimeter of Swiss manufacture was more likely to detect motility (1.23% increase/dB increase in SEL) with respect to a U.S. model (0.4% increase/dB increase in SEL). The probability of motility occurring within 5 min of a noise event was 0.90 (Swiss model) and 0.84 (U.S. model) for the indoor SEL. No such relationship with outdoor SEL values was reliably observed. The linear relationship between indoor SEL and the percentages of subjects exhibiting motility following a noise event was:

$$\%\ \text{motility (Swiss)} = -23.74 + 1.23(\text{SEL}), \quad (1)$$

$$\%\ \text{motility (U . S .)} = 47.16 + 0.4(\text{SEL}), \quad (2)$$

where applicable indoor SEL values are in the range of 50–100 dB.

Fidell *et al.* (2000) reported another pre/post study of the effects of aircraft noise on sleep in anticipation of an expected increase in air traffic at a general aviation airport (DeKalb-Peachtree). Indoor and outdoor sound levels were again monitored in the sleeping quarters of 22 participants during a total of 686 subject nights before, during, and following the summer Olympic games in Atlanta. The number

of noise events between 76 and 80 dB L_{max} increased slightly during the games. The number of events prior to the game in the range of 61–75 dB L_{max} was greater than during the games, but fewer following them.

Behaviorally confirmed awakenings were greatest (1.8 per night) prior to the games, and dropped slightly to 1.3 per night during the games, and to 1.0 per night following the games. Indoor SEL predicted actimetrically monitored arousals (at a 5% rate of increase per 10 dB increase in SEL), while outdoor SEL predicted behavioral awakenings (1.3% rate of increase per 10 dB increase in SEL). The variability in this response was greater at the higher outdoor SEL values, so that the prevalence of awakening at 100 dB ranged from 0% to 20%, but only from 0% to 2% at 60 dB. Even at high noise levels most people in this study were not awakened by aircraft overflights.

C. Passchier-Vermeer *et al.*, 2002

Passchier-Vermeer *et al.* (2002) reported a study of sleep disturbance conducted in the vicinity of Amsterdam's Schiphol Airport (AAS). Aircraft noise in this study was monitored within the bedrooms of 418 subjects and at several outdoor locations over an 11 day period. Aircraft noise exposure levels varied with distance from AAS. Subjects ranged in age from 18 to 81 years, and were evenly divided by gender. They answered morning and evening questions regarding sleep quality, recalled awakenings due to aircraft noise, and their annoyance due to aircraft noise. Motility was monitored throughout full 24 h days for the duration of the study. The wrist-mounted actimeter was also equipped with an event marker, which subjects pressed to indicate that they had been awakened. Subjects reported their subjective sleepiness at five designated time periods over the course of the day. They also performed a reaction time test intended to assess the effects of sleep loss on performance.

The effects of aircraft noise on sleep were assessed on "instantaneous," 24 h, and long-term time scales. For instantaneous effects, the probability of aircraft noise-induced motility (and onset of motility) was actimetrically measured during consecutive 15 s interval epochs and related to indoor L_{max} and SEL.[1] The authors defined aircraft noise-induced motility as movement occurring within **any** 15 s interval of an aircraft noise event and aircraft noise-induced *onset* of motility as movement within a 15 s epoch immediately following an interval in which movement had **not** occurred during an aircraft noise event.

The analysis of effects on a 24 h time scale included the sleep period time and the waking period after the overnight sleep period. In addition to the instantaneous measures, Passchier-Vermeer *et al.* (2002) examined perceived sleep quality, BCA, and questionnaire responses. Noise metrics of interest were the indoor equivalent aircraft sound level and the number of aircraft during sleep period time.

For the long-term time scale, the authors considered variables aggregated over the 11 nights of the study, including mean motility and responses obtained from morning questionnaires for this time scale. Noise metrics included indoor aircraft sound levels assessed over 11 sleep periods for individual subjects, and outdoor metrics representative of long-term nighttime aircraft noise exposure at 15 designated locations. The equivalent indoor aircraft sound level from 2300 to 0700 h was used to assess aircraft noise effects on responses to questionnaire items.

For the instantaneous effects, Passchier-Vermeer *et al.* (2002) found that aircraft noise events increased probabilities of both motility and onset of motility. The probability of motility increased with increasing indoor L_{max}, such that at 68 dB, the probability of motility during an aircraft noise event was about three times greater than the probability of motility in the absence of aircraft noise. The authors showed that an indoor L_{max} of 32 dB and an indoor SEL of 38 dB were the thresholds for increased probability of motility. The corresponding thresholds for probability of *onset* of motility were indoor L_{max} of 32 dB and indoor SEL of 40 dB.

Equations (3) through (6) from this study predicted motility and onset of motility during the 15 s epoch of an aircraft noise event, where the indoor L_{max} occurs

$$P(\text{motility}) = 0.00063(L_{max} - 32) + 0.0000314(L_{max} - 0.0000314)^2, \quad (3)$$

$$P(\text{motility}) = 0.000532(\text{SEL} - 38) + 0.0000268(\text{SEL} - 0.0000268)^2, \quad (4)$$

$$P(\text{onset of motility}) = 0.000415(L_{max} - 32) + 0.00000884(L_{max} - 0.00000884)^2, \quad (5)$$

$$P(\text{onset of motility}) = 0.000273(\text{SEL} - 40) + 0.00000357(\text{SEL} - 0.00000357)^2, \quad (6)$$

where applicable indoor L_{max} values are in the range 32–70 dB for Eqs. (3) and (5). The applicable SEL values are in the range of 38–80 dB and 40–80 dB for Eqs. (4) and (6), respectively. Passchier-Vermeer *et al.* indicated that these L_{max} and SEL thresholds for increases in the probabilities of motility and onset of motility are about 15 dB lower than those estimated by Ollerhead *et al.* (1992). Passchier-Vermeer *et al.* also found that subjects indicated an awakening, by way of marker pressing, in 5951 of the 7 864 899 15 s epochs assessed in this study (i.e., 0.0757%). Of these marker presses, 763 (0.0807%) occurred during the 945 939 15 s epochs that coincided with an aircraft noise event and 5188 (0.075%) occurred during the 6 918 960 15 s epochs monitored outside of the aircraft noise event. This difference was reported to be statistically different using a 1 tailed test, at $p < 0.05$.

Consistent with an observation made by Fidell *et al.* (1995b), Passchier-Vermeer *et al.* (2002) found that the instantaneous measures were strongly influenced by the average equivalent indoor ambient sound level assessed over the 11 sleep periods. When indoor equivalent levels were low, the probability of motility due to aircraft noise was greater, especially at the higher L_{max} levels. This suggests that people

accustomed to sleeping in quieter quarters may be more likely to experience motility when exposed to intruding noise than people who customarily sleep in noisier quarters. The authors also showed that the probability of motility increased as a function of time after sleep onset. After 7 h of sleep, aircraft noise was 1.3 times more likely to increase motility as at the start of the sleep period. Thus, the probability of motility was found to be greater at the end than at the beginning of the night. It is unknown whether this finding is related to sleep stage, and also why motility due to aircraft noise peaked at 46 years of age.

Factors that had no demonstrable effect on the instantaneous measures included the type of aircraft noise event (takeoff vs. landing), median sound level within the bedroom during sleep in the absence of aircraft noise, or the estimated equivalent indoor aircraft sound level from 2300 to 0700 h. Individual factors such as a subject's gender or attitude towards aircraft noise were not related to the extent of disturbance.

On a 24 h time scale, Passchier-Vermeer *et al.* (2002) observed a statistically significant increase in mean motility during sleep, number of BCA and number of recalled awakenings due to aircraft noise as a function of indoor equivalent aircraft sound level, and the number of aircraft during the sleep period time. Although statistically significant, the increase in BCAs and recalled awakenings was reported as being small. Mean motility over the night was higher: (1) when average noise within the bedroom *not* due to aircraft noise increased, (2) when the transmission loss from outdoors to indoors was low, (3) when subjects indicated a difficulty falling asleep due to aircraft noise, and (4) in subjects who attributed awakenings to aircraft noise exposure. Motility was 15% higher in subjects who recalled being awakened each night by aircraft noise compared to subjects that never indicated such an awakening. The study found that when aircraft noise was noted as the cause for difficulty in falling asleep (increasing sleep latency time), the delay to sleep onset was about 15 min.

While perceived sleep quality was reduced as mean motility increased, equivalent indoor aircraft sound levels and number of aircraft events were not statistically related to perceived sleep quality. Compared to mean motility and sleep latency time, perceived difficulty falling asleep had a stronger influence on perceived sleep quality, sleepiness during the waking state (i.e., fatigue), the number of subjectively recalled awakenings, and the number of BCA. Aircraft noise had a slight impact on self-reported sleepiness the following day at 1000 h, but not at any other time point. It should, however, be noted that 1000 h was the first sampling period and that it is entirely possible that results may have been different at earlier sampling times. Aircraft noise exposure at night did not have any statistically significant impact on the speed of responding as measured by the Wilkinson reaction time task. This simple cognitive task was conducted prior to retiring for bed and required the subject to button press as fast as possible following presentation of a visual stimulus.

Over the long term, it was found that when the average sound level within the bedroom over the 11 day study period due to aircraft noise increased, mean motility (as determined over the 11 day period) was also higher and sleep latency time was longer. Mean motility was also related to the frequency of recalled awakenings, BCA, the use of sleeping medication, self-reported sleep quality, the number of general sleep complaints, and the number of health complaints. The number of health complaints collected from the questionnaires at the end of the 11 day study period increased, regardless of age, from 2.5 to 4 on a scale from 0 to 13, if the average sound level in the bedroom due to aircraft noise increased from 5 to 35 dB during the sleep period over the 11 study days.

Difficulty falling asleep due to aircraft noise was reported in 12 of the 4600 subject nights (0.26%). On 21 occasions, subjects reported being awakened by aircraft noise at the end of the sleep period time (0.46%). During the nighttime, 159 (~2%) of reported awakenings were attributed to aircraft noise. Window position was altered 121 times during the nighttime and on 13 occasions (10.7%) the window was closed because of aircraft noise.

Subjects were asleep before 2300 h in about one-third of the study nights. However, there were only minor effects (~4%), from aircraft noise between 2300 and 2400 h, on

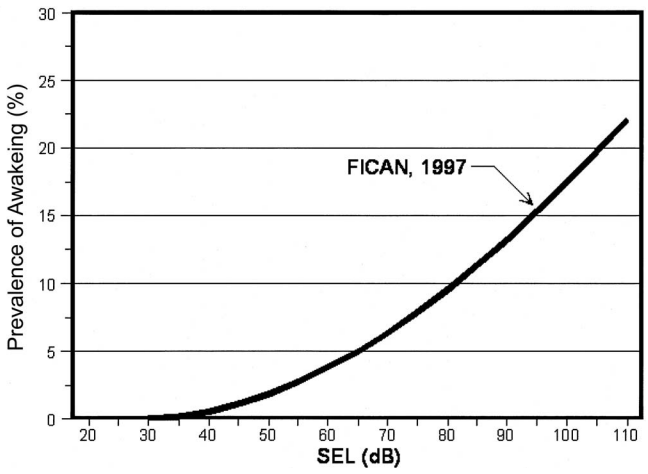


FIG. 1. Powell's analysis (FICAN, 1997) of upper limit of field observations of sleep disturbance as a function of indoor sound exposure levels.

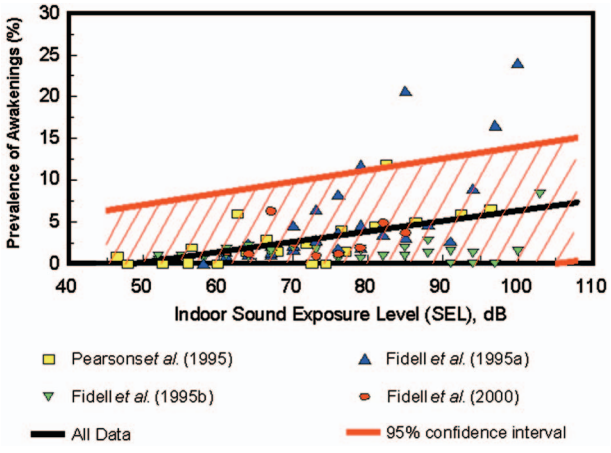


FIG. 2. ANSI S12.9-2000/Part 6 relationship between prevalence of awakening due to single aircraft noise intrusions and indoor sound exposure levels, with 95% confidence interval on prediction equation.

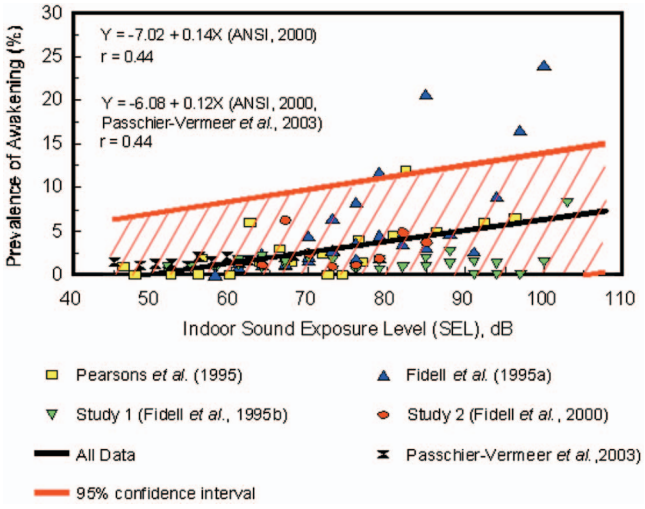


FIG. 3. ANSI S12.9-2000/Part 6 relationship and field data with observations of Passchier-Vermeer *et al.* (2003).

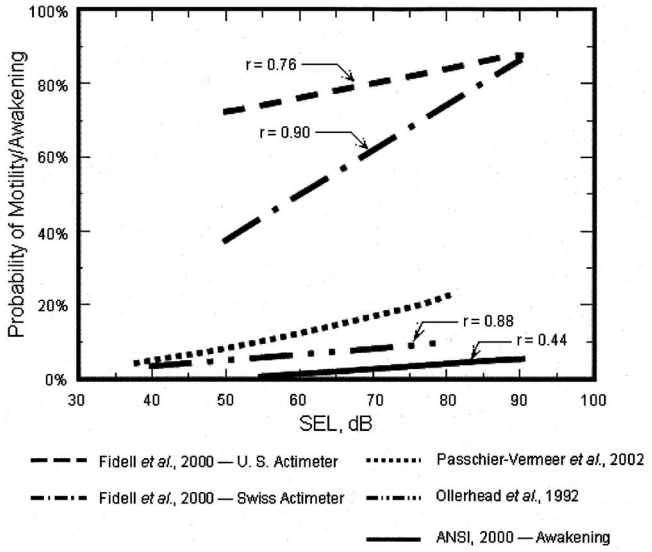


FIG. 4. Functions relating motility to indoor sound exposure levels inferred from four field studies, in comparison with ANSI S12.9-2000/Part 6 awakening relationship.

endpoints such as motility, BCAs, and number of recalled awakenings. Between 0600 and 0700 h, in about 50% of subject nights, participants were still sleeping; however, aircraft noise at this time contributed about 27% of all effects and 27% of the aircraft noise events occurred within this time period. There was no association between outdoor aircraft noise metrics and aircraft noise-induced increase in the probability of motility.

IV. SUMMARY RELATIONSHIPS

The U.S. Federal Interagency Committee on Noise (FICON) proposed an interim dose-response curve in 1992 to predict the percentage of people that might be awakened by noise based on findings arising from both laboratory and field studies. Field studies published since 1992 have suggested that the prevalence of awakening in familiar sleeping quarters is considerably smaller than observed in the laboratory and the initial curve proposed by FICON likely overestimated awakenings in exposed populations (Pearsons *et al.*, 1995).

Two relationships more recent than that of FICON summarized much of the behavioral awakening data described in this article. The first [shown in Eq. (7) and Fig. 1], developed by Powell but published by FICAN (1997), does not purport to be a dosage-effect relationship, but simply an upper limit on some of the behavioral awakening data. The FICAN relationship is not a formal policy position of the U.S. government, but a recommendation intended to protect the public from sleep disturbance in *any* degree.

Prevalence of Awakening due to individual aircraft

$$\text{noise intrusions} = 0.0087 \times (\text{SEL} - 30)^{1.79}, \qquad (7)$$

where applicable indoor SEL values are in the range of 40–110 dB.

The second of the post-FICON relationships, adopted by ANSI (2000), is the result of a regression analysis on a superset of the FICAN behavioral awakening data, as seen in Fig. 2 and in Eq. (8).

Prevalence of Awakening due to individual aircraft

$$\text{noise intrusions} = -7.02 + 0.14(\text{SEL}), \qquad (8)$$

where applicable indoor SEL values are in the range of 50–100 dB.

Neither the FICAN nor the ANSI relationship includes the findings of Passchier-Vermeer *et al.* (2002, 2003). Figure 3 revises the ANSI (2000) relationship by including the behavioral awakening data reported by Passchier-Vermeer *et al.* (2002, 2003). The regression on the combined ANSI and Passchier-Vermeer data differs little from that of the ANSI standard, as shown in Eq. (9). Neither of these relationships accounts for as much as 20% of the variance in the relationship between indoor SEL and the prevalence of awakenings.

Prevalence of Awakenings due to single aircraft

$$\text{noise intrusions} = -6.08 + 0.12(\text{SEL}), \qquad (9)$$

where applicable indoor SEL values are in the range of 45–110 dB.

Figure 4 shows linear regressions between motility and indoor SEL for four studies, including one in which two different actimeters were employed. The figure also includes the ANSI relationship between awakening and indoor SEL for purposes of comparison. The regression equations for the four relationships between motility and indoor SEL may be found in Table II.

Although the summary relationships for the motility findings from the four studies all suggest at least a superficially greater sensitivity to indoor SEL than the ANSI (behavioral awakening) relationship, the motility findings of the various studies do not agree well with one another. This lack of agreement is probably due in large part to differences in details of measurement, analysis and definitions of "motility" and "onset of motility" in the various studies.

TABLE II. Comparison of relationships inferred from five recent field studies of aircraft noise-induced sleep disturbance between indoor noise levels and arousal, motility, and awakening.

Reference	Prediction of arousal	Prediction of motility from noise events	Prediction of onset of motillity	Prediction of awaking
Passchier-Vermeer *et al.* (2002, 2003)	Not applicable	P(motility)=0.00063(L_{max}−32)+0.0000314(L_{max} 0.0000314)^2 P(motility)=0.000532(SEL −38)+0.0000268(SEL− 0.0000268)^2	P(onset of motility)= 0.000415(L_{max}−32)+ 0.00000884(L_{max}− 0.00000884)^2 P(onset of motility)= 0.000273(SEL−40)+ 0.00000357(SEL− 0.00000357)^2	Prevalence of event-related awakening (%) =0.51+ 0.000353(SEL)^2 (Note: Equation pertains to original data from Passchier-Vermeer *et al.* (2002) who recommend that prevalence of event-related awakening be reduced by 1.523% to obtain the prevalence of *aircraft* noise-induced, event- related awakening)
Ollerhead *et al.* (1992)	Not applicable	% of events leading to motility=−2.96+ 0.162(SEL)	Not applicable	Prevalence of event-related awakening (%) =0.4*[−2.96+ 0.162([SEL)]
Fidell *et al.* (1995a)	Not applicable	Not applicable	Not applicable	Prevalence of event-related awakening (%) =−10.24+0.167 (SEL)
Fidell *et al.* (1995b)	Prevalence of event-related arousal per Cole *et al.* (1992) (%)=1.306+ 0.279(SEL)	US actimeter: %of events leading to motility =47.16+0.4(SEL) Swiss actimeter: %of events leading to motility =−23.74+ 1.23(SEL)	Not applicable	Prevalence of event-related awakening (%) =−15.041+0.246(SEL)
Fidell *et al.* (2000) (incombination with Fidell, 1995a and b)	Prevalence of event-related arousal per Cole *et al.* (1992) (%) =4.579+ 0.218(SEL)	US actimeter: % of events leading to motility = 53.041+0.386(SEL)	Not applicable	Prevalence of event-related awakening (%) =−17.371+0.294(SEL)

V. CONCLUSIONS

Table II summarizes the major findings of the reviewed studies. Findings about noise-induced sleep disturbance differ considerably both with respect to the measure of sleep disturbance (motility or behavioral awakening), and by study. The findings of the reviewed studies are inconclusive about the effects of aircraft noise on changes in sleep states that do not result in awakenings. Neither behavioral awakening nor motility measurements are capable of detecting more subtle interference with sleep quality, such as brief changes in sleep stage or so-called "microarousals," that might also reflect a state of disrupted sleep.

On the other hand, the findings are in reasonable agreement with respect to several consequences of nighttime aircraft noise in habitually exposed populations: Spontaneous (that is, nonaircraft related) awakenings are more common than aircraft noise-induced awakenings in airport neighborhoods; a small percentage of habitually exposed people are actually awakened from sleep by aircraft noise intrusions; and although AN-ISD increases as a function of time in bed, this observation is confounded by the fact that sleep in general becomes more easily disturbed, such that noise intrusions in the latter part of the night are more likely to disturb sleep than in the earlier part of the night.

A. Noise measurements as predictors of sleep disturbance

The clearest and most consistent relationships between measurements of aircraft noise levels and sleep disturbance were observed between indoor sound levels and BCAs. Except in one case (Fidell *et al.*, 1995b), neither long nor short term measures of outdoor noise levels were reliably associated with any measure of sleep disturbance.

One of the more consistent findings from the reviewed field studies was that a single event noise metric such as SEL or L_{max} could be used to describe the effect of aircraft noise on sleep. However, summary relationships between indoor SEL and the prevalence of awakening due to single aircraft noise intrusions have shallow slopes, and do not account for more than about 20% of the variance in the relationship between sleep disturbance and acoustic measures of aircraft noise exposure. As such, single event metrics do not by

themselves provide robust guidance for regulatory purposes. Also, given the mediating factors for awakenings (e.g., time after sleep onset and ambient noise levels in sleeping quarters) the summary relationships between indoor SEL and awakenings should not be over-interpreted for other predictive purposes.

Regarding the effects of the totality of nighttime aircraft noise events on awakenings, Passchier-Vermeer *et al.* (2002) was the only reviewed study that showed a statistically significant relationship, albeit small, between subjective and objective measures of awakening and both the indoor equivalent sound level due to aircraft and the number of aircraft events.

Such practical limitations on interpretations of empirical findings about noise effects are not unique to aircraft noise and sleep disturbance. For example, Fidell and Silvati (2004) reviewed similar difficulties of interpretation of findings about noise-induced annoyance and complaints.

B. Motility as an index of sleep disturbance

Motility may be a more sensitive measure of sleep disturbance than behavioral awakening, in that motility can be associated with indoor SELs of noise intrusions at lower levels than can awakenings. This may be too simple a characterization of motility as an indicator of noise-induced sleep disturbance, however. Relationships between indoor sound exposure levels and motility inferred in the reviewed studies are inconsistent with one another. Also, clock time, time after sleep onset, and ambient noise levels in the sleeping quarters also mediate the likelihood of noise-induced motility.

Interpretations of motility findings are complicated by methodological differences in their measurement. For example, Passchier-Vermeer *et al.* (2002) monitored motility in 15 s epochs, while Ollerhead *et al.* (1992) and Fidell *et al.* (1995a, 1995b, 2000) assessed motility in 30 s epochs. Analyses conducted in epochs of greater duration may underestimate the effect of AN-ISD as measured by onset of motility, especially for aircraft noise intrusions of greater SEL (and hence, of potentially longer durations). In such cases, the onset of motility may occur in the interval prior to the L_{max} of a noise intrusion. Since the onset of motility might not be noted in a 30 s interval that includes the L_{max}, onset of motility might be attributed to nonaircraft noise events rather than to an aircraft noise event.

The convenience of onset of motility as a metric of sleep disturbance for regulatory purposes is further hindered by its complexity of interpretation. It is conceivable, for example, that focusing on the *onset* of motility among those not yet disturbed by ongoing noise intrusions of long duration could lead to a perverse focus on the effects of noise intrusions on hardier sleepers — those not already awakened or otherwise disturbed by intruding noises of shorter duration or lower sound level.

ACKNOWLEDGMENTS

The authors extend their appreciation to Ken Hume, John Bradley, Larry Finegold, and Willy Passchier-Vermeer for their feedback on earlier drafts. Portions of this review have been presented by the senior author in "Aircraft noise and sleep disturbance: A review of field studies," Journal of the Canadian Acoustical Association Proceedings, Ottawa, Canada, October, 2004.

Nomenclature

AAS = Amsterdam's Schiphol airport
AN-ISD = aircraft noise-induced sleep disturbance
BCA = behaviorally confirmed awakening
DEN = Stapleton international airport
DIA = Denver international airport
EEG = electroencephalographic
FICAN = U.S. federal interagency committee on aircraft noise
FICON = U.S. federal interagency committee on noise
LAX = Los Angeles international airport
PSG = polysomnography
REM = rapid eye movement
SEL = sound exposure level
SWS = slow wave sleep

[1]Indoor equivalent sound level of an aircraft noise event, normalized to 1 s, taken over the time interval that the aircraft sound level is greater than the indoor $L_{max}-10$ dB

American National Standards Institute (ANSI). (**2000**). "Quantities and procedures for description and measurement of environmental sound—Part 6: Methods for estimation of awakenings associated with aircraft noise events heard in homes," ANSI S12.9–2000/Part 6.

Ancoli-Israel, S., Cole, R., Alessi, C., Chambers, M., Moorcroft, W., and Pollak, C. P. (**2003**). "The role of actigraphy in the study of sleep and circadian rhythms," Sleep **26**, 342–392.

Anderson, G. S., and Miller, N. P. (**2005**). "A pragmatic re-analysis of sleep disturbance data," *Noise-Con 2005*, Minneapolis, MN.

Cole, R., Kripke, D. F., Gruen, W., Mullaney, D., and Gillin, J. C. (**1992**). "Automatic sleep/wake identification from wrist activity," Sleep (N.Y.) **15**, 461–469.

De Gennaro, L., Vecchio, F., Ferrara, M., Curcio, G., Rossini, P. M., and Babiloni, C. (**2005**). "Antero-posterior functional coupling at sleep onset: Changes as a function of increased sleep pressure," Brain Res. Bull. **65**, 133–140.

Dinges, D. F., Pack, F., Williams, K., Gillen, K. A., Powell, J. W., Ott, G. E., Aptowicz, C., and Pack, A. I. (**1997**). "Cumulative sleepiness, mood disturbance, and psychomotor vigilance performance decrements during a week of sleep restricted to 4–5 hours per night," Sleep **20**, 267–267.

Federal Interagency Committee on Aviation Noise (FICAN). (**1997**). "Effects of aviation noise on awakenings from sleep," http://www.fican.org/pages/sleepdst.html (accessed on 08/30/2005).

Fidell, S., Pearsons, K., Tabachnick, B. G., Howe, R., Silvati, L. (**1995a**). "Field study of noise-induced sleep disturbance," J. Acoust. Soc. Am. **98**, 1025–1033.

Fidell, S., Howe, R., Tabachnick, B. G., Pearsons, K., and Sneddon, M. (**1995b**). "Noise-induced sleep disturbance in residences near two civil airports," NASA Contractor Report No. 198252.

Fidell, S., Pearsons, K., Tabachnick, B. G., and Howe, R. (**2000**). "Effects on sleep disturbance of changes in aircraft noise near three airports," J. Acoust. Soc. Am. **107**, 2535–2547.

Fidell, S., and Silvati, L. (**2004**). "Parsimonious alternatives to regression analysis for characterizing prevalence rates of aircraft noise annoyance," Noise Control Eng. J. **52**, 56–68.

Finegold., L., and Elias, B. (**2002**). "A predictive model of noise induced awakenings from transportation noise sources," *Proceedings for the International Congress and Exposition on Noise Control Engineering*, N444.

Harrison, Y., and Home, J. A. (**2000**). "The impact of sleep deprivation on decision making: A review," J. Exp. Psychol., Appl. **6**, 236–249.

Health Canada. (**2001**). "Noise from civilian aircraft in the vicinity of airports-Implications for human health 1. Noise, stress and cardiovascular

disease," http://www.hc-sc.gc.ca/ewh-semt/pubs/noise-bruit/01hecs-secs256/index-e.html (accessed on 10/30/2005).
Horne, J., Pankhurst, F., Reyner, L., Hume, K., and Diamond, I. (**1994**). "A field study of sleep disturbance: Effects of aircraft noise and other factors on 5,742 nights of actimetrically monitored sleep in a large subject sample," Sleep **17**, 146–159.
Horonjeff, R., Fidell, S., Teffeteller, S., and Green, D. M. (**1982**). "Behavioral awakening as functions of duration and detectability of noise intrusions in the home," J. Sound Vib. **84**, 327–336.
Hume, K., Van, F., and Watson, A. (**2003**). "Effects of aircraft noise on sleep: EEG-based measurements," TSO, UK, available at: http://www.cate.mmu.ac.uk/documents/Publications/CAAreport0603.pdf (accessed 08/08/2006).
Michaud, D. S., Miller, S., Ferrarotto, C., Konkle, A. T. M., Keith, S. E., and Campbell, K. C. (**2005**). "Waking levels of salivary biomarkers are altered following sleep in a lab with no further increase associated with simulated nighttime noise exposure," Noise & Health **8**(29), 1–15.
Nordic Council of Ministers. (**1994**). (Hygge, S. ed.) "Health effects of community noise," Nordic Council of Ministers, Copenhagen.
Ollerhead, J. B., Jones, C. J., Cadoux, R. E., Woodley, A., Atkinson, B. J., Horne, J. A., Pankhurst, F., Reyner, L., Hume, K. I., Van, F., Watson, Al, Diamond, I. D., Egger, P., Holmes, D., and McKean, J. (**1992**). "Report of a field study of aircraft noise and sleep disturbance," Department of Safety, Environment and Engineering, Civil Aviation Authority, London.
Passchier-Vermeer, W., Vos, H., Steenbekkers, J., van der Ploeg, F., and Groothuis-Oudshoorn, K. (**2002**). "Sleep disturbance and aircraft noise exposure: Exposure-effect relationships," TNO Inro Report No. 2002.027, 1–245.
Passchier-Vermeer (**2003**). "Night-time noise events and awakening," TNO Inro Report No. 2003-32, 1–61.
Pearsons, K., Barber, D., Tabachnick, B. G., and Fidell, S. (**1995**). "Predicting noise-induced sleep disturbance," J. Acoust. Soc. Am. **97**, 331–338.
Pilcher, J. J., and Huffcutt, A. I. (**1996**). "Effects of sleep deprivation on performance: A meta-analysis," Sleep **19**, 318–326.
Rylander, R., Sorensen, S., and Berglund, K. (**1972**). "Sonic boom effects on sleep—a field experiment on military and civilian populations," J. Sound Vib. **24**, 41–50.
van Dongen, H. P., Maislin, G., Mullington, J. M., and Dinges, D. F. (**2003**). "The cumulative cost of additional wakefulness: Dose-response effects on neurobehavioral functions and sleep physiology from chronic sleep restriction and total sleep deprivation," Sleep **26**, 117–126.
Williamson, A. M., and Feyer, A. M. (**2000**). "Moderate sleep deprivation produces impairments in cognitive and motor performance equivalent to legally prescribed levels of alcohol intoxication," Occup. Environ. Med. **57**, 649–655.

LETTERS TO THE EDITOR

Linear and nonlinear measures of ocean acoustic environmental sensitivity (L)

Stan E. Dosso
School of Earth and Ocean Sciences, University of Victoria, Victoria, British Columbia V8W 3P6, Canada

Peter M. Giles and Gary H. Brooke
General Dynamics Canada, 3785 Richmond Road, Ottawa, Ontario K2H 5B7, Canada

Diana F. McCammon
McCammon Acoustical Consulting, 475 Baseline Rd., Waterville, Nova Scotia B0P 1V0, Canada

Sean Pecknold and Paul C. Hines
Defence Research and Development Canada-Atlantic, Dartmouth, Nova Scotia B2Y 3Z7, Canada

(Received 2 June 2006; revised 4 October 2006; accepted 5 October 2006)

This letter defines linear, linearized, and nonlinear measures of environmental sensitivity for ocean acoustic propagation that account for realistic uncertainties in various environmental parameters (water-column sound-speed profile and seabed geoacoustic properties). Simple interpretations of sensitivity are based on the implicit assumption of a linear relationship between parameter sensitivity and parameter uncertainty. This assumption is examined by comparing the three sensitivity measures over a range of parameter uncertainties about the actual assumed environmental uncertainty. Sensitivity range and depth dependencies are illustrated for realistic geoacoustic uncertainties and oceanographic variability of the sound-speed profile. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2382719]

PACS number(s): 43.30.Pc [AIT] Pages: 42–45

I. INTRODUCTION

It is well known that the various physical parameters of the ocean environment (water column and seabed) influence propagating acoustics fields to varying degrees. However, there has been relatively little work in defining practical, quantitative measures of environmental sensitivity.[1] Kessel[1] defined a sensitivity measure based on the relative change in magnitude of the acoustic pressure for a specific environmental change, spatially averaged over the water column to produce stable, representative values.

The goal of this work is to quantify and examine the effect on acoustic data of realistic variations in environmental parameters (e.g., due to errors in parameter estimates and spatial/temporal variability). This requires transferring uncertainties in environmental parameters through the forward problem (propagation modeling) and quantifying the resulting uncertainties in the data. In this approach, a parameter's sensitivity depends on the degree of uncertainty assigned to that parameter. Thus, a given parameter could be considered sensitive in one environmental setting where its uncertainty is large (and hence its potential variability can significantly affect acoustic fields), but insensitive in another setting where it is well known (and its small uncertainty has a correspondingly small potential effect on the fields).

Since the uncertainties of various environmental parameters are often only roughly known, an important component of this approach involves understanding how parameter sensitivities vary with parameter uncertainties. Simple interpretations of sensitivity implicitly assume a linear relationship: For Gaussian-distributed parameter uncertainties, the data uncertainty is also Gaussian distributed with a standard deviation that scales directly with the parameter standard deviation. However, for nonlinear problems these assumptions do not necessarily hold, and we are not aware of work to date examining linearity in ocean acoustic sensitivity studies. This letter defines and examines linear, linearized, and nonlinear sensitivity measures.

II. SENSITIVITY MEASURES

The interpretation of sensitivity is well defined for linear problems, which provide a basis for addressing nonlinear problems. For a linear problem, the change in datum d_i due a change δm_j to the jth model parameter m_j is given by

$$\delta d_i^j = d_i(m_j + \delta m_j) - d_i(m_j) \tag{1}$$

$$= \frac{\partial d_i}{\partial m_j} \delta m_j, \tag{2}$$

where the dependence on parameters that are unchanged is suppressed. The partial derivatives in Eq. (2) are sometimes themselves referred to as sensitivities. However, these derivatives do not account for the variation in parameter uncertainties. Equation (2) indicates how uncertainties transfer from parameters to data for a linear problem: If the uncertainty for parameter m_j is Gaussian distributed with standard deviation σ_j, then the corresponding uncertainty for datum d_i is also Gaussian with standard deviation given by $|\partial d_i / \partial m_j| \sigma_j$ (i.e., the derivative acts as a magnification factor). A linear sensitivity measure that is independent of data/parameter scales and units and accounts for parameter uncertainties is

$$S_{ij} = \frac{|\delta d_i^j(\delta m_j = \sigma_j)|}{|d_i|}. \tag{3}$$

For linear problems, the data perturbation δd_i^j in Eq. (3) can be computed using Eqs. (1) and (2). The advantage to using Eq. (2) is that, given the required partial derivatives, changes to the data can be computed for any parameter change without solving the forward problem; however, this can be offset by the difficulty in computing accurate derivatives. Equation (3) represents the ratio of the standard deviation of the ith datum (due to the uncertainty in the jth parameter) to its expected value and is equivalent to the coefficient of variation, a standard statistical measure used for comparing the variability of potentially disparate quantities. Further, linearity provides a simple understanding of how this sensitivity scales with the parameter uncertainty; e.g., if σ_j is doubled, S_{ij} doubles.

The above concepts can be extended to weakly nonlinear problems. Expanding the data functional for a parameter perturbation about the background model and neglecting second-order terms leads to an approximate local linear relationship between data and parameter perturbations

$$\delta d_i^j = d_i(m_j + \delta m_j) - d_i(m_j) \tag{4}$$

$$\approx \frac{\partial d_i(m_j)}{\partial m_j} \delta m_j. \tag{5}$$

It is possible to use Eq. (5) in a linearized sensitivity measure. Analytic approaches to computing partial derivatives have been derived for normal-mode propagation models;[2,3] for other propagation models derivatives can be estimated numerically. However, it is preferable to calculate δd_i^j according to Eq. (4), applying the forward model, as this makes no linear approximation and is applicable for any propagation model. For improved accuracy, a linearized sensitivity measure can be defined using a two-sided perturbation average

$$S_{ij} = \frac{1}{2}\left[\frac{|d_i(m_j + \sigma_j) - d_i(m_j)|}{|d_i|} + \frac{|d_i(m_j - \sigma_j) - d_i(m_j)|}{|d_i|} \right]. \tag{6}$$

For nonlinear problems, Gaussian-distributed model uncertainties do not necessarily imply Gaussian data uncertainties. Strong nonlinearity can lead to complicated data uncertainty distributions and preclude a simple linearized interpretation of sensitivity. To address nonlinearity explicitly, the full data uncertainty distribution can be sampled using a Monte Carlo approach to draw random model perturbations δm_j from a Gaussian distribution with zero mean and standard deviation σ_j, and computing the corresponding data perturbation δd_i^j for each sample. A nonlinear sensitivity measure is based on the root-mean-square perturbation

$$S_{ij} = \frac{[\langle |d_i(m_j + \delta m_j) - d_i(m_j)|^2 \rangle]^{1/2}}{|d_i|}, \tag{7}$$

where $\langle \cdot \rangle$ represents an ensemble average over the Monte Carlo samples. For a linear problem, the linear, linearized, and nonlinear sensitivity measures are identical (given sufficient Monte Carlo sampling).

For nonlinear problems it also does not necessarily follow that the sensitivity S_{ij} scales with σ_j in a simple way. The linearity of parameter sensitivities are examined here by comparing the linear, linearized, and nonlinear sensitivity measures over a range of parameter standard deviations σ_j about the actual environmental uncertainty. An approximately linear response is characterized by good agreement between the measures and by a linear increase in S_{ij} with σ_j. A useful measure for the latter criterion is the ratio of the sensitivity to the relative parameter uncertainty (referred to here as the relative sensitivity)

$$R_{ij} = \frac{S_{ij}}{\sigma_j / m_j}, \tag{8}$$

which is constant over σ_j for a linear problem.

Propagating acoustic fields are highly variable functions of source-receiver geometry, and hence the sensitivity at a point may not be representative of the overall sensitivity of the local field.[1] Effective sensitivity stabilization is achieved here by spatial averaging of the numerator and denominator of the sensitivity measures, as described in the following section.

III. EXAMPLES

This section provides examples of the sensitivity measures and linearity analysis outlined above. The environmental parameters and uncertainties considered are based on the Malta Plateau, a well-studied shallow-water region of the Mediterranean Sea. The environmental model, illustrated in Fig. 1, is comprised of a 131 m water column over-top a seabed consisting of three homogeneous layers. The ocean sound-speed profile (SSP), measured at the site, includes a strong negative gradient in the near-surface waters and a weak sound channel with its axis near mid-water depth. The three seabed layers are characterized by sound speeds v_1, v_2, v_3, densities ρ_1, ρ_2, ρ_3, and attenuation coefficients α_1, α_2, α_3. The upper two sediment layers are of thicknesses h_1 and h_2 (the basement layer is assumed to be semi-infinite). Geoa-

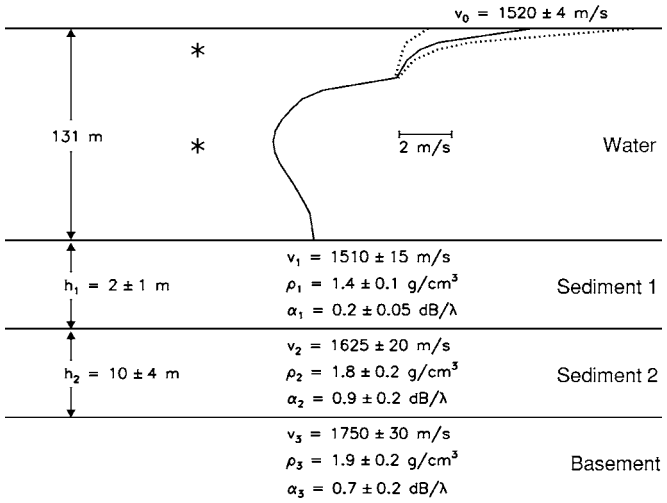


FIG. 1. Schematic diagram of the Malta Plateau environment, including assumed values and standard deviations for the water-column sound-speed profile and seabed geoacoustic parameters (see text). Asterisks indicate source depths considered in this paper.

coustic parameters values and uncertainties that are representative of existing knowledge of the Malta Plateau region are given in Fig. 1.

While it is straightforward to quantify sensitivity to individual points of the SSP, the goal of this work is to consider physically realistic environmental variability. Hence, the SSP uncertainty is taken to represent oceanographic variability due to surface heating/cooling and wind mixing, with the effects decaying exponentially with depth over the top 30 m, as shown in Fig. 1. This variability is represented in terms of the standard deviation of the surface sound speed v_0.

The linearity of a number of the most sensitive environmental parameters is investigated in Fig. 2, which compares the linear, linearized, and nonlinear sensitivity measures for a range of parameter standard deviation values about the actual environmental uncertainties (given in Fig. 1). In each case, the acoustic frequency is 1200 Hz, the source-receiver range is 10 km, and the receiver depth is 15 m. Source depths are chosen to illustrate different propagation conditions; for the geoacoustic parameters the source depth is 65 m (mid-water column), while for the SSP parameter v_0 the source depth is 15 m, within the variable near-surface layer (the sensitivities of all environmental parameters for both source depths are considered later). Propagation modeling was carried out using a normal-mode model which provides complex acoustic fields for multiple ranges and source/receiver depths in a single run, allowing for efficient spatial averaging. The results shown in all figures are averaged over ±100 m in range (five-point average) and ±5 m in source and receiver depths (three-point averages). Nonlinear sensitivities were computed using Monte Carlo samples of 300 random parameter perturbations drawn from Gaussian distributions with the standard deviations indicated on the parameter axes of Fig. 2. The distributions were truncated in cases where the random perturbations extended over unphysical values (e.g., negative layer thicknesses and attenuations); the effect of this on the results is negligible. Repetitions of the Monte Carlo sampling (with different random perturbations sequences)

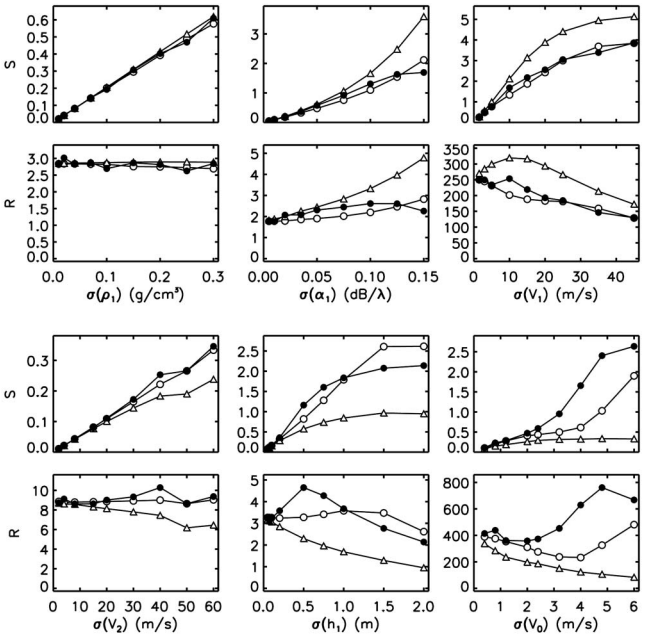


FIG. 2. Sensitivity measures S and R as a function of standard deviations for parameters ρ_1, α_1, v_1, v_2, h_1, and v_0. Nonlinear, linearized, and linear sensitivity measures are indicated by filled circles, open circles, and triangles, respectively. Range is 10 km; receiver depth is 15 m; source depth is 15 m for v_0, 65 m for other parameters.

yielded essentially identical results, indicating that the process converged to stable nonlinear sensitivity estimates.

Figure 2 indicates substantially different sensitivity behavior for different environmental parameters. For instance, the sensitivity to the upper sediment density ρ_1 behaves in a linear manner, with close agreement between the three sensitivity measures for all values of the parameter standard deviation $\sigma(\rho_1)$. Further, the sensitivity S increases linearly with $\sigma(\rho_1)$, illustrated further by the relative sensitivity R which is essentially constant over the range of $\sigma(\rho_1)$ values. In contrast, the sensitivity results for the remaining environmental parameters in Fig. 2 appear nonlinear to varying degrees, with significant differences between the three sensitivity measures and generally nonuniform rates of increase in sensitivity S with parameter standard deviations (also indicated by nonconstant R). The linearized sensitivity provides a better approximation to the nonlinear sensitivity than does the linear sensitivity for all parameters and all standard deviations. It is also worth noting that the magnitude of the sensitivity varies significantly with parameter.

Figure 2 indicates how parameter sensitivities vary with parameter standard deviations at a point. Figure 3 illustrates the nonlinear, linearized, and linear sensitivities for selected parameters as a function of range and depth, with the parameter standard deviation fixed at the assumed environmental uncertainties for the Malta Plateau region (Fig. 1). For the three geoacoustic parameters, the source is at 65 m depth near the sound-channel axis, while for the SSP parameter v_0, the source is at 15 m depth in the near-surface layer.

Considering first the nonlinear sensitivity measure (left column of Fig. 3), the sensitivities for geoacoustic parameters v_1, v_2, and h_1 are lowest at mid-water depths as a result of reduced bottom interaction due to sound-channel propaga-

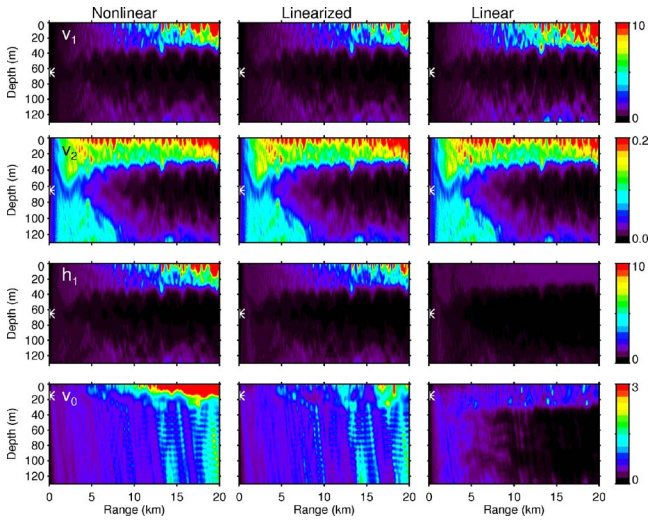


FIG. 3. (Color online) Range-depth sensitivity sections for parameters v_1, v_2, h_1, and v_0 (rows) and nonlinear, linearized, and linear sensitivity measures (columns). Source depth indicated by asterisks.

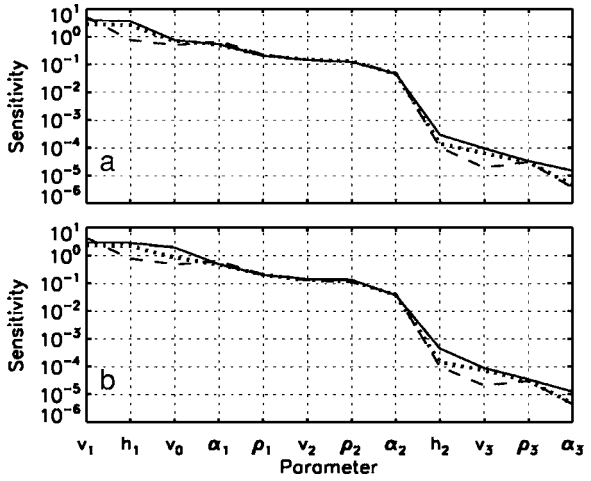


FIG. 4. Parameter sensitivities averaged over receiver depths of 0–30 m for source depths of (a) 65 m and (b) 15 m. Nonlinear, linearized, and linear sensitivity measures are indicated by solid, dotted, and dashed lines, respectively.

tion, while sensitivities are highest near the surface due to the strong bottom interaction resulting from the downward-refracting sound-speed gradient at this depth. Sensitivities increases with range as the effects of repeated bottom interactions accumulate. The sensitivity to the second sediment layer sound speed v_2 includes high values at short ranges, as steep propagation paths (which attenuate with range) interact significantly with the deeper layer. The highest sensitivity for the SSP parameter v_0 is confined to depths over which the SSP varies, with sensitivity increasing with range.

The agreement of the linearized and linear sensitivities with the nonlinear sensitivity in Fig. 3 is good for parameter v_1 and excellent for v_2, with generally better agreement for the linearized measure. The linearized sensitivity is in reasonable agreement with the nonlinear sensitivity for parameter h_1, but in poor agreement for v_0. The linear sensitivity differs substantially from the nonlinear sensitivity for h_1 and v_0, consistent with the single-point results of Fig. 2.

Figure 3 illustrates the detailed range-depth dependence of environmental parameter sensitivities for the Malta Plateau model. In some applications it may be desirable to consider the sensitivity more broadly, e.g., to quantify the overall environmental sensitivity of acoustic fields over some depth-range region. This can be accomplished in a straightforward manner by averaging sensitivities over the desired region (distinct from spatial averaging to stabilize sensitivities). For example, Fig. 4 shows the three sensitivity measures for all environmental parameters averaged over 0–30 m receiver depth and 0–20 km range for source depths of 15 and 65 m. Sensitivities for the different parameters vary over almost six orders of magnitude, with the most sensitive parameters being those of the upper seabed layer and SSP, followed by the second and then third seabed layers. The sensitivity results for the two source depths in Fig. 4 are generally similar, with the largest difference for the SSP parameter v_0. The linearized and linear sensitivities agree with the nonlinear sensitivity to within an order of magnitude for all parameters, although significant differences occur for the highest and lowest sensitivity parameters. (Differences for the low sensitivity parameters may be due to numerical precision limitations.) In all cases the linearized estimate provides a better approximation to the nonlinear sensitivity measure. Figure 4 indicates that, in the present case, variations in the third-layer parameters within their assumed uncertainties are essentially irrelevant to ocean acoustic propagation, and variations in the second-layer parameters are, at most, only marginally relevant.

IV. SUMMARY

The problem of quantifying ocean acoustic environmental sensitivity in a meaningful manner may be considered that of transferring uncertainties in environmental parameters to uncertainties in the acoustic responses (data), and quantifying the results. Monte Carlo sampling provides a general approach to transferring uncertainties, and formed the basis for a nonlinear sensitivity measure. More efficient, but approximate, linear and linearized sensitivities were also defined based on one-sided and two-sided parameter perturbations equivalent to the assumed parameter uncertainty (standard deviation). The three sensitivity measures were compared for a realistic environmental model and parameter uncertainties representing the Malta Plateau. The three measures agreed to within an order of magnitude for all parameters, with the linearized measure representing a better approximation to the nonlinear sensitivity. However, significantly nonlinear behavior was observed for several of the most sensitive parameters.

[1]R. T. Kessel, "A mode-based measure of field sensitivity to geoacoustic parameters in weakly range-dependent environments," J. Acoust. Soc. Am. **105**, 122–129 (1999).
[2]P. Gerstoft, "Inversion of acoustic data using a combination of genetic algorithms and Gauss-Newton approach," J. Acoust. Soc. Am. **97**, 2181–2190 (1995).
[3]R. T. Kessel, "The variation of modal wave numbers with geoacoustic parameters in layered media," J. Acoust. Soc. Am. **102**, 2690–2696 (1997).

Convex optimization based time-domain broadband beamforming with sidelobe control (L)

Shefeng Yan,[a] Chaohuan Hou, and Xiaochuan Ma
Institute of Acoustics, Chinese Academy of Sciences, 100080, Beijing, People's Republic of China

Yuanliang Ma
Institute of Acoustic Engineering, Northwestern Polytechnical University, 710072, Xi'an, People's Republic of China

(Received 3 June 2006; revised 15 October 2006; accepted 22 October 2006)

An approach to time-domain broadband beamforming with sidelobe control is proposed. The array response is expressed as a linear function of the finite impulse response filters tap weights. The filters are designed by minimizing beamformer output power while maintaining the distortionless response in the direction of the desired signal and guaranteeing the sidelobes to be below some given threshold values. Norm constraint on tap weights is also imposed to improve the robustness of the beamformer. Computationally efficient convex formulation for the beamformer design problem is derived using second-order cone programming. Simulation results demonstrate the satisfactory performance of the proposed approach. © *2007 Acoustical Society of America.*
[DOI: 10.1121/1.2400622]

PACS number(s): 43.60.Fg, 43.30.Wi [EJS] Pages: 46–49

I. INTRODUCTION

Array signal processing has wide applications in sonar, radar, wireless communications, etc. Broadband beamforming is one of the most important tasks in array signal processing.[1,2] Broadband beamforming includes discrete Fourier transform implementation of a frequency-domain beamformer and finite impulse response (FIR) implementation of a time-domain beamformer. The frequency-domain implementation is not suitable for some applications due to its associated delay. A more important advantage of the time-domain implementation is that we can update the beamformer when each new snapshot arrives.

The Frost processor[3] and the generalized sidelobe canceller beamformer[4] are two types of traditional broadband FIR beamformers. However, the sidelobes of these beamformers are not controlled, and this can lead to severe performance degradation in the case of unexpected (i.e., suddenly appearing) interferers. Godara[5] has also developed frequency-domain techniques for computing the tap weights in the FIR beamformer. However, the algorithm may lead to large number of taps, and hence high computational cost, since it requires that the number of taps should be equal to the number of frequency bins. Furthermore, the sidelobes cannot be controlled strictly.

Second-order cone programming (SOCP) has been used to develop narrowband adaptive beamforming algorithms with sidelobe control by Liu.[6] In this letter, we develop the narrowband adaptive beamforming to a broadband one. A convex optimization based approach to designing broadband FIR beamformer with controlled sidelobes is proposed.

[a]Author to whom correspondance should be addressed. Electronic mail: sfyan@ieee.org

II. BACKGROUND

Consider an M-element array with a known arbi trary geometry. Assume that $D+1<M$ broadband plane waves impinge on the array from directions $\Theta=[\theta_0,\theta_1,\ldots,\theta_d,\ldots,\theta_D]$. The time series received at the mth element can be modeled as

$$\tilde{x}_m(t)=\beta s_0[t-\tau_m(\theta_0)]+\sum_{d=1}^{D} s_d[t-\tau_m(\theta_d)]+\tilde{\nu}_m(t),$$
$$m=1,\ldots,M, \tag{1}$$

where $\{s_d(t)\}_{d=0}^{D}$ are the $D+1$ source signals as observed at an arbitrarily chosen reference point, $\tau_m(\theta_d)$ is the propagation delay between the mth sensor and the reference point associated with the dth source, and $\tilde{\nu}_m(t)$ is the additive noise at the mth sensor. We assume that the first term in Eq. (1) corresponds to the signal of interest (SOI) and that of the second term to D interferences. β is a binary parameter indicating whether the SOI is present in the training snapshots.

Figure 1 illustrates the conventional Frost beamforming structure consisting of M sensors and L taps per sensor. T_s is the time delay between any two adjacent taps. A group of pre-delays T_m, $m=1,\ldots,M$ are attached after the sensors. The pre-steering delay is used to align the signals to the look direction, i.e., $T_m=-\tau_m(\theta_0)$. Here, the look direction is assumed to be exactly equal to θ_0. Since the pre-delays are usually not an integer, mechanical or electronic pre-steering is applied.

The pre-delayed sampled data from the mth sensor is $x_m(n)=\tilde{x}_m(t-T_m)|_{t=nT_s}$. The input to the other tap weights is given by

$$x_{m\ell}(n)=x_m[n-(\ell-1)]=\tilde{x}_m[t-T_m-(\ell-1)T_s]\big|_{t=nT_s},$$

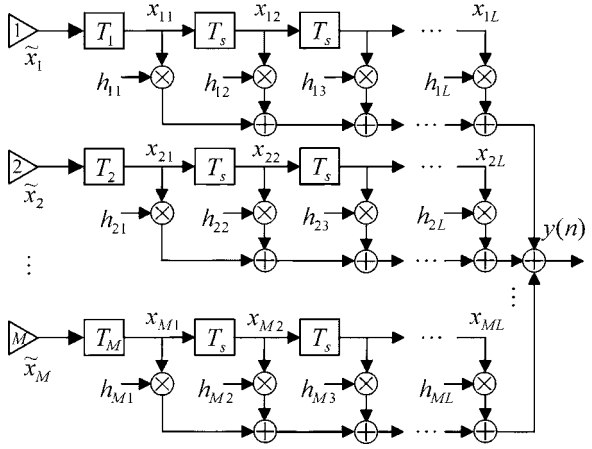


FIG. 1. Conventional Frost beamformer.

$$m = 1, \ldots, M, \quad \ell = 1, \ldots, L. \tag{2}$$

All the data samples are weighted and summed together to obtain the desired output signal. Let $h_{m\ell}$ be the adjustable weight of the ℓth tap following behind the mth sensor. We assume that the weights are complex. The time series of the beamformer output is given by

$$y(n) = \sum_{m=1}^{M} \sum_{\ell=1}^{L} h_{m\ell} x_m[n-(\ell-1)] = \sum_{m=1}^{M} \sum_{\ell=1}^{L} h_{m\ell} x_{m\ell}(n). \tag{3}$$

We can define two $M \times L$ matrices

$$\mathbf{X}(n) = \begin{bmatrix} x_{11}(n) & x_{12}(n) & \cdots & x_{1L}(n) \\ x_{21}(n) & x_{22}(n) & \cdots & x_{2L}(n) \\ \vdots & \vdots & \ddots & \vdots \\ x_{M1}(n) & x_{M2}(n) & \cdots & x_{ML}(n) \end{bmatrix}$$

and

$$\mathbf{H} = \begin{bmatrix} h_{11} & h_{12} & \cdots & h_{1L} \\ h_{21} & h_{22} & \cdots & h_{2L} \\ \vdots & \vdots & \ddots & \vdots \\ h_{M1} & h_{M2} & \cdots & h_{ML} \end{bmatrix}.$$

Let us introduce the notations

$$\mathbf{x}(n) = \mathrm{vec}\{\mathbf{X}(n)\}, \tag{4}$$

$$\mathbf{h} = \mathrm{vec}\{\mathbf{H}\}, \tag{5}$$

where vec$\{\cdot\}$ denotes the vectorization operator which stacks the columns of a matrix on top of each other, and $\mathbf{x}(n)$ represents the nth snapshot. Then Eq. (3) can be written as

$$y(n) = \mathbf{h}^T \mathbf{x}(n), \tag{6}$$

where $(\cdot)^T$ denotes the transpose.

III. PROPOSED BEAMFORMER

Let $\mathbf{h}_m = [h_{m1}, h_{m2}, \ldots, h_{mL}]^T$. The complex frequency response of the FIR filter is given by

$$H_m(f) = \sum_{\ell=1}^{L} h_{m\ell} \mathrm{e}^{-j(\ell-1)2\pi f T_s} = \mathbf{e}^T(f)\mathbf{h}_m, \tag{7}$$

where $\mathbf{e}(f) = [1, \mathrm{e}^{-j2\pi f T_s}, \ldots, \mathrm{e}^{-j(L-1)2\pi f T_s}]^T$.

Let $\kappa_m = \mathrm{e}^{-j2\pi f T_m}$, the actual weight corresponding to the mth sensor at frequency f is

$$w_m(f) = H_m(f)\kappa_m = \mathbf{e}^T(f)\mathbf{h}_m\kappa_m, \quad m = 1,2, \ldots, M. \tag{8}$$

Thus, the weight vector employed at frequency f is

$$\begin{aligned} \mathbf{w}(f) &= [w_1(f), w_2(f), \ldots, w_M(f)]^T \\ &= [\mathbf{e}^T(f)\mathbf{h}_1\kappa_1, \mathbf{e}^T(f)\mathbf{h}_2\kappa_2, \ldots, \mathbf{e}^T(f)\mathbf{h}_M\kappa_M]^T \\ &= \{[\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_M]^T \mathbf{e}(f)\} \circ \boldsymbol{\kappa} = [\mathbf{H}\mathbf{e}(f)] \circ \boldsymbol{\kappa}, \end{aligned} \tag{9}$$

where $\circ$ denotes the Hadamard (i.e., element-wise) product of two vectors, and $\boldsymbol{\kappa} = [\kappa_1, \kappa_m, \ldots, \kappa_M]^T$.

The array response (frequency-wavenumber response) of the beamformer can be written as follows:

$$p(f,\theta) = \mathbf{w}^T(f)\mathbf{a}(f,\theta) = \mathbf{a}^T(f,\theta)\mathbf{w}(f), \tag{10}$$

where $\mathbf{a}(f,\theta)$ is the array manifold vector that is given by

$$\mathbf{a}(f,\theta) = [\mathrm{e}^{-j2\pi f \tau_1(\theta)}, \cdots, \mathrm{e}^{-j2\pi f \tau_M(\theta)}]^T. \tag{11}$$

Using Eqs. (9) and (11) in Eq. (10) gives

$$\begin{aligned} p(f,\theta) &= [\mathbf{a}(f,\theta) \circ \boldsymbol{\kappa}]^T \mathbf{H}\mathbf{e}(f) \\ &= \{\mathbf{e}(f) \otimes [\mathbf{a}(f,\theta) \circ \boldsymbol{\kappa}]\}^T \mathbf{h} \\ &= \mathbf{u}^T(f,\theta)\mathbf{h} = \mathbf{h}^T\mathbf{u}(f,\theta), \end{aligned} \tag{12}$$

where $\otimes$ denotes the Kronecker product and $\mathbf{u}(f,\theta) = \mathbf{e}(f) \otimes [\mathbf{a}(f,\theta) \circ \boldsymbol{\kappa}]$. We define $\mathbf{u}(f,\theta)$ as the array manifold vector corresponding to the stacked vector $\mathbf{x}$. Its $m+(\ell-1)M$ element is

$$[\mathbf{u}(f,\theta)]_{m+(\ell-1)M} = \mathrm{e}^{-j2\pi f[T_m + \tau_m(\theta) + (\ell-1)T_s]} = \mathrm{e}^{-j2\pi f \xi_{m\ell}}, \tag{13}$$

where $\xi_{m\ell} = T_m + \tau_m(\theta) + (\ell-1)T_s$.

Let $\mathbf{R}$ denote the $ML \times ML$ theoretical covariance matrix of the stacked array output vector. We assume that $\mathbf{R}$ has the following form:

$$\mathbf{R} = E[\mathbf{x}(n)\mathbf{x}^H(n)] = \sum_{d=0}^{D} \mathbf{R}_d + \mathbf{R}_\nu, \tag{14}$$

where $(\cdot)^H$ denotes the Hermitian transpose, $\mathbf{R}_d$ is the broadband signal covariance matrix corresponding to the dth signal and $\mathbf{R}_\nu$ is the noise covariance matrix.

Consider a single plane-wave signal, say the dth source, the covariance matrix $\mathbf{R}_d$ is given by

$$\mathbf{R}_d = \int_{f_L}^{f_U} S_d(f)\mathbf{u}(f,\theta)\mathbf{u}^H(f,\theta)df, \tag{15}$$

where $S_d(f)$ is the source spectrum of the dth signal, which occupy the frequency band $[f_L, f_U]$. We assume that the signal has a flat spectrum over the frequency band with total power σ_s^2. Then,

$$\mathbf{R}_d = \frac{\sigma_s^2}{f_U - f_L} \int_{f_L}^{f_U} \mathbf{u}(f,\theta)\mathbf{u}^H(f,\theta)df. \tag{16}$$

Using Eq. (13), the $m+(\ell-1)M$, $p+(q-1)M$ element (here $p=1,2,\ldots,M$, $q=1,2,\ldots,L$) is

$$[\mathbf{R}_d]_{m+(\ell-1)M,p+(q-1)M} = \frac{\sigma_s^2}{f_U - f_L}\int_{f_L}^{f_U} \mathrm{e}^{-j2\pi f(\xi_{m\ell}-\xi_{pq})}df$$
$$= \frac{\sigma_s^2}{B}\,\mathrm{sinc}[\pi B(\xi_{m\ell}-\xi_{pq})]$$
$$\times \mathrm{e}^{-j2\pi f_c(\xi_{m\ell}-\xi_{pq})}, \quad (17)$$

where $B=f_U-f_L$ and $f_c=(f_U+f_L)/2$.

Assume that the noise has a flat spectrum over the same frequency band with total power σ_n^2, and is uncorrelated from sensor to sensor. The $m+(\ell-1)M$, $p+(q-1)M$ element of noise covariance matrix $\mathbf{R}_\nu$ is given by

$$[\mathbf{R}_\upsilon]_{m+(l-1)M,p+(q-1)M} = \frac{\sigma_n^2}{B}\delta_{mp}\int_{f_L}^{f_U} \mathrm{e}^{-j2\pi f(l-q)T_s}df$$
$$= \frac{\sigma_n^2}{B}\delta_{mp}\mathrm{sinc}[\pi B(l-q)T_s]$$
$$\times \mathrm{e}^{-j2\pi f_c(l-q)T_s}, \quad (18)$$

where δ_{mp} is the Kronecker delta function.

The theoretical data covariance matrix can be obtained by substituting Eqs. (17) and (18) into Eq. (14).

Note that the SOI, interferers, and the additive noise need not have flat spectra over the band of interest. In fact, our beamformer is a data dependent one and we do not require any *priori* information about their spectral densities. In practical applications, the covariance matrix $\mathbf{R}$ is unavailable. Therefore, the sample covariance matrix

$$\hat{\mathbf{R}} = \frac{1}{N}\sum_{n=1}^{N}\mathbf{x}(n)\mathbf{x}^H(n) \quad (19)$$

is used instead of Eq. (14). Here N is the number of snapshots.

The beamformer output power is given by

$$P_o = E[y(n)y^H(n)] = E[\mathbf{h}^T\mathbf{x}(n)\mathbf{x}^H(n)\mathbf{h}^*] = \mathbf{h}^T\mathbf{R}\mathbf{h}^*, \quad (20)$$

where $(\cdot)^*$ denotes the conjugate.

For our FIR beamformer, the output power is minimized while the SOI is maintained to be undistorted and the sidelobes are controlled. In order to improve its robustness, white noise gain constraint is applied to limit the norm of FIR filter tap weights to some known constant $\Delta>0$.

$$\sum_{m=1}^{M}\sum_{\ell=1}^{L}\|\mathbf{h}_{m\ell}\|^2 \leqslant \Delta^2 \quad \text{or } \|\mathbf{h}\| \leqslant \Delta. \quad (21)$$

Consequently, the beamforming problem is formulated as follows:

$$\min_{\mathbf{h}} \quad \mathbf{h}^T\mathbf{R}\mathbf{h}^*.$$

$$\text{subject to } |\mathbf{u}^T(f,\theta_0)\mathbf{h}| = 1, \quad \forall f \in [f_L,f_U],$$

$$|\mathbf{u}^T(f,\theta)\mathbf{h}| \leqslant \varepsilon, \quad \forall\, \theta \in \Theta_{\mathrm{SL}}, \quad \forall f \in [f_L,f_U],$$

$$\|\mathbf{h}\| \leqslant \Delta, \quad (22)$$

where ε controls the sidelobes, and Θ_{SL} is the sidelobe regions. The sidelobe regions at different frequency can be different.

Note that our beamformer does not require that $T_m=-\tau_m(\theta_0)$. Therefore, we can let $T_m=\mathrm{int}[-\tau_m(\theta_0)]$, where int$[\cdot]$ denotes round towards nearest integer. Consequently, the pre-delay can be implemented by a set of tapped delay lines, which is more convenient than the Frost processor.

We can discretize the frequency band $[f_L,f_U]$ using a finite number of frequencies $f_k \in [f_L,f_U]$, $k=1,\ldots,K$. The frequencies are often uniformly spaced and the narrowband condition should be satisfied within each frequency bin:

$$B_s \cdot \Delta T_{\max} \ll 1, \quad (23)$$

where B_s is the bandwidth of each frequency bin and $\Delta T_{\max}$ is the maximum travel time between any two elements in the array.

Let $\mathbf{R}=\mathbf{U}^H\mathbf{U}$ be the Cholesky factorization. We obtain

$$\mathbf{h}^T\mathbf{R}\mathbf{h}^* = (\mathbf{U}^*\mathbf{h})^T(\mathbf{U}^*\mathbf{h})^* = \|\mathbf{U}^*\mathbf{h}\|^2. \quad (24)$$

Noting that minimizing $\|\mathbf{U}^*\mathbf{h}\|^2$ is equivalent to minimizing $\|\mathbf{U}^*\mathbf{h}\|$ and introducing a new auxiliary variable γ, we can rewrite Eq. (22) as

$$\min_{\mathbf{h}} \gamma, \text{ subject to } \|\mathbf{U}^*\mathbf{h}\| \leqslant \gamma,$$

$$|\mathbf{u}^T(f_k,\theta_0)\mathbf{h}| = 1, \quad f_k \in [f_L,f_U],\ k=1,\ldots,K,$$

$$|\mathbf{u}^T(f_k,\theta_{i,k})\mathbf{h}| \leqslant \varepsilon, \quad \theta_{i,k} \in \Theta_{\mathrm{SL},k},\ i=1,\ldots,I_k,$$

$$f_k \in [f_L,f_U],\ k=1,\ldots,K,$$

$$\|\mathbf{h}\| \leqslant \Delta, \quad (25)$$

where $\Theta_{\mathrm{SL},k}$, $k=1,\ldots,K$ is the sidelobe regions at frequency bin f_k. $\theta_{i,k} \in \Theta_{SL,k}$, $i=1,\ldots,I_k$ is the angular grid chosen that properly approximates the sidelobe region $\Theta_{\mathrm{SL},k}$ by a finite number of directions. The choice of I_k is determined by the required accuracy of approximation.

The optimization problems (25) can be formulated as the convex SOCP problem, which can be solved using a SOCP solver (e.g., SeDuMi,[7] also refer to the Ref. 6). A review of the applications of SOCP has been presented by Lobo.[8]

The output signal-to-interference-plus-noise ratio (SINR) is

$$\mathrm{SINR} = \frac{\mathbf{h}^T\mathbf{R}_0\mathbf{h}^*}{\mathbf{h}^T\left(\sum_{d=1}^{D}\mathbf{R}_d + \mathbf{R}_\nu\right)\mathbf{h}^*}. \quad (26)$$

IV. SIMULATION RESULTS

Consider a linear array of $M=12$ uniformly spaced elements, with a half-wavelength spacing at the upper frequency f_U. The normalized design band is $[f_L,f_U]=[0.16,0.32]$ (The normalized sampling frequency was 1). In this simulation, the design band $[f_L,f_U]$ is decomposed into $K=33$ uniform frequency bins. Unless otherwise stated, we

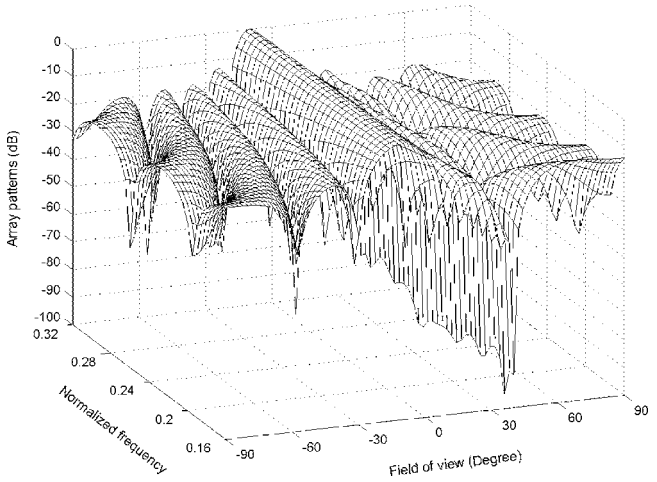


FIG. 2. Array patterns of the FIR beamformer without sidelobe control.

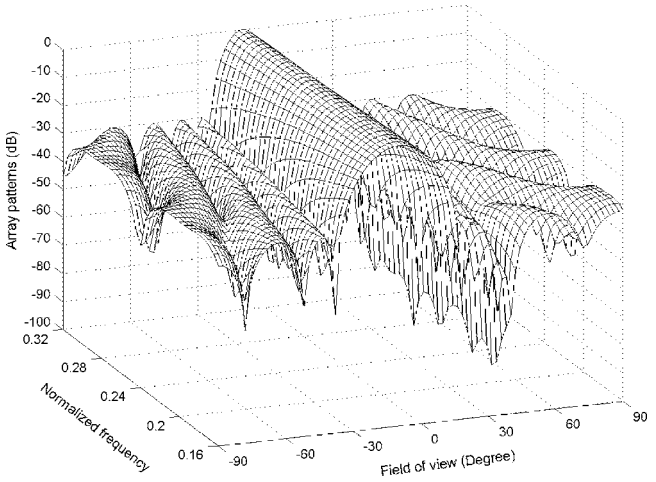


FIG. 3. Array patterns of the FIR beamformer with sidelobe control.

use the beamforming methods with the theoretical array covariance matrix **R** for convenience. The direction of the SOI is $\theta_0=-10°$ with respect to the broadside of the array. The signal-to-noise ratio is equal to 0 dB. A single interferer is assumed to impinge on the array from the direction of 40° with an interference-to-noise ratio of 30 dB in each sensor. It is assumed that the length of the FIR filters is $L=8$, and the white noise constraint value is $\Delta=1$. A uniform grid of 2° is used to discretize the field of view $[-90°,90°]$.

In the first example, we use a broadband FIR beamforming method without sidelobe control, i.e., solving an optimization problem similar to Eq. (22) except the sidelobe constraints. Using Eq. (12) by replacing **h** with the obtained FIR filters, the array responses as a function of frequency and the angle are calculated on the chosen grid of points. The obtained FIR broadband array patterns are shown in Fig. 2. It is seen that the array patterns have their mainlobes in the SOI direction and deep nulls in the direction of interferer. However, the sidelobe level is about −12.5 dB, which is a little higher. From Eq. (26), the output SINR is 10.7 dB.

In the second example, we use the proposed broadband FIR beamforming method with sidelobe control. Assume that the beamwidth decreases as the frequency increases. The sidelobe regions are $[-90°,-35°]\cup[15°,90°]$ and $[-90°,-25°]\cup[5°,90°]$ at the frequency f_L and f_U, respectively. It is assumed that $\varepsilon=0.001$.

The SOCP problem (25) is solved to obtain the FIR filters. The array responses as a function of frequency and the angle are calculated and shown in Fig. 3. It is seen that the array patterns also have their mainlobes and nulls in the direction of SOI and interferer, respectively. More importantly, the obtained sidelobes of our approach are strictly guaranteed to be below −30 dB.

It is also seen that the main lobe becomes a little wider. The output SINR is 10.1 dB, which indicates a 0.6 dB degradation over that without sidelobe control. However, these costs are acceptable in practical applications. The reason for degradation is that the performance parameters for a beamformer, i.e., the beamwidth, sidelobe level and the output SINR, are mutually correlated. Our algorithm provides a suitable compromise among these conflicting objectives.

V. CONCLUSION

An approach to broadband FIR beamforming with sidelobe control has been proposed. Our approach minimizes the beamformer output power, while maintaining distortionless response in the direction of signal and providing that the array pattern sidelobes are below some given threshold value. Convex SOCP-based formulation of our beamformer is derived. Results of computer simulations show that the proposed approach can guarantee that the sidelobes are strictly below the threshold. Furthermore, our FIR beamformer has integral tapped pre-delay, in contrast to the Frost beamfomer, where expensive mechanical or electronic presteering is utilized since its pre-delays are not integers. For simplicity, plane wave approximation is used is this letter. For the case of multipath that occurs with vertical arrays and horizontal arrays near endfire in ocean waveguides, our approach also works if we replace the array manifold vectors with the matched-field replica vectors.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China under Grant No. 60602055.

[1]D. B. Ward, R. A. Kennedy, and R. C. Williamson, "Theory and design of broadband sensor arrays with frequency invariant far-field beam patterns," J. Acoust. Soc. Am. **97**, 1023–1034 (1995).
[2]Z. S. Wang, J. Li, P. Stoica, T. Nishida, and M. Sheplak, "Constant-beamwidth and constant-powerwidth wideband robust Capon beamformers for acoustic imaging," J. Acoust. Soc. Am. **116**, 1621–1631 (2004).
[3]O. L. Frost III, "An algorithm for linearly constrained adaptive array processing," Proc. IEEE **60**, 926–935 (1972).
[4]L. J. Griffiths and C. W. Jim, "An alternative approach to linearly constrained adaptive beamforming," IEEE Trans. Antennas Propag. **30**, 27–34 (1982).
[5]L. C. Godara, "Application of the fast Fourier transform to broadband beamforming," J. Acoust. Soc. Am. **98**, 230–240 (1995).
[6]J. Liu, A. B. Gershman, Z. Q. Luo, and K. M. Wong, "Adaptive beamforming with sidelobe control: A second-order cone programming approach," IEEE Signal Process. Lett. **10**, 331–334 (2003).
[7]J. F. Sturm, "Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones," Optim. Methods Software **11-12**, 625–653 (1999).
[8]M. Lobo, L. Vandenberghe, S. Boyd, and H. Hebret, "Applications of second-order cone programming," Linear Algebr. Appl. **284**, 193–228 (1998).

Ultrasonic characterization of microstructure evolution during processing

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

Zheng Li
Department of Mechanics and Engineering Science, Peking University, Beijing 100871, China

(Received 12 June 2006; revised 27 September 2006; accepted 5 October 2006)

Many cold-working processes for polycrystalline metals cause alignment of the grains with a single symmetry axis. This type of microstructure is called fiber texture. The existence of a preferred orientation of the grains has a significant influence on the propagation and scattering of ultrasonic waves, which are often used for material inspection. Knowledge of the wave attenuation of such textured materials is of both theoretical and practical interest to nondestructive testing and materials characterization. In this article, the quantitative relations between fiber texture and wave attenuations of hexagonal crystals are presented. The texture is characterized by a Gaussian distribution function that contains a single parameter that governs the transition of the texture from perfectly aligned crystals to statistically isotropic. Under this assumption, the materials of interest have a varying degree of transverse isotropy representative of processing conditions. Simple expressions for the attenuations of the three modes of waves are given in a concise, generalized representation. Finally, numerical results are presented and discussed in terms of the directional, frequency, and texture dependence. The results presented are expected to improve the understanding of the microstructure evolution during thermomechanical processing. [DOI: 10.1121/1.2382749]

PACS number(s): 43.20.Bi, 43.20.Gp, 43.35.Cg [TDM] Pages: 50–59

I. INTRODUCTION

The macroscopic anisotropy of the physical properties of polycrystalline materials is generally determined by two factors, the anisotropy of the crystals, which is described by the single-crystal elastic constants, and the distribution of the crystals in space, which is called the texture. The simplest anisotropic symmetry class to be considered is that of medium with a single symmetry axis. The direction of the single symmetry axis is defined as the fiber direction. In this case, the fiber direction is not necessarily perpendicular to the sample surface, but typically is associated with processing conditions (e.g., rolling direction). The knowledge that the material microstructure directly affects the macroscopic material properties was a turning point in the field of materials manufacturing. Specific types of manufacturing processes are used to produce the corresponding microstructure in a controlled fashion. Many heat treatment processes, such as annealing, are used to relieve the internal stress state that develops during cold working, which allows the microstructure to rearrange itself to a state of lower energy. During such processing, individual crystals in a polycrystalline aggregate undergo orientation changes. Often, the recrystallization process creates material texture, or preferred orientation of grains. The degree and type of texture are best described quantitatively by the orientation distribution function (ODF). The distribution of the orientation occurring in the recrystallization textures of polycrystalline titanium and zinc may be represented by a Gaussian ODF (Li, 2000).

Ultrasonic techniques provide information about the interior microstructure due to the penetration of ultrasonic waves. In recent years, major advances in ultrasonic monitoring nondestructive evaluation (NDE) demonstrate a potential to characterize recrystallization processes. Previous analysis of wave propagation and scattering in polycrystals with fiber texture was focused mainly on samples with cubic crystal symmetry (Hirsekorn, 1985, 1986; Turner, 1999; Ahmed and Thompson, 1996). However, crystal symmetries other than cubic are also important. For example, the fiber texture of hexagonal crystals has been observed for a variety of materials, such as titanium, zinc, magnesium, ice, and many others.

The scattering of elastic waves by grains of polycrystals has received considerable attention. Contributions for cubic symmetry with uniformly distributed orientations of grains were made by Hirsekorn (1982, 1983), Stanke and Kino (1984), and Weaver (1990). The problem of wave propagation and scattering in the case of polycrystalline grains with an aligned [001] axis has been examined by Ahmed and Thompson (1996) and Turner (1999). In that case, the average medium is statistically transversely isotropic. Ahmed and Thompson (1992, 1996) also studied correlations defined by both equiaxed grains and grains with elongation. Most recently, wave attenuation in the case of orthorhombic-cubic symmetry was investigated by Yang and Turner (2004).

[a]Author to whom correspondence should be addressed; Electronic mail: lyang4@unl.edu

In this article, quantitative relations between fiber texture and wave attenuations of hexagonal crystals are presented. The texture is characterized by a Gaussian distribution function that contains a single parameter that governs the transition of the texture from perfectly aligned grains to statistically isotropic. Under this assumption, the materials of interest have a varying degree of transverse isotropy representing various states of processing conditions. Simple expressions for the attenuations of the three modes of waves are given in a concise representation. The resulting attenuations are presented and discussed in terms of the directional, frequency, and texture dependence. The results presented are expected to improve the understanding of the microstructure evolution during thermomechanical processing. In addition, the present formulation may be used to study diffuse ultrasonic problems in a straightforward manner.

II. EFFECTIVE ELASTIC STIFFNESS

For textured materials, a detailed description of polycrystalline material properties in the sample requires a knowledge of the orientation distribution of all crystallites in the sample. The orientation of a given single crystallite is specified by the three Euler angles θ, ψ, and φ. The orientation distribution of crystallite grains with preferred directions in the sample can be described by the orientation distribution function (ODF), $F(\theta,\psi,\varphi)$, which is the probability density function in terms of the three Euler angles. To discuss the orientation of a grain, a set of crystallite-fixed axes X_i is chosen for a given grain. One may choose the sample-fixed axes x_i in polycrystals. The crystallite axes X_i and the sample axes x_i are related through a rotation matrix using the three Euler angles. In general, the elastic modulus tensor for a single hexagonal crystallite is given by

$$c_{ijkl} = \lambda^h \delta_{ij}\delta_{kl} + \mu^h(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}) + A(\delta_{ij}\hat{e}_k\hat{e}_l + \delta_{kl}\hat{e}_i\hat{e}_j) + B(\delta_{ik}\hat{e}_j\hat{e}_l + \delta_{il}\hat{e}_j\hat{e}_k + \delta_{jk}\hat{e}_i\hat{e}_l + \delta_{jl}\hat{e}_i\hat{e}_k) + D\hat{e}_i\hat{e}_j\hat{e}_k\hat{e}_l, \quad (1)$$

where the unit vector $\hat{\mathbf{e}}$ is defined as the crystal sixfold symmetry axis. The five coefficients in Eq. (1) are given in terms of single hexagonal crystallite elastic constants c_{11}, c_{33}, c_{44}, c_{12}, c_{13}, c_{66} as

$$\lambda^h = c_{12} = c_{11} - 2c_{66}, \quad \mu^h = c_{66},$$

$$A = c_{13} - c_{12}, \quad B = c_{44} - c_{66},$$

$$D = c_{11} + c_{33} - 2c_{13} - c_{44}. \quad (2)$$

The average medium is characterized by the average stiffness tensor. If the average elastic stiffness tensor, $\langle \mathbf{c} \rangle$, represents a transversely isotropic medium, it may be written as a function of Kronecker deltas and the unit vector $\hat{\mathbf{n}}$, which defines the fiber direction (or uniaxial symmetry axis). It can be expressed by

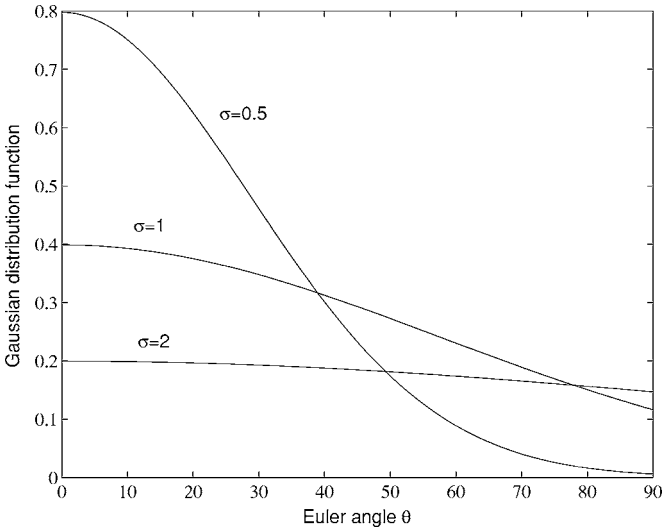


FIG. 1. Gaussian distribution function with various parameters σ.

$$\begin{aligned} C_{ijkl} = \langle c_{ijkl} \rangle &= \Gamma_1\delta_{ij}\delta_{kl} + \Gamma_2(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}) \\ &+ \Gamma_3(\delta_{ij}\hat{n}_k\hat{n}_l + \delta_{kl}\hat{n}_i\hat{n}_j) + \Gamma_4(\delta_{ik}\hat{n}_j\hat{n}_l + \delta_{il}\hat{n}_j\hat{n}_k \\ &+ \delta_{jk}\hat{n}_i\hat{n}_l + \delta_{jl}\hat{n}_i\hat{n}_k) + \Gamma_5\hat{n}_i\hat{n}_j\hat{n}_k\hat{n}_l \\ &= \Gamma_1 J_1 + \Gamma_2 J_2 + \Gamma_3 J_3 + \Gamma_4 J_4 + \Gamma_4 J_5, \end{aligned} \quad (3)$$

where Eq. (3) serves to define the tensors J_i with coefficients Γ_i $(i=1,2,3,4,5)$. Relations between Γ_i and C_{ijkl} are derived below. The ensemble average stiffness of the medium is defined explicitly by

$$\langle c_{ijkl} \rangle = \frac{1}{8\pi^2}\int_0^{\pi}\int_0^{2\pi}\int_0^{2\pi} c_{ijkl}F(\sigma,\theta,\psi,\varphi)\sin\theta\, d\psi\, d\varphi\, d\theta, \quad (4)$$

where c_{ijkl} is given in Eq. (1).

As discussed above, an orientation distribution function (ODF) $F(\sigma,\theta,\psi,\varphi)$ is introduced to represent uniquely the crystallite orientation distribution. Here, a Gaussian distribution function is adopted as the ODF, which is a reasonable approximation for the fiber texture associated with hexagonal materials (Li, 2000). That is, the crystallite distribution depends only on Euler angle θ and is independent of angles φ and ψ,

$$F(\sigma,\theta) = F_0 \exp\left(-\frac{\theta^2}{2\sigma^2}\right), \quad (5)$$

where σ is a single parameter that governs the transition of the texture from perfectly aligned grains to statistically isotropic. The normalization coefficient F_0 is defined by $(1/8\pi^2)\int_0^{\pi}\int_0^{2\pi}\int_0^{2\pi} F(\sigma,\theta,\psi,\varphi)\sin\theta\, d\psi\, d\varphi\, d\theta = 1$. Thus, it may be expressed as

$$F_0 = \frac{\sqrt{\pi}\sigma}{4\sqrt{2}}\exp\left(-\frac{\sigma^2}{2}\right)\left[2\,\mathrm{Erfi}\left(\frac{\sigma}{\sqrt{2}}\right) - \mathrm{Erfi}\left(\frac{-i\pi+\sigma^2}{\sqrt{2}\sigma}\right) - \mathrm{Erfi}\left(\frac{i\pi+\sigma^2}{\sqrt{2}\sigma}\right)\right], \quad (6)$$

where Erfi is the imaginary error function. The shape of this type of distribution function for several values of σ is shown

in Fig. 1. Two extreme cases are observed. One is a quasi-single crystal with perfectly aligned grains which occurs as $\sigma \to 0$. The other limiting case is a statistically isotropic polycrystal with randomly oriented grains which occurs as $\sigma \to \infty$. Other degrees of texture between these limits can be realized by varying σ. Thus, such a distribution function allows a single parameter to model this transition.

In order to obtain the coefficients of Γ_i, the following identities are needed:

$$\langle \hat{e}_i \hat{e}_j \rangle = \delta_{ij} \left[\frac{1}{2}(I_\sigma^0 - I_\sigma^2) \right] + \hat{n}_i \hat{n}_j \left[\frac{1}{2}(-I_\sigma^0 + 3I_\sigma^2) \right],$$

$$\langle \hat{e}_i \hat{e}_j \hat{e}_k \hat{e}_l \rangle = (J_1 + J_2 - J_3 - J_4 + 3J_5) \left[\frac{1}{8}(I_\sigma^0 - 2I_\sigma^2 + I_\sigma^4) \right] + (J_5) I_\sigma^4 + (J_3 + J_4 - 6J_5) \left[\frac{1}{2}(I_\sigma^2 - I_\sigma^4) \right], \quad (7)$$

where the coefficients I_σ^m $(m=0,2,4,6,8)$ are defined by

$$I_\sigma^m = \frac{1}{2} \int_0^\pi F(\sigma, \theta) \cos^m \theta \sin \theta \, d\theta. \quad (8)$$

Using the identities given in Eq. (7), the five coefficients Γ_i in Eq. (3) are obtained and expressed as

$$\Gamma_1 = \lambda^h + 2AM_1 + DM_3,$$

$$\Gamma_2 = \mu^h + 2BM_1 + DM_3,$$

$$\Gamma_3 = AM_2 - DM_3 + DM_5,$$

$$\Gamma_4 = BM_2 - DM_3 + DM_5,$$

$$\Gamma_5 = 3DM_3 + DM_4 - 6DM_5,$$

where M_i are expressed in terms of the coefficients in Eq. (8) as

$$M_1 = \frac{1}{2}(I_\sigma^0 - I_\sigma^2), \quad M_2 = \frac{1}{2}(-I_\sigma^0 + 3I_\sigma^2), \quad M_5 = I_\sigma^4,$$

$$M_3 = \frac{1}{8}(I_\sigma^0 - 2I_\sigma^2 + I_\sigma^4), \quad M_4 = \frac{1}{2}(I_\sigma^2 - I_\sigma^4). \quad (9)$$

Hence, the five effective elastic constants of the ensemble average medium are given by

$$C_{11} = \Gamma_1 + 2\Gamma_2, \quad C_{33} = \Gamma_1 + 2\Gamma_2 + 2\Gamma_3 + 4\Gamma_4 + \Gamma_5,$$

$$C_{44} = \Gamma_2 + \Gamma_4, \quad C_{12} = \Gamma_1, \quad C_{13} = \Gamma_1 + \Gamma_3, \quad (10)$$

$$C_{66} = \frac{1}{2}(C_{11} - C_{12}).$$

It is known that the identities given in Eq. (7) can be simplified under the two limiting cases of interest as (Fedorov, 1968)

$$\langle \hat{e}_i \hat{e}_j \rangle = \frac{1}{3}\delta_{ij}, \quad \langle \hat{e}_i \hat{e}_j \hat{e}_k \hat{e}_l \rangle = \frac{1}{15}(J_1 + J_2), \quad \text{when } \sigma \to \infty,$$

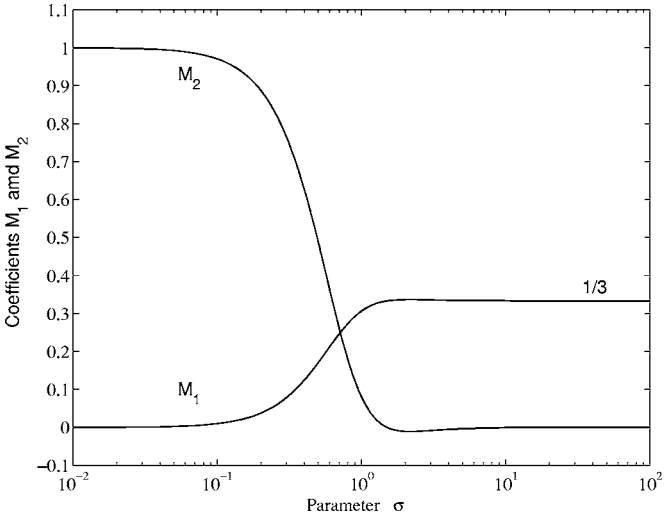


FIG. 2. Coefficients M_1 and M_2 as a function of texture parameter σ.

and $\langle \hat{e}_i \hat{e}_j \rangle = \hat{n}_i \hat{n}_j$, $\langle \hat{e}_i \hat{e}_j \hat{e}_k \hat{e}_l \rangle = \hat{n}_i \hat{n}_j \hat{n}_k \hat{n}_l$, when $\sigma \to 0$. (11)

The identities in Eq. (7) may be rewritten as

$$\langle \hat{e}_i \hat{e}_j \rangle = M_1 \delta_{ij} + M_2 \hat{n}_i \hat{n},$$

$$\langle \hat{e}_i \hat{e}_j \hat{e}_k \hat{e}_l \rangle = S_1(J_1 + J_2) + S_2(J_3 + J_4) + S_3 J_5, \quad (12)$$

where

$$S_1 = \frac{1}{8}(I_\sigma^0 - 2I_\sigma^2 + I_\sigma^4),$$

$$S_2 = \frac{1}{8}(-I_\sigma^0 + 6I_\sigma^2 - 5I_\sigma^4), \quad S_3 = \frac{1}{8}(3I_\sigma^0 - 30I_\sigma^2 + 35I_\sigma^4).$$

The five coefficients M_1 and M_2, and S_1, S_2, and S_3 are shown in Figs. 2 and 3, respectively. It is observed as $\sigma \to \infty$, the coefficients M_2, and S_2 and S_3 approach zero, and as $\sigma \to 0$, the coefficients M_1, and S_1 and S_2 approach zero. Thus, the expected limiting behavior for **C** is observed. For the statistically isotropic case, the average elastic stiffness of a hexagonal crystallite may be simplified as (Li and Thompson, 1990)

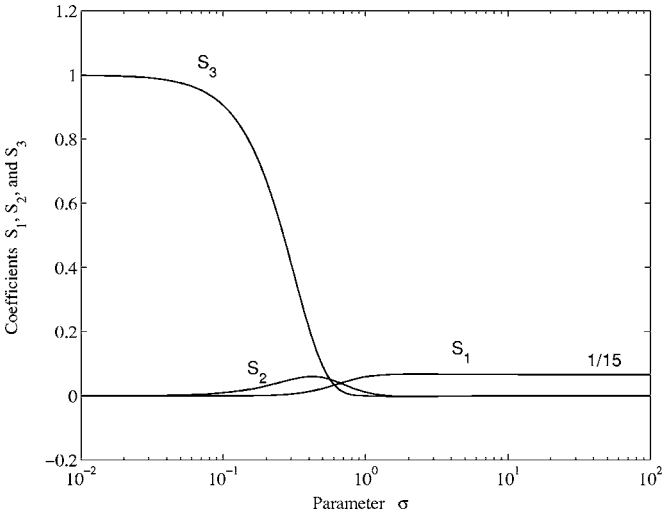


FIG. 3. Coefficients S_1, S_2, and S_3 as a function of texture parameter σ.

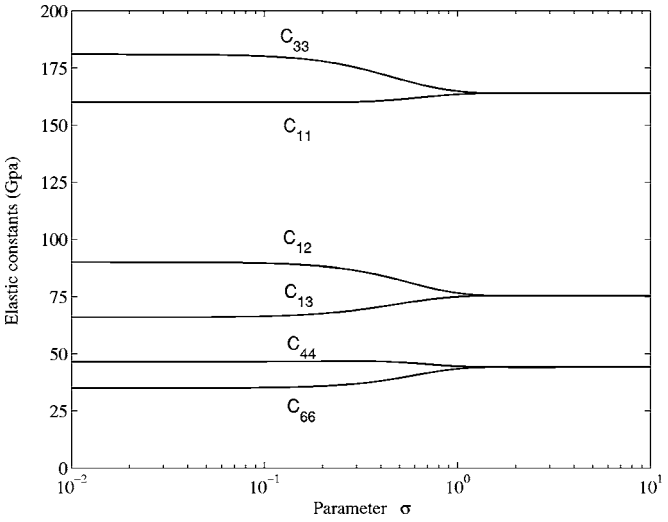


FIG. 4. Effective elastic constants as a function of texture parameter σ.

$$C_{ijkl} = \bar{\lambda}\delta_{ij}\delta_{kl} + \bar{\mu}(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}) = \left(\lambda^h + \frac{2}{3}A + \frac{D}{15}\right)\delta_{ij}\delta_{kl} + \left(\mu^h + \frac{2}{3}B + \frac{D}{15}\right) \times (\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}). \quad (13)$$

Example results of the effective elastic constants are now shown for a titanium polycrystal. The material constants of the single crystallite used here are $c_{11}=160$ GPa, $c_{12}=90$ GPa, $c_{13}=66$ GPa, $c_{33}=181$ GPa, and $c_{44}=46.5$ GPa. In Fig. 4, the elastic constants are presented as a function of σ, where σ governs the crystal alignment from perfectly aligned ($\sigma=0$) to randomly aligned ($\sigma\to\infty$). It is clearly seen from Fig. 4 that as σ approaches zero, the grains in polycrystals become perfectly aligned. In this case, the polycrystal behaves as a quasi-single crystal, the elastic constants of which reduce to the appropriate single-crystal constants. When σ approaches infinity, the grains in the polycrystal are randomly oriented as expected. In such a case, the five independent elastic constants of the polycrystal reduce to two independent elastic constants. The transition between these two limits is seen clearly. Because the focus here is on wave propagation and scattering phenomena, the directional dependence of the wave speeds is of importance. Thus, slowness surfaces are presented for various parameter σ for the shear horizontal, quasilongitudinal, and quasishear waves, respectively, in Figs. 5, 6, and 7. The transition of the texture from perfectly aligned to statistically isotropic is clear.

III. ATTENUATION

The scattering of elastic waves, often characterized by the attenuation, in a polycrystal results from the misalignment of the grains. To calculate the attenuations, the relevant inner products on the covariance of the moduli fluctuations are required (Turner, 1999). These may be written as

$$\Xi^{\cdots\hat{\mathbf{u}}\hat{\mathbf{p}}\hat{\mathbf{s}}\hat{\mathbf{v}}}_{\cdots\hat{\mathbf{u}}\hat{\mathbf{p}}\hat{\mathbf{s}}\hat{\mathbf{v}}} = \Xi^{ijkl}_{\alpha\beta\gamma\delta}\hat{u}_\beta\hat{u}_k\hat{p}_\alpha\hat{p}_l\hat{s}_i\hat{s}_\delta\hat{v}_\gamma\hat{v}_j, \quad (14)$$

where the covariance of the moduli fluctuations $\Xi^{ijkl}_{\alpha\beta\gamma\delta}$ is given explicitly by

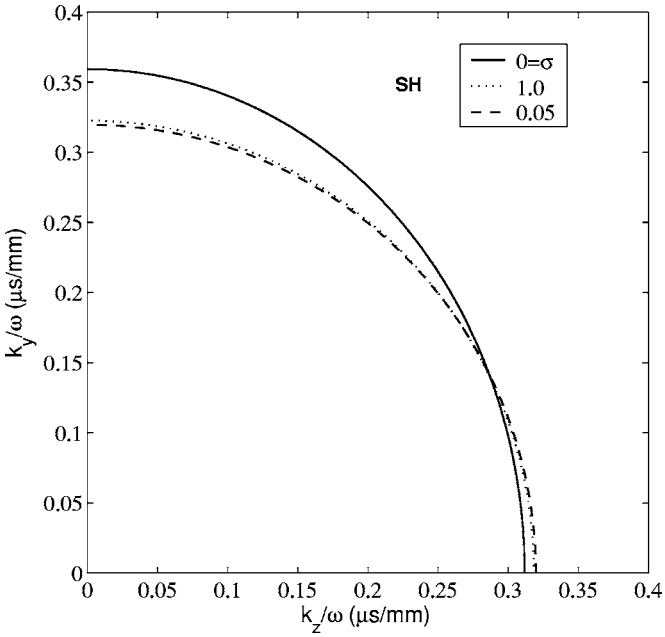


FIG. 5. Slowness surface of the shear horizontal wave (*SH*).

$$\Xi^{\alpha\beta\gamma\delta}_{ijkl} = \langle c_{ijkl}c_{\alpha\beta\gamma\delta}\rangle - \langle c_{ijkl}\rangle\langle c_{\alpha\beta\gamma\delta}\rangle. \quad (15)$$

The brackets, $\langle\ \rangle$, denote an ensemble average over all orientations of grains. The average elastic stiffness tensor, $\langle\mathbf{c}\rangle$, is given in Eq. (4) for the case of interest here. The first term in Eq. (15) $\langle\mathbf{cc}\rangle$ is defined by

$$\langle c_{ijkl}c_{\alpha\beta\gamma\delta}\rangle = \frac{1}{8\pi^2}\int_0^\pi\int_0^{2\pi}\int_0^{2\pi} c_{ijkl}c_{\alpha\beta\gamma\delta}F(\sigma,\theta) \times \sin\theta\, d\psi\, d\varphi\, d\theta. \quad (16)$$

In order to carry out the calculations in Eq. (16), the following general identities are given:

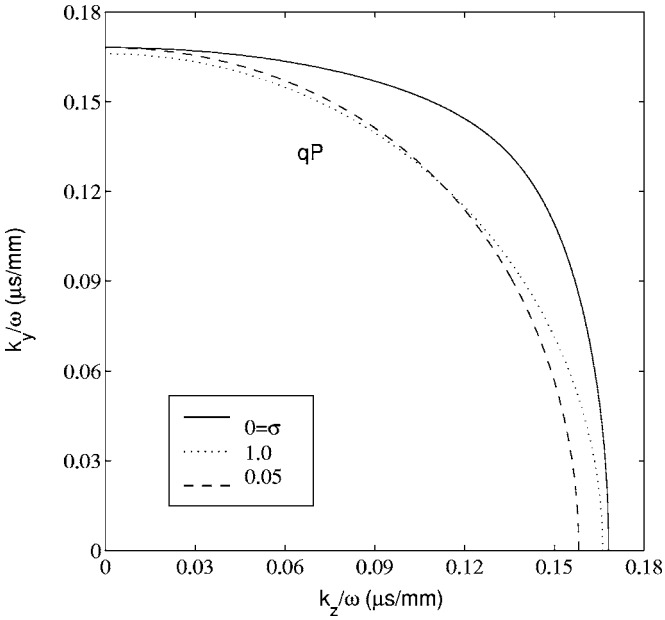


FIG. 6. Slowness surface of the quasilongitudinal wave (*qP*).

$$\langle \hat{e}_i\hat{e}_j\hat{e}_k\hat{e}_l\hat{e}_\alpha\hat{e}_\beta\rangle = \left[\frac{1}{48}(I_\sigma^0 - 3I_\sigma^2 + 3I_\sigma^4 - I_\sigma^6)\right]\{\Delta_{ij}\Delta_{kl}\Delta_{\alpha\beta} + \text{all permutations-15 terms in all}\}$$
$$+\left[\frac{1}{8}(I_\sigma^2 - 2I_\sigma^4 + I_\sigma^6)\right]\{\hat{n}_i\hat{n}_j[\Delta_{kl}\Delta_{\alpha\beta} + \Delta_{k\beta}\Delta_{l\alpha} + \Delta_{k\alpha}\Delta_{l\beta}] + \text{all permutations-15 terms in all}\}$$
$$+\left[\frac{1}{2}(I_\sigma^4 - I_\sigma^6)\right]\{\hat{n}_i\hat{n}_j\hat{n}_k\hat{n}_l\Delta_{ij} + \text{all permutations-15 terms in all}\} + [I_\sigma^6]\hat{n}_i\hat{n}_j\hat{n}_k\hat{n}_l\hat{n}_\alpha\hat{n}_\beta, \quad (17)$$

$$\langle \hat{e}_i\hat{e}_j\hat{e}_k\hat{e}_l\hat{e}_\alpha\hat{e}_\beta\hat{e}_\gamma\hat{e}_\delta\rangle = \left[\frac{1}{384}(I_\sigma^0 - 4I_\sigma^2 + 6I_\sigma^4 - 4I_\sigma^6 + I_\sigma^8)\right]\{\Delta_{ij}\Delta_{kl}\Delta_{\alpha\beta}\Delta_{\gamma\delta} + \text{all permutations-105 terms in all}\}$$
$$+\left[\frac{1}{48}(I_\sigma^2 - 3I_\sigma^4 + 3I_\sigma^6 - I_\sigma^8)\right]\{\hat{n}_i\hat{n}_j\Delta_{kl}\Delta_{\alpha\beta}\Delta_{\gamma\delta} + \text{all permutations-420 terms in all}\}$$
$$+\left[\frac{1}{8}(I_\sigma^4 - 2I_\sigma^6 + I_\sigma^8)\right]\{\hat{n}_i\hat{n}_j\hat{n}_k\hat{n}_l\Delta_{\alpha\beta}\Delta_{\gamma\delta} + \text{all permutations-210 terms in all}\}$$
$$+\left[\frac{1}{2}(I_\sigma^6 - I_\sigma^8)\right]\{\hat{n}_i\hat{n}_j\hat{n}_k\hat{n}_l\hat{n}_\alpha\hat{n}_\beta\Delta_{\gamma\delta} + \text{all permutations-28 terms in all}\} + [I_\sigma^8]\hat{n}_i\hat{n}_j\hat{n}_k\hat{n}_l\hat{n}_\alpha\hat{n}_\beta\hat{n}_\gamma\hat{n}_\delta, \quad (18)$$

where $\Delta_{mn} = \delta_{mn} - \hat{n}_m\hat{n}_n$. In Fig. 8, examples of a few terms of $\langle \mathbf{cc}\rangle$ are shown. As expected those values reach a constant value, respectively, as $\sigma \to \infty$. Using the identities given in Eqs. (7), (17), and (18), the covariance $\Xi_{ijkl}^{\alpha\beta\gamma\delta}$ can be written in a general form, which is not presented here due to brevity. Example values of the covariance $\Xi_{ijkl}^{\alpha\beta\gamma\delta}$ of interest are plotted in Fig. 9. It is observed from Fig. 9 that the covariance $\Xi_{ijkl}^{\alpha\beta\gamma\delta}$ reduces to two constants when σ becomes larger.

For the inner products $\Xi_{\cdots\hat{\mathbf{u}}\hat{\mathbf{p}}\hat{\mathbf{s}}\hat{\mathbf{v}}}^{\cdots\hat{\mathbf{u}}\hat{\mathbf{p}}\hat{\mathbf{s}}\hat{\mathbf{v}}}$ presented in Eq. (14), the vectors $\hat{\mathbf{p}}$ and $\hat{\mathbf{s}}$, respectively, represent the incoming and outgoing propagation directions. The vectors $\hat{\mathbf{u}}$ and $\hat{\mathbf{v}}$ are vectors defining the polarization directions of the particular waves. Without loss of generality, the vectors are defined with respect to a general xyz coordinate system as shown in Fig. 10. The vectors $\hat{\mathbf{n}}$, $\hat{\mathbf{p}}$, and $\hat{\mathbf{s}}$ are given by

$$\hat{\mathbf{n}} = \hat{\mathbf{z}}, \quad \hat{\mathbf{p}} = \hat{\mathbf{x}}\sin\Theta + \hat{\mathbf{y}}\cos\Theta,$$
$$\hat{\mathbf{s}} = \hat{\mathbf{x}} \quad \sin\Theta'\cos\phi' + \hat{\mathbf{y}}\sin\Theta'\sin\phi' + \hat{\mathbf{z}}\cos\Theta'. \quad (19)$$

The polarization vectors $\hat{\mathbf{u}}$ and $\hat{\mathbf{v}}$ as shown in Fig. 10 are given elsewhere (Turner, 1999). Substituting those unit vectors into Eq. (14), and combining with Eq. (15), the general inner products $\Xi_{\cdots\hat{\mathbf{u}}\hat{\mathbf{p}}\hat{\mathbf{s}}\hat{\mathbf{v}}}^{\cdots\hat{\mathbf{u}}\hat{\mathbf{p}}\hat{\mathbf{s}}\hat{\mathbf{v}}}$ may be calculated. An example inner product associated with the attenuation of a shear horizontal wave is presented in the Appendix. If the tensorial and spatial components of covariance are assumed to be independent, the spatial correlation function W is not dependent on the tensorial part. For simplicity, the correlation function W is assumed to have an exponential form

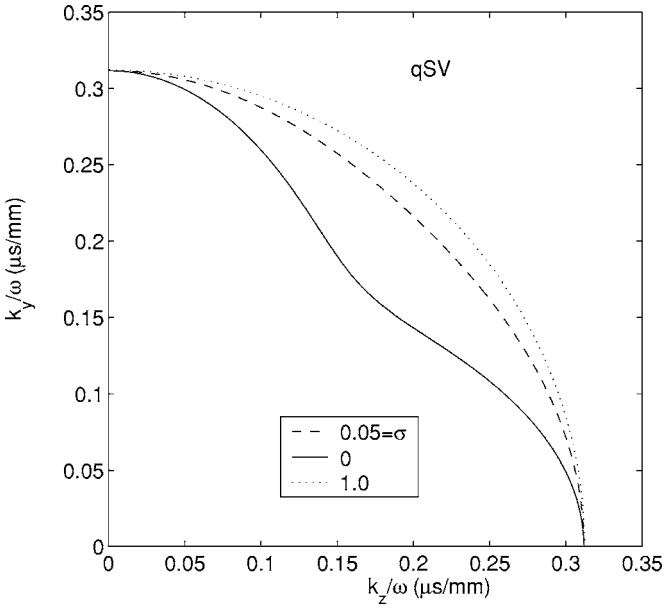


FIG. 7. Slowness surface of the quasi-shear vertical wave (qSV).

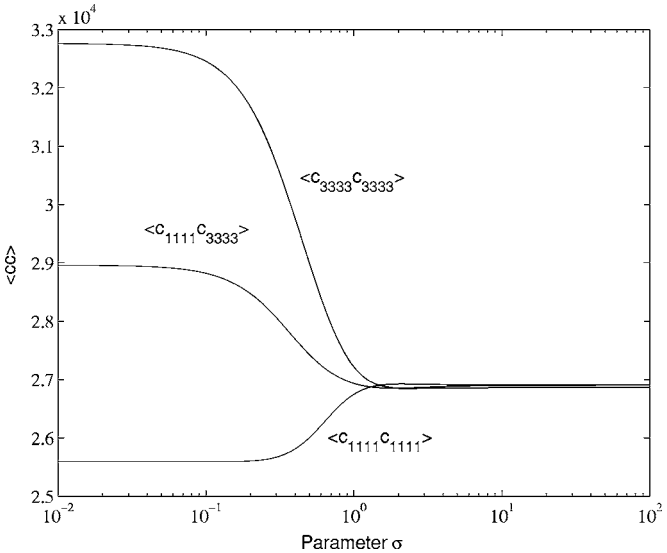


FIG. 8. Various terms $\langle \mathbf{cc}\rangle$ as a function of σ.

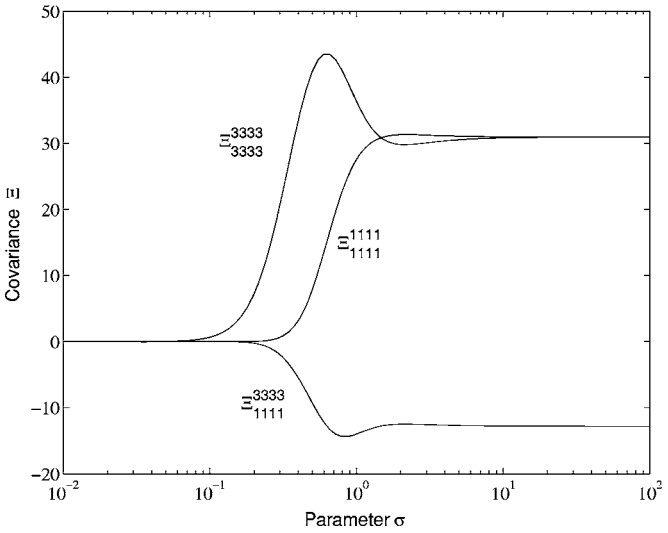


FIG. 9. Covariance Ξ as a function of σ.

$$W(r) = e^{-r/L},$$

where L is the spatial correlation length, which is of the order of the grain radius of the crystallites. In general, the simple exponential form of the correlation function must be modified for the polycrystalline materials with texture since grain elongation often exists. A more general expression (Ahmed and Thompson, 1992) will be used in future work. The influence of the choice of correlation function on the attenuations is not the subject of the present work.

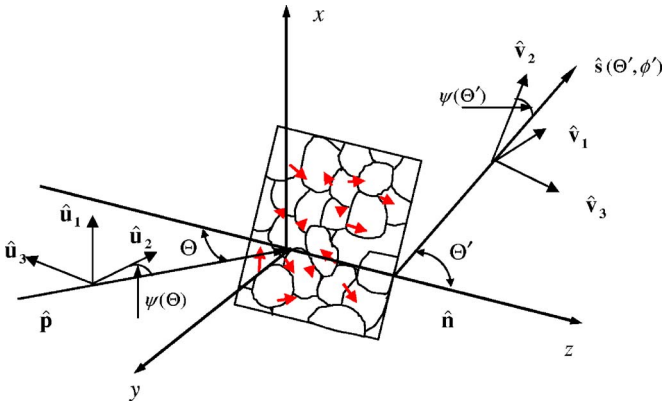


FIG. 10. (Color online) Geometry for the propagation direction $\hat{\mathbf{p}}$, the scattered direction $\hat{\mathbf{s}}$, and the respective polarization directions $\hat{\mathbf{u}}$ and $\hat{\mathbf{v}}$.

Using the expressions given above, the resulting dimensionless attenuations are written in a general form as (Yang and Turner, 2004)

$$\alpha_\beta(\hat{\mathbf{p}})L = \frac{x_\beta^4 c_\beta(\hat{\mathbf{p}})}{2\rho^2}\left\{\int_{4\pi}\frac{\Xi^{\cdots\hat{\mathbf{u}}_K\hat{\mathbf{p}}\hat{\mathbf{s}}\hat{\mathbf{v}}_1}_{\cdots\hat{\mathbf{u}}_K\hat{\mathbf{p}}\hat{\mathbf{s}}\hat{\mathbf{v}}_1}(\hat{\mathbf{p}},\hat{\mathbf{s}})}{(1+x_\beta^2(\hat{\mathbf{p}})+x_{SH}^2(\hat{\mathbf{s}})-2x_\beta(\hat{\mathbf{p}})x_{SH}(\hat{\mathbf{s}})\hat{\mathbf{p}}\cdot\hat{\mathbf{s}})^2}\frac{1}{c_{SH}^5(\hat{\mathbf{s}})}d^2\hat{s}\right.$$
$$+\int_{4\pi}\frac{\Xi^{\cdots\hat{\mathbf{u}}_K\hat{\mathbf{p}}\hat{\mathbf{s}}\hat{\mathbf{v}}_2}_{\cdots\hat{\mathbf{u}}_K\hat{\mathbf{p}}\hat{\mathbf{s}}\hat{\mathbf{v}}_2}(\hat{\mathbf{p}},\hat{\mathbf{s}})}{(1+x_\beta^2(\hat{\mathbf{p}})+x_{qP}^2(\hat{\mathbf{s}})-2x_\beta(\hat{\mathbf{p}})x_{qP}(\hat{\mathbf{s}})\hat{\mathbf{p}}\cdot\hat{\mathbf{s}})^2}\frac{1}{c_{qP}^5(\hat{\mathbf{s}})}d^2\hat{s}$$
$$\left.+\int_{4\pi}\frac{\Xi^{\cdots\hat{\mathbf{u}}_K\hat{\mathbf{p}}\hat{\mathbf{s}}\hat{\mathbf{v}}_3}_{\cdots\hat{\mathbf{u}}_K\hat{\mathbf{p}}\hat{\mathbf{s}}\hat{\mathbf{v}}_3}(\hat{\mathbf{p}},\hat{\mathbf{s}})}{(1+x_\beta^2(\hat{\mathbf{p}})+x_{qSV}^2(\hat{\mathbf{s}})-2x_\beta(\hat{\mathbf{p}})x_{qSV}(\hat{\mathbf{s}})\hat{\mathbf{p}}\cdot\hat{\mathbf{s}})^2}\frac{1}{c_{qSV}^5(\hat{\mathbf{s}})}d^2\hat{s}\right\}, \tag{20}$$

where K is defined as the polarization for the wave type β (1, 2, or 3 for wave types SH, qP, and qSV, respectively). The normalized frequency x_β is defined by $x_\beta = \omega L/c_\beta$, where ω is the frequency and c_β is the wave velocity of each type. In Eq. (20), it can be seen that the integrals are over the unit sphere, which is defined by unit vector $\hat{\mathbf{s}}$. Further details of the scattering model can be reviewed by the reader in the articles of Weaver (1990) and Turner (1999). For the extreme case of statistical isotropy ($\sigma\to\infty$), the dimensionless longitudinal and transverse attenuations can be given as follows:

$$\alpha_L = \alpha_{LL}L + \alpha_{LT}L, \quad \alpha_T = \alpha_{TL}L + \alpha_{TT}L, \tag{21}$$

where

$$\alpha_{LL}L = \frac{x_L^4}{2c_L^4}\int_{-1}^{+1}\frac{\Xi^{\cdots\hat{\mathbf{p}}\hat{\mathbf{p}}\hat{\mathbf{s}}\hat{\mathbf{s}}}_{\cdots\hat{\mathbf{p}}\hat{\mathbf{p}}\hat{\mathbf{s}}\hat{\mathbf{s}}}}{[1+2x_L^2(1-\chi)]^2}d\chi,$$

$$\alpha_{LT}L = \frac{x_T^4}{2c_L^3c_T}\int_{-1}^{+1}\frac{[\Xi^{\cdots\hat{\mathbf{p}}\hat{\mathbf{p}}\hat{\mathbf{s}}\hat{\mathbf{s}}}_{\cdots\hat{\mathbf{p}}\hat{\mathbf{p}}\hat{\mathbf{s}}\hat{\mathbf{s}}_2}+\Xi^{\cdots\hat{\mathbf{p}}\hat{\mathbf{p}}\hat{\mathbf{s}}\hat{\mathbf{s}}}_{\cdots\hat{\mathbf{p}}\hat{\mathbf{p}}\hat{\mathbf{s}}\hat{\mathbf{s}}_3}]}{(1+x_L^2+x_T^2-2x_Lx_T\chi)^2}d\chi,$$

$$\alpha_{TL}L = \frac{1}{2}\left(\frac{c_T}{c_L}\right)^2\alpha_{LT}L,$$

$$\alpha_{TT}L = \frac{x_T^4}{4c_T^4}\times\int_{-1}^{+1}\frac{[\Xi^{\cdots\hat{\mathbf{p}}_2\hat{\mathbf{p}}\hat{\mathbf{s}}\hat{\mathbf{s}}_2}_{\cdots\hat{\mathbf{p}}_2\hat{\mathbf{p}}\hat{\mathbf{s}}\hat{\mathbf{s}}_2}+\Xi^{\cdots\hat{\mathbf{p}}_3\hat{\mathbf{p}}\hat{\mathbf{s}}\hat{\mathbf{s}}_3}_{\cdots\hat{\mathbf{p}}_3\hat{\mathbf{p}}\hat{\mathbf{s}}\hat{\mathbf{s}}_3}+\Xi^{\cdots\hat{\mathbf{p}}_2\hat{\mathbf{p}}\hat{\mathbf{s}}\hat{\mathbf{s}}_3}_{\cdots\hat{\mathbf{p}}_2\hat{\mathbf{p}}\hat{\mathbf{s}}\hat{\mathbf{s}}_3}+\Xi^{\cdots\hat{\mathbf{p}}_3\hat{\mathbf{p}}\hat{\mathbf{s}}\hat{\mathbf{s}}_2}_{\cdots\hat{\mathbf{p}}_3\hat{\mathbf{p}}\hat{\mathbf{s}}\hat{\mathbf{s}}_2}]}{[1+2x_T^2(1-\chi)]^2}$$
$$\times d\chi, \tag{22}$$

with $\chi = \hat{\mathbf{p}}\cdot\hat{\mathbf{s}} = \cos\Phi$, where Φ is the angle separating the propagation direction $\hat{\mathbf{p}}$ and the scattering direction $\hat{\mathbf{s}}$. The detailed expressions for the inner products in Eq. (22) are given in the Appendix.

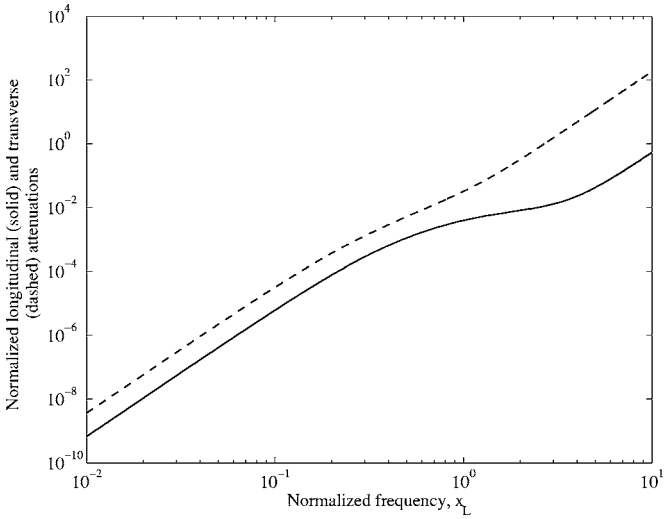


FIG. 11. Normalized longitudinal α_L and transverse α_T attenuations in terms of normalized frequency x_L.

Numerical results of the attenuations are now presented for titanium, using the single elastic constants given above. First, Fig. 11 presents the normalized longitudinal and transverse attenuations which are given in Eq. (21) as a function of dimensionless frequency, x_L, for the extreme case of statistically isotropy ($\sigma \to \infty$). It is shown that the transverse attenuation is always greater than the longitudinal attenuation as shown in cubic crystals. Next, the general attenuations as a function of texture parameter σ are presented. To compute the attenuations of three wave types, numerical integrations are employed in Eq. (20). The inner products on the covariance of the moduli fluctuations are obtained by Eq. (14). Using the covariance and wave propagation vectors obtained, the inner products of each wave type are calculated numerically. Then, the attenuations of the three wave types as a function of texture parameter σ are obtained.

The texture dependence of the attenuations for a given dimensionless frequency $x_{SH}=0.5$ is presented. Figure 12 shows the dimensionless attenuation of the shear horizontal wave (*SH*) as a function of the texture parameter σ for given

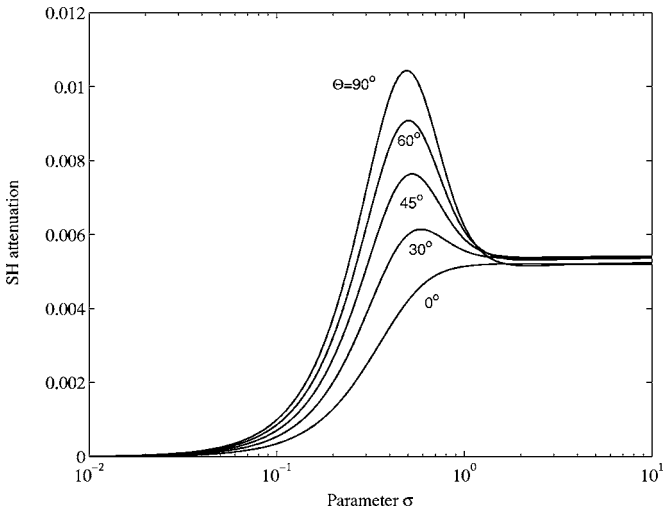


FIG. 12. Attenuations of the shear horizontal wave (*SH*) versus texture parameter σ with various wave propagation directions Θ and the given frequency $x_{SH}=0.5$.

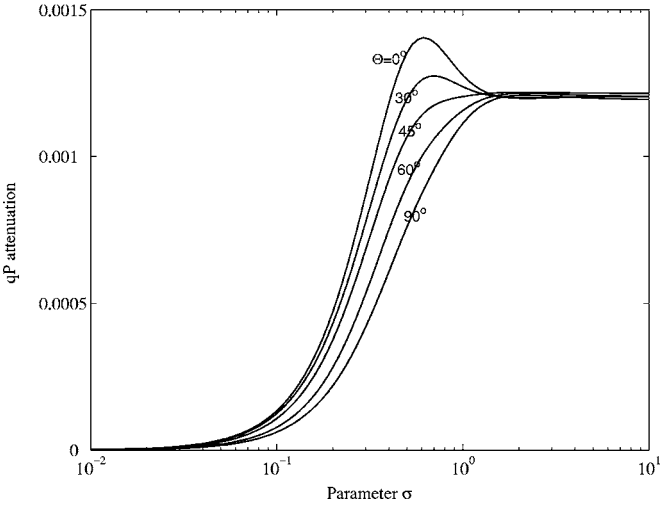


FIG. 13. Attenuations of the quasi-longitudinal wave (*qP*) versus texture parameter σ with various wave propagation directions Θ and the given frequency $x_{SH}=0.5$.

wave propagation directions Θ. It is observed that attenuation is zero when the texture parameter σ is small. Since in this limit the hexagonal crystals are perfectly aligned, they behave as a single crystal without scattering attenuation as expected. After the zero attenuation region, the attenuation increases to reach a maximum, then decreases to constant attenuation. The attenuations along various propagation directions all recover the isotropic limit as expected. The slight error observed in the isotropic limit is due to limits of the numerical integrations. It is also observed from Fig. 12 that there are different maxima for different wave propagation directions such that it is sensitive to the propagation direction due to texture. In particular, it is seen that there is no peak presented for $\Theta=0°$. Such information is useful for monitoring microstructure evolution during processing.

In Fig. 13, the dimensionless quasi-longitudinal attenuation (*qP*) is plotted versus the texture parameter σ for the given dimensionless frequency and wave propagation direction. It is observed that the attenuation shows a similar track with texture parameter σ to *SH* wave as well. The attenuation is zero for σ small, then increases to reach a maximum value. It is interesting to see that there are no maxima except at $\Theta=0°$ and $\Theta=30°$ appearing in other selected propagation directions. The dependence of these maxima on single-crystal parameters is the subject of future research. The relationship between the attenuation of the quasi-shear vertical wave (*qSV*) and the texture parameter σ is presented in Fig. 14. It is seen that there are maxima displayed for some propagation directions. However, the variation of these peaks of the *qSV* wave is much smaller than those of the *SH* and *qP* waves. Therefore, it might be concluded that the *SH* and *qP* waves are more sensitive to the grain orientation during processing. The attenuation results provide good motivation to use experiment measurements to detect changes of microstructure of hexagonal materials during the cold-working process. In addition, the experimental data could be used with various representative samples of cold working to clarify the peak observed in the theoretical predictions as part of future research.

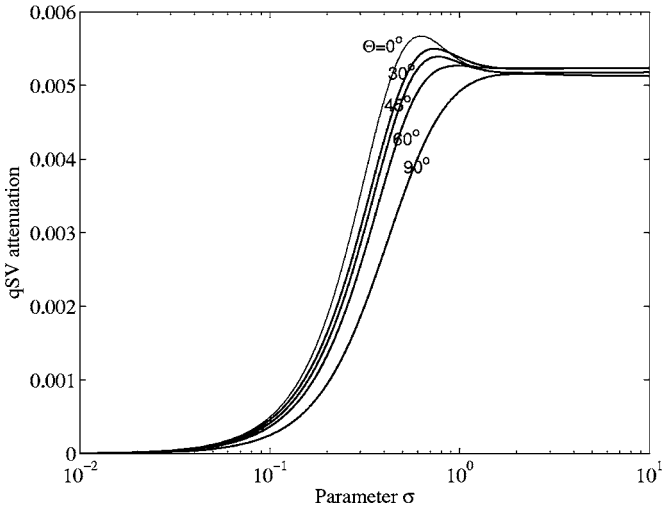


FIG. 14. Attenuations of the quasi-shear vertical wave (*qSV*) versus texture parameter σ with various wave propagation directions Θ and the given frequency x_{SH}=0.5.

IV. CONCLUSIONS

In this article, the wave propagation and scattering in hexagonal polycrystalline materials with fiber texture was discussed. The quantitative relations between fiber texture and wave attenuations of hexagonal crystals were presented. The texture is characterized by a Gaussian distribution function that contains a single parameter that governs the transition of the texture from perfectly aligned crystals to statistically isotropic. Under this assumption, the materials of interest have a varying degree of transverse isotropy representative of processing conditions. Simple expressions for the attenuations of the three modes of waves are given in a concise representation. Finally, numerical results for titantium were presented and discussed in terms of the directional, frequency, and texture dependence. The results show that the attenuations of each wave type can be considerably affected during processing since the material's microstructure is changing. The results presented are expected to improve the understanding of the texture variations during processing. In experimental measurements, if one measures the attenuations at several times during the annealing process, it might be possible to obtain the texture and grain size information using this model, which is a future research topic.

ACKNOWLEDGMENTS

We gratefully acknowledge the financial support of the U.S. Department of Energy (Grant No. DE-FG02-01ER45890).

APPENDIX: EXAMPLE INNER PRODUCTS

As an example, the inner products of the shear horizontal wave are given in the following. For the *SH* wave, we have

$$\begin{aligned}\Xi^{\ldots\hat{\mathbf{u}}_1\hat{\mathbf{p}}\hat{\mathbf{s}}\hat{\mathbf{v}}_1}_{\ldots\hat{\mathbf{u}}_1\hat{\mathbf{p}}\hat{\mathbf{s}}\hat{\mathbf{v}}_1} &= \sin^2\Theta\sin^2\Theta'[a_1\sin^4\phi' - a_1\sin^2\phi' + a_2] + a_3\\ &\quad + \sin\Theta\cos\Theta\sin\Theta'\cos\Theta'[a_4\sin^3\phi'\\ &\quad + a_5\sin\phi'] + \cos^2\Theta\cos^2\Theta'[a_6\sin^2\phi' + a_7];\end{aligned}$$

$$\begin{aligned}\Xi^{\ldots\hat{\mathbf{u}}_1\hat{\mathbf{p}}\hat{\mathbf{s}}\hat{\mathbf{v}}_2}_{\ldots\hat{\mathbf{u}}_1\hat{\mathbf{p}}\hat{\mathbf{s}}\hat{\mathbf{v}}_2} &= \sin^2\Theta\sin^2\Theta'\sin^2\gamma'[a_1\sin^2\phi'\cos^2\phi' + a_8]\\ &\quad + \sin^2\Theta(\cos^2\Theta'\sin^2\gamma' + \sin^2\Theta'\cos^2\gamma')[a_4]\\ &\quad + \sin^2\Theta\cos^2\Theta'\cos^2\gamma'[a_9]\\ &\quad + \sin^2\Theta\sin\Theta'\cos\Theta'\sin\gamma'\cos\gamma'[a_{10}]\\ &\quad + \sin\Theta\cos\Theta(\cos^2\Theta'\sin\gamma'\cos\gamma'\\ &\quad + \sin\Theta'\cos\Theta'\cos^2\gamma')[a_{11}\sin\phi']\\ &\quad + \sin\Theta\cos\Theta(\sin^2\Theta'\sin\gamma'\cos\gamma'\\ &\quad + \sin\Theta'\cos\Theta'\sin^2\gamma')[a_{12}\sin^3\phi' + a_{13}\sin\phi']\\ &\quad + \cos^2\Theta\cos^2\Theta'\cos^2\gamma'[a_{14}]\\ &\quad + \cos^2\Theta\sin^2\Theta'\sin^2\gamma'[a_{15}\cos^2\phi' + a_{16}]\\ &\quad + \cos^2\Theta(\sin^2\Theta'\cos^2\gamma' + \cos^2\Theta'\sin^2\gamma')\\ &\quad \times[a_{17}\cos^2\phi' + a_{18}]\\ &\quad + \cos^2\Theta\sin\Theta'\cos\Theta'\sin\gamma'\cos\gamma'\\ &\quad \times[a_{19}\cos^2\phi' + a_{20}];\end{aligned}$$

$$\begin{aligned}\Xi^{\ldots\hat{\mathbf{u}}_1\hat{\mathbf{p}}\hat{\mathbf{s}}\hat{\mathbf{v}}_3}_{\ldots\hat{\mathbf{u}}_1\hat{\mathbf{p}}\hat{\mathbf{s}}\hat{\mathbf{v}}_3} &= \sin^2\Theta\sin^2\Theta'\cos^2\gamma'[a_1\sin^2\phi'\cos^2\phi' + a_8]\\ &\quad + \sin^2\Theta(\cos^2\Theta'\cos^2\gamma' + \sin^2\Theta'\sin^2\gamma')[a_4]\\ &\quad + \sin^2\Theta\cos^2\Theta'\sin^2\gamma'[a_9]\\ &\quad - \sin^2\Theta\sin\Theta'\cos\Theta'\sin\gamma'\cos\gamma'[a_{10}]\\ &\quad + \sin\Theta\cos\Theta(-\cos^2\Theta'\sin\gamma'\cos\gamma'\\ &\quad + \sin\Theta'\cos\Theta'\sin^2\gamma')[a_{11}\sin\phi']\\ &\quad + \sin\Theta\cos\Theta(-\sin^2\Theta'\sin\gamma'\cos\gamma'\\ &\quad + \sin\Theta'\cos\Theta'\cos\sin^2\gamma')[a_{12}\sin^3\phi'\\ &\quad + a_{13}\sin\phi'] + \cos^2\Theta\cos^2\Theta'\cos^2\gamma'[a_{14}]\\ &\quad + \cos^2\Theta\sin^2\Theta'\cos^2\gamma'[a_{15}\cos^2\phi' + a_{16}]\\ &\quad + \cos^2\Theta(\sin^2\Theta'\sin^2\gamma' + \cos^2\Theta'\cos^2\gamma')\\ &\quad \times[a_{17}\cos^2\phi' + a_{18}]\\ &\quad - \cos^2\Theta\sin\Theta'\cos\Theta'\sin\gamma'\cos\gamma'\\ &\quad \times[a_{19}\cos^2\phi' + a_{20}];\end{aligned}$$

where the coefficients are given by

$$\begin{aligned}a_1 &= -4(2BM_1 + DM_3)^2 + 32B^2M_3 + 24D^2N_1\\ &\quad + 48BDM_6,\end{aligned}$$

$$\begin{aligned}a_2 &= -(2BM_1 + DM_3)^2 + 9D^2N_1 + 8B^2M_3 + 12BDM_6\\ &\quad - 3D^2N_2 - B^2M_4 - 2BDM_7,\end{aligned}$$

$$a_3 = 3D^2N_2 + B^2M_4 + 2BDM_7,$$

$$\begin{aligned}a_4 &= -4BDM_2M_3 - 8BDM_1M_3 - 8BDM_1M_4\\ &\quad - 4D^2M_3M_4 + 12BDM_6 + 16B^2M_3 + 16B^2M_4\\ &\quad - 8B^2M_1M_2 + 36BDM_7 - 16B^2M_1^2 + 24D^2N_2,\end{aligned}$$

$$a_5 = -12D^2N_2 - 22BDM_7 - 6BDM_6 - 8B^2M_3 + 2BDM_2M_3 + 4BDM_1M_3 + 2D^2M_3M_4 + 8B^2M_1^2 - 10B^2M_4 + 4B^2M_1M_2 + 4BDM_1M_4,$$

$$a_6 = B^2M_5 - 4B^2M_1^2 - D^2M_4^2 + 2B^2M_4 + 4BDM_7 - 2BDM_2M_4 - 4BDM_1M_4 + 2D^2N_3 + 2B^2M_3 + 2BDM_8 - 4B^2M_1M_2 - B^2M_2^2,$$

$$a_7 = D^2N_3 + B^2M_3 - B^2M_4 - 3D^2N_2,$$

$$a_8 = 4ABM_3 + 12BDM_6 + 4B^2M_3 + 6ADM_6 + A^2M_3 + 15D^2N_1,$$

$$a_9 = A^2M_3 + 2ADM_7 + D^2N_3,$$

$$a_{10} = 4ABM_3 + 2B^2M_4 + 8BDM_7 + 6ADM_6 + 2A^2M_3 + 2ADM_7 + 12D^2N_2,$$

$$a_{11} = 2ABM_4 + 4D^2N_3 + 6BDM_7 + 2ABM_3 + 2BDM_8 + 4ADM_7 + 4B^2M_4,$$

$$a_{12} = -24D^2N_2 + 16B^2M_1^2 - 16B^2M_4 + 8BDM_1M_4 - 36BDM_7 + 4BDM_2M_3 + 8BDM_1M_3 + 4D^2M_3M_4 - 12BDM_6 + 8B^2M_1M_2 - 16B^2M_3,$$

$$a_{13} = 36D^2N_2 - 16B^2M_1^2 + 16B^2M_4 + 2ABM_3 - 8BDM_1M_4 + 42BDM_7 - 4BDM_2M_3 - 8BDM_1M_3 - 4D^2M_3M_4 + 2ABM_4 + 18BDM_6 + 4ADM_7 - 8B^2M_1M_2 + 20B^2M_3,$$

$$a_{14} = 4BDM_8 + D^2N_4 + A^2M_4 + 4ABM_4 + 2ADM_8 + 4B^2M_4,$$

$$a_{15} = 12BDM_7 + 4ABM_4 + 4ADM_7 + 4B^2M_4 + 12D^2N_2,$$

$$a_{16} = A^2M_4 + 2ADM_7 + 3D^2N_2,$$

$$a_{17} = 4BDM_7 + 2BDM_8 - 4BDM_1M_4 - 2BDM_2M_4 - D^2M_4^2 - 4B^2M_1^2 + B^2M_5 + 2B^2M_4 + 2B^2M_3 - 4B^2M_1M_2 - B^2M_2^2 + 2D^2N_3,$$

$$a_{18} = 2BDM_7 + B^2M_3 + D^2N_3,$$

$$a_{19} = 16BDM_7 - 8BDM_1M_4 - 4BDM_2M_4 + 8BDM_8 - 2D^2M_4^2 - 8B^2M_1^2 + 4B^2M_3 - 2B^2M_2^2 + 4ABM_4 + 4ADM_7 + 2B^2M_5 + 12B^2M_4 - 8B^2M_1M_2 + 8D^2N_3,$$

$$a_{20} = 8BDM_7 + 2B^2M_3 + 4ABM_4 + 2A^2M_4 + 2ADM_8 + 2ADM_7 + 4D^2N_3;$$

with

$$M_6 = \left[\frac{1}{48}(I_\sigma^0 - 3I_\sigma^2 + 3I_\sigma^4 - I_\sigma^6)\right], \quad M_7 = \left[\frac{1}{8}(I_\sigma^2 - 2I_\sigma^4 + I_\sigma^6)\right],$$

$$M_8 = \left[\frac{1}{2}(I_\sigma^4 - I_\sigma^6)\right], \quad M_9 = I_\sigma^6,$$

$$N_1 = \left[\frac{1}{384}(I_\sigma^0 - 4I_\sigma^2 + 6I_\sigma^4 - 4I_\sigma^6 + I_\sigma^8)\right],$$

$$N_2 = \left[\frac{1}{48}(I_\sigma^2 - 3I_\sigma^4 + 3I_\sigma^6 - I_\sigma^8)\right], \quad N_3 = \left[\frac{1}{8}(I_\sigma^4 - 2I_\sigma^6 + I_\sigma^8)\right],$$

$$N_4 = \left[\frac{1}{2}(I_\sigma^6 - I_\sigma^8)\right], \quad N_5 = I_\sigma^8,$$

and M_i $(i=1,2,3,4,5)$ are given in Eq. (9). For the quasi-longitudinal and quasi-shear waves, the inner products are not presented here due to brevity.

For the limiting case of statistical isotropy $(\sigma \to \infty)$, the inner products required to compute the attenuations given in Eq. (22) are expressed as

$$\Xi_{\ldots\hat{\mathbf{p}}\hat{\mathbf{p}}\hat{\mathbf{s}}\hat{\mathbf{s}}}^{\ldots\hat{\mathbf{p}}\hat{\mathbf{p}}\hat{\mathbf{s}}\hat{\mathbf{s}}} = \left[\frac{4}{3}T_4 + \frac{52}{15}T_7\right] + \left[4T_4 + \frac{64}{3}T_5 + 16T_6 + \frac{208}{15}T_7\right]\chi^2 + \left[\frac{16}{3}T_6 + \frac{28}{15}T_7\right]\chi^4,$$

$$\Xi_{\ldots\hat{\mathbf{p}}\hat{\mathbf{p}}\hat{\mathbf{s}}\hat{\mathbf{s}}_2}^{\ldots\hat{\mathbf{p}}\hat{\mathbf{p}}\hat{\mathbf{s}}\hat{\mathbf{s}}_2} = \Xi_{\ldots\hat{\mathbf{p}}_2\hat{\mathbf{p}}\hat{\mathbf{s}}\hat{\mathbf{s}}}^{\ldots\hat{\mathbf{p}}_2\hat{\mathbf{p}}\hat{\mathbf{s}}\hat{\mathbf{s}}} = [T_4 + 2T_7] + [4T_5 + 4T_6 + 4T_7]\chi^2,$$

$$\Xi_{\ldots\hat{\mathbf{p}}\hat{\mathbf{p}}\hat{\mathbf{s}}\hat{\mathbf{s}}_3}^{\ldots\hat{\mathbf{p}}\hat{\mathbf{p}}\hat{\mathbf{s}}\hat{\mathbf{s}}_3} = \Xi_{\ldots\hat{\mathbf{p}}_3\hat{\mathbf{p}}\hat{\mathbf{s}}\hat{\mathbf{s}}}^{\ldots\hat{\mathbf{p}}_3\hat{\mathbf{p}}\hat{\mathbf{s}}\hat{\mathbf{s}}} = [T_4 + 4T_5 + 4T_6 + 6T_7] + \left[\frac{16}{3}T_6 + \frac{28}{15}T_7\right]\chi^2 + \left[\frac{16}{3}T_6 + \frac{28}{15}T_7\right]\chi^4,$$

$$\Xi_{\ldots\hat{\mathbf{p}}_2\hat{\mathbf{p}}\hat{\mathbf{s}}\hat{\mathbf{s}}_2}^{\ldots\hat{\mathbf{p}}_2\hat{\mathbf{p}}\hat{\mathbf{s}}\hat{\mathbf{s}}_2} = [4T_6 + 4T_7] + [T_3 + 3T_6 + 2T_7]\chi^2,$$

$$\Xi_{\ldots\hat{\mathbf{p}}_3\hat{\mathbf{p}}\hat{\mathbf{s}}\hat{\mathbf{s}}_3}^{\ldots\hat{\mathbf{p}}_3\hat{\mathbf{p}}\hat{\mathbf{s}}\hat{\mathbf{s}}_3} = \left[\frac{4}{3}T_6 + \frac{52}{15}T_7\right] - \left[\frac{16}{3}T_6 + \frac{28}{15}T_7\right]\chi^2 + \left[\frac{16}{3}T_6 + \frac{14}{15}T_7\right]\chi^4,$$

$$\Xi_{\ldots\hat{\mathbf{p}}_2\hat{\mathbf{p}}\hat{\mathbf{s}}\hat{\mathbf{s}}_3}^{\ldots\hat{\mathbf{p}}_3\hat{\mathbf{p}}\hat{\mathbf{s}}\hat{\mathbf{s}}_2} = \Xi_{\ldots\hat{\mathbf{p}}_3\hat{\mathbf{p}}\hat{\mathbf{s}}\hat{\mathbf{s}}_2}^{\ldots\hat{\mathbf{p}}_2\hat{\mathbf{p}}\hat{\mathbf{s}}\hat{\mathbf{s}}_3} = [T_6 + 2T_7],$$

where

$$T_1 = -\frac{8}{45}A^2 - \frac{16}{315}AD - \frac{16}{4725}D^2,$$

$$T_2 = -\frac{8}{315}AD - \frac{8}{315}BD - \frac{8}{45}AB - \frac{16}{4725}D^2,$$

$$T_3 = -\frac{8}{45}B^2 - \frac{16}{315}BD - \frac{16}{4725}D^2,$$

$$T_4 = \frac{A^2}{15} + \frac{2}{105}AD + \frac{D^2}{945},$$

$$T_5 = \frac{AB}{15} + \frac{AD}{105} + \frac{BD}{105} + \frac{D^2}{945},$$

$$T_6 = \frac{B^2}{15} + \frac{2BD}{105} + \frac{D^2}{945},$$

$$T_7 = \frac{D^2}{945}.$$

Ahmed, S., and Thompson, R. B. (**1992**). "Effects of preferred grain orientation and grain elongation on ultrasonic wave propagation in stainless steel," in *Review of Progress in Quantitative NDE*, Vol. **11**, 1999–2006, edited by D. O. Thompson, and D. E. Chimenti (Plenum, New York).

Ahmed, S., and Thompson, R. B. (**1996**). "Propagation of elastic waves in equiaxed stainless-steel polycrystals with aligned [001] axes," J. Acoust. Soc. Am. **99**, 2086–2096.

Fedorov, A. F. (**1968**). *Theory of Elastic Waves in Crystals* (Plenum, New York).

Hirsekorn, S. (**1982**). "The scattering of ultrasonic waves by polycrystals," J. Acoust. Soc. Am. **72**, 1021–1031.

Hirsekorn, S. (**1983**). "The scattering of ultrasonic waves by polycrystals. II. Shear waves," J. Acoust. Soc. Am. **73**, 1160–1163.

Hirsekorn, S. (**1985**). "The scattering of ultrasonic waves in polycrystalline materials with texture," J. Acoust. Soc. Am. **77**, 832–843.

Hirsekorn, S. (**1986**). "Directional dependence of ultrasonic propagation in textured polycrystals," J. Acoust. Soc. Am. **79**, 1269–1279.

Li, J. Y. (**2000**). "The effective electroelastic moduli of textured piezoelectric polycrystalline aggregates," J. Mech. Phys. Solids **48**, 529–552.

Li, Y., and Thompson, R. B. (**1990**). "Relations between elastic constants C_{ij} and texture parameters for hexagonal materials," J. Appl. Phys. **67**, 2663–2665.

Stanke, F. E., and Kino, G. S. (**1984**). "A unified theory for elastic wave propagation in polycrystalline materials," J. Acoust. Soc. Am. **75**, 665–681.

Turner, J. A. (**1999**). "Elastic wave propagation and scattering in heterogeneous, anisotropic media: Textured polycrystalline materials," J. Acoust. Soc. Am. **106**, 541–552.

Weaver, R. L. (**1990**). "Diffusion of ultrasound in polycrystals," J. Mech. Phys. Solids **38**, 55–86.

Yang, L., and Turner, J. A. (**2004**). "Attenuation of ultrasonic waves in rolled metals," J. Acoust. Soc. Am. **116**, 3319–3327.

Modeling the excitation of guided waves in generally anisotropic multilayered media

Alexander Velichko[a] and Paul D. Wilcox[b]
Department of Mechanical Engineering, University of Bristol, Bristol, BS8 1TR, United Kingdom

(Received 1 May 2006; revised 16 October 2006; accepted 17 October 2006)

The design of transducers to excite and detect guided waves is a fundamental part of a nondestructive evaluation or structural health monitoring system and requires the ability to predict the radiated guided wave field of a transmitting transducer. For most transducers, this can be performed by making the assumption that the transducer is weakly coupled and then integrating the Green's function of the structure over the area of the transducer. The majority of guided wave modeling is based on two-dimensional (2D) formulations where plane, straight-crested waves are modeled. Several techniques can be readily applied to obtain the solution to the forced 2D problem in terms of modal amplitudes. However, for transducer modeling it is desirable to obtain the complete three-dimensional (3D) field, which is particularly challenging in anisotropic materials. In this paper, a technique for obtaining a far-field asymptotic solution to the 3D Green's function in terms of the modal solutions to the forced 2D problem is presented. Results are shown that illustrate the application of the technique to isotropic (aluminium) and anisotropic (cross-ply and unidirectional composite) plates. Where possible, results from the asymptotic model are compared to those from 3D time-marching finite element simulations and good agreement is demonstrated.
 [DOI: 10.1121/1.2390674]

PACS number(s): 43.20.Bi, 43.35.Cg, 43.35.Zc [RMW] Pages: 60–69

I. INTRODUCTION

The analysis of guided waves in multilayered media has been the subject of a considerable amount of research for over a century. The first solutions of the unforced modal problem considered a two-dimensional (2D) cross section through the waveguide. In this formulation, the media are assumed to be in a state of plane strain and the guided wave modes predicted are plane, straight-crested waves with wave fronts perpendicular to the cross section. In the current paper, the plane-strain formulation for straight-crested guided waves is referred to as the 2D formulation. Much research has been devoted to analyzing the dispersion relationships for guided waves using the 2D formulation and a number of methods of solution have been developed including global and transfer matrix methods[1–3] and semianalytical finite element (SAFE) methods.[4,5] Numerical solutions using some of these methods are well established and commercially available.[6] Although less well known, the tools for predicting the amplitude of excited guided waves based on a 2D formulation are also well established. For example, the forced problem can be solved directly by using integral transforms,[7–9] the SAFE method[10] or by using modal expansion and the principle of reciprocity.[11,12]

In practice the modal solution obtained from a 2D formulation provides an adequate basis for understanding many aspects of wave propagation in real three-dimensional (3D) structures. However, the 2D formulation is a much less satisfactory basis for modeling the radiated guided wave field from a finite sized transducer, since the 2D formulation inherently requires the force distribution to extend infinitely in the plane perpendicular to the cross section. To accurately model a transducer a 3D formulation is required. This has been addressed by a number of researchers for specific cases. There are several approaches to finding the solution of the 3D forced problem. For example, the 3D wave field due to surface load can be calculated by using multiple integral transforms coupled with matrix methods for isotropic[13] and anisotropic[14] materials or using a modal expansion method.[15] It is also possible to use the finite element method[16] or other numerical methods.[17] In some particular cases the analytical expressions for the 3D solutions can be obtained. For example, for an isotropic plate and axisymmetric normal surface loading such formulas have been presented by Ditri *et al.*[18]

In this paper, the 3D Green's function is written in such a way that its far-field asymptotic solution can be expressed in terms of the modal expansion of a forced 2D system, which, as previously noted, can be obtained by a number of established methods. A technique is therefore provided for numerically computing the 3D excited guided wave field from a finite sized transducer using only the dispersion relationships and mode shapes obtained from 2D formulations. For the case of an isotropic plate such a method has been presented by Wilcox.[19] In the recent paper by Moulin *et al.*[20] the particular case of normal surface force on the isotropic plate is considered. Based on existing analytical expressions for solutions to the 3D and 2D problems, the authors then derived the relationship between them. However, analytical solutions are available for only a few specific cases, and direct comparison between 3D and 2D solutions becomes impossible for the case of anisotropic layered media. The

[a]Author to whom correspondence should be addressed. Electronic mail: a.velichko@bristol.ac.uk
[b]Electronic mail: p.wilcox@bristol.ac.uk

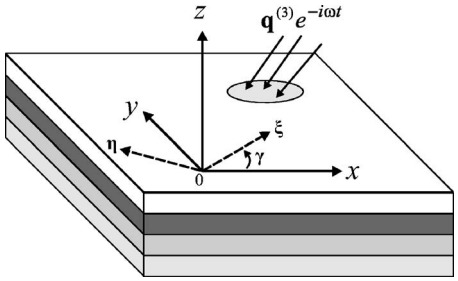


FIG. 1. System geometry.

technique proposed here is applicable to generally anisotropic layered media, although the relationship between 2D and 3D solutions is more complex than in isotropic or transversely isotropic layered media.

II. THEORY

A. Formulation of general 3D problem

A planar multilayered system consisting of N generally anisotropic layers is considered with Cartesian coordinate axes (x,y,z) defined with the z axis normal to the plane of the layers. An arbitrary time harmonic load $\mathbf{q}^{(3)}\ e^{-i\omega t}$ is applied to the upper surface of the system at $z=0$. The system is illustrated schematically in Fig. 1. The resulting time-harmonic displacement field in the system due to $\mathbf{q}^{(3)}$ is denoted by $\mathbf{u}^{(3)}$. The function $\mathbf{u}^{(3)}$ is related to $\mathbf{q}^{(3)}$ by the 3D Green's function $\mathbf{g}^{(3)}(x,y,z)$:

$$\mathbf{u}^{(3)}(x,y,z,\mathbf{q}^{(3)}) = \int\int \mathbf{g}^{(3)}(x-x',y-y',z) \times \mathbf{q}^{(3)}(x',y')dx'dy'. \quad (1)$$

The Green's function, $\mathbf{g}^{(3)}$, can be written in terms of its 2D spatial Fourier transform, $\mathbf{G}^{(3)}$, as

$$\mathbf{g}^{(3)}(x,y,z) = \frac{1}{4\pi^2}\int\int \mathbf{G}^{(3)}(k_x,k_y,z)e^{i(k_x x+k_y y)}dk_x dk_y, \quad (2)$$

where the matrix $\mathbf{G}^{(3)}(k_x,k_y,z)$ is the Green's function for straight-crested waves propagating in the direction given by the components k_x,k_y of the wave vector.

B. 2D problem

A new coordinate system (ξ,η,z) is defined that is a rotation of the original coordinate system by an angle γ about the vertical axis z (Fig. 1).

A special case of $\mathbf{q}^{(3)}$ may be defined as $\mathbf{q}^{(2)}(\xi)$ which is invariant in the η direction. The displacement field due to this loading is defined as $\mathbf{u}^{(2)}(\gamma,\xi,z,\mathbf{q}^{(2)})$. The loading $\mathbf{q}^{(2)}$ and displacement $\mathbf{u}^{(2)}$ represent the case of 2D excitation. The relationship between $\mathbf{u}^{(2)}$ and $\mathbf{q}^{(2)}$ may be written as the convolution integral

$$\mathbf{u}^{(2)}(\gamma,\xi,z,\mathbf{q}^{(2)}) = \int \mathbf{g}^{(2)}(\gamma,\xi-\xi',z)\mathbf{q}^{(2)}(\xi')d\xi', \quad (3)$$

where $\mathbf{g}^{(2)}$ is the 2D Green's function; $\mathbf{g}^{(2)}$ can be written in terms of its one-dimensional spatial Fourier transform, $\mathbf{G}^{(2)}$, as

$$\mathbf{g}^{(2)}(\gamma,\xi,z) = \frac{1}{2\pi}\int_{-\infty}^{+\infty}\mathbf{G}^{(2)}(\gamma,k,z)e^{ik\xi}dk. \quad (4)$$

By using the residues technique, the integration in Eq. (4) can be reduced to the sum of residuals

$$\mathbf{g}^{(2)}(\gamma,\xi,z) = \sum_m \mathbf{E}_m^{(2)}(\gamma,z)e^{ik_m\xi},$$
$$\mathbf{E}_m^{(2)}(\gamma,z) = i\ \mathrm{res}\ \mathbf{G}^{(2)}(\gamma,k,z)\big|_{k=k_m}, \quad (5)$$

where $\mathbf{E}_m^{(2)}$ is defined as the 2D modal excitability matrix.

The real poles of the matrix $\mathbf{G}^{(2)}$ represent the propagating waves while the complex poles represent nonpropagating waves that decay exponentially with propagation distance from the source. In this paper only the contributions from real poles are considered.

To solve the 2D problem using the described integral transforms method it is necessary to calculate matrix $\mathbf{G}^{(2)}$. The modal solution of the forced 2D problem can also be obtained by a number of other established methods. For example, the reciprocity approach[11] leads to an alternative expression for mode amplitude $\mathbf{E}_m^{(2)}$ that is useful because it can be computed directly from the mode shape. The expression for propagating mode amplitude (see, for example, Núñez *et al.*[21]) is

$$\mathbf{E}_m^{(2)}(\gamma,z) = \frac{i\omega}{4P_m}\mathbf{u}_m(\gamma,z)\cdot\mathbf{u}_m^{*T}(\gamma,z=0), \quad (6)$$

where * denotes complex conjugation, T represents transpose and $\mathbf{u}_m=(u_{mx},u_{my},u_{mz})^T$ is the displacement field distribution for mth mode. Coefficient P_m is the average power flow of the mode, given by[11]

$$P_m = \frac{\omega}{2}\ \mathrm{Im}\int(\mathbf{T}_m\mathbf{u}_m^*)\mathbf{n}dz, \quad (7)$$

where $\mathbf{T}_m$ is the stress tensor and $\mathbf{n}=(1,0,0)^T$ is the direction of mode propagation.

C. Relation between 2D and 3D problems for straight-crested waves

Consider now the relationship between 2D and 3D Green's functions for straight-crested waves $\mathbf{G}^{(2)}$ and $\mathbf{G}^{(3)}$.

The transformation from the coordinate system (x,y,z) to the new coordinate system (ξ,η,z) is represented by the matrix $\mathbf{A}$

$$\begin{pmatrix}\xi\\ \eta\\ z\end{pmatrix} = \mathbf{A}(\gamma)\begin{pmatrix}x\\ y\\ z\end{pmatrix}, \quad \mathbf{A}(\gamma) = \begin{pmatrix}\cos\gamma & \sin\gamma & 0\\ -\sin\gamma & \cos\gamma & 0\\ 0 & 0 & 1\end{pmatrix}. \quad (8)$$

The Fourier variables, k_ξ and k_η, in the new coordinate system are related to those in the original coordinate system by

$$k_x = k_\xi\cos\gamma - k_\eta\sin\gamma, \quad k_y = k_\xi\sin\gamma + k_\eta\cos\gamma \quad (9)$$

The Green's matrix in the Fourier domain, $\mathbf{G}^{(3)}\times(k_x,k_y,z)$, is transformed as

$$\mathbf{G}_{\gamma}^{(3)}(k_{\xi},k_{\eta},z)=\mathbf{A}\mathbf{G}^{(3)}(k_{\xi}\cos\gamma-k_{\eta}\sin\gamma,k_{\xi}\sin\gamma+k_{\eta}\cos\gamma,z)\mathbf{A}^{-1}. \tag{10}$$

Here $\mathbf{G}_{\gamma}^{(3)}(k_{\xi},k_{\eta},z)$ is the Green's function for straight-crested wave with wave vector $(k_{\xi},k_{\eta},0)$. On the other hand, $\mathbf{G}^{(2)}(\gamma,k,z)$ is Green's function for straight-crested wave with wave vector $(k,0,0)$. Hence,

$$\mathbf{G}^{(2)}(\gamma,k,z)=\mathbf{G}_{\gamma}^{(3)}(k,0,z),$$

or

$$\mathbf{G}^{(2)}(\gamma,k,z)=\mathbf{A}\mathbf{G}^{(3)}(k\cos\gamma,k\sin\gamma,z)\mathbf{A}^{-1}. \tag{11}$$

D. Far-field asymptotic solution to 3D problem

The double wave number integral (2) can be evaluated numerically and different calculation methods have been developed.[22,23] But numerical evaluation of this type of integral is very difficult due to the presence of singularities and the high oscillation of integrand in the far-field zone. An alternative approach is asymptotical analysis of the integral, which is used in this paper.

Expression (5) gives the expansion of the 2D solution in terms of 2D modes. In this section the analogous mode expansion of the 3D solution in the far-field will be derived. As in the 2D case the mode amplitudes are proportional to the residuals for real poles of the matrix $\mathbf{G}^{(3)}$. Then by using Eq. (11) it is possible to obtain the relationship between mode amplitudes in 2D and 3D cases.

First a change of global coordinates from Cartesian (x,y,z) to cylindrical polar (r,φ,z) is defined:

$$x=r\cos\varphi,\quad y=r\sin\varphi,\quad k_x=k\cos\gamma,\quad k_y=k\sin\gamma.$$

The expression (2) for the 3D Green's function $\mathbf{g}^{(3)}$ can therefore be written in (r,φ,z) coordinates as:

$$\mathbf{g}^{(3)}(r,\varphi,z)=\frac{1}{4\pi^2}\int_{\varphi-\pi/2}^{\varphi+3\pi/2}\int_{\Gamma}\mathbf{G}^{(3)}(k\cos\gamma,\ k\sin\gamma,z)\times e^{ikr\cos(\gamma-\varphi)}k\,dk\,d\gamma. \tag{12}$$

The contour of integration Γ coincides with the real positive half axis except for real poles. In these points it diverges in the complex plane k in accordance with the principle of limiting absorption.[7,13]

The integral with respect to γ is divided into two parts: from $\varphi-\pi/2$ to $\varphi+\pi/2$ and from $\varphi+\pi/2$ to $\varphi+3\pi/2$. In the second integral γ is changed to $\gamma+\pi$ and k to $-k$. Then

$$\mathbf{g}^{(3)}(r,\varphi,z)=\frac{1}{4\pi^2}\int_{\varphi-\pi/2}^{\varphi+\pi/2}\int_{\Gamma\cup-\Gamma}\mathbf{G}^{(3)}(k\cos\gamma,\ k\sin\gamma,z)\times e^{ikr\cos(\gamma-\varphi)}k\,dk\,d\gamma. \tag{13}$$

The integration with respect to k can be performed by using the residues theory. The contours Γ and $-\Gamma$ can be closed in the upper half plane of k and the integrals are reduced to the sum of residues of the poles and integrals along the imaginary axis. Propagating modes are described by the real poles only and decrease as $r^{-1/2}$ as $r\to\infty$. The contributions of the residues in complex poles and integrals along the imaginary axis have a larger order of decrease. Therefore,

$$\mathbf{g}^{(3)}=\frac{1}{2\pi}\sum_{m}\int_{\varphi-\pi/2}^{\varphi+\pi/2}\mathbf{G}_{m}^{(3)}(\gamma,z)|k_m(\gamma)|e^{ik_m(\gamma)r\cos(\gamma-\varphi)}d\gamma+O(r^{-3/2}),\quad r\to\infty,$$

$$\mathbf{G}_{m}^{(3)}(\gamma,z)=i\ \mathrm{res}\,\mathbf{G}^{(3)}(k\cos\gamma,k\sin\gamma,z)\big|_{k=k_m(\gamma)}. \tag{14}$$

Integrals with respect to γ are then calculated using the stationary phase method. This enables the final expression for the far-field asymptote to the 3D problem to be written as:

$$\mathbf{g}^{(3)}(r,\varphi,z)=\frac{1}{\sqrt{r}}\sum_{m}\mathbf{E}_{m}^{(3)}(\varphi,z)e^{ir\Phi_m(\gamma_m,\varphi)}+O(r^{-3/2}),\quad r\to\infty,$$

$$\mathbf{E}_{m}^{(3)}(\varphi,z)=B_m(\varphi)\mathbf{G}_{m}^{(3)}(\gamma_m,z), \tag{15}$$

where $\mathbf{E}_{m}^{(3)}$ is defined as the 3D modal excitability matrix.

The phase functions Φ_m and coefficients B_m are given by

$$\Phi_m(\gamma,\varphi)=k_m(\gamma)\cos(\gamma-\varphi),$$

$$B_m=\frac{|k_m(\gamma_m)|}{\sqrt{2\pi|b_m|}}\exp\left\{i\frac{\pi}{4}\mathrm{sgn}\,b_m\right\},\quad b_m=\frac{\partial^2\Phi_m(\gamma_m,\varphi)}{\partial\gamma^2}. \tag{16}$$

The angle $\gamma_m\equiv\gamma_m(\varphi)$ is found from the following equation:

$$\frac{\partial\Phi_m(\gamma,\varphi)}{\partial\gamma}=0,\quad \varphi-\frac{\pi}{2}\le\gamma\le\varphi+\frac{\pi}{2}. \tag{17}$$

The angles φ and γ_m are related by

$$\varphi=\gamma_m+\arctan\frac{c'(\gamma_m)}{c(\gamma_m)}, \tag{18}$$

here $c(\gamma)=\omega/k_m(\gamma)$ is phase velocity.

E. Relation between 2D and 3D problems

In summary, the 2D Green's function is

$$\mathbf{g}^{(2)}(\gamma,\xi,z)=\sum_{m}\mathbf{E}_{m}^{(2)}(\gamma,z)e^{ik_m\xi},$$

and the far-field asymptotic to the 3D Green's function is

$$\mathbf{g}^{(3)}(r,\varphi,z)=\frac{1}{\sqrt{r}}\sum_{m}\mathbf{E}_{m}^{(3)}(\varphi,z)e^{ir\Phi_m(\gamma_m,\varphi)}.$$

Using expression (11), the modal excitability matrices in the 2D and 3D cases can be related by

$$\mathbf{E}_{m}^{(3)}(\varphi,z)=B_m(\varphi)\mathbf{A}^{-1}(\gamma_m)\mathbf{E}_{m}^{(2)}(\gamma_m,z)\mathbf{A}(\gamma_m). \tag{19}$$

The far-field solution for a particular mode in the φ direction is therefore intimately related to the appropriate 2D solution for the same mode in the γ_m direction.

The phase function Φ_m can be written as $\Phi_m=\mathbf{k}_m\cdot\mathbf{n}$. In this expression $\mathbf{k}_m=(k_m(\gamma_m)\cos\gamma_m,k_m(\gamma_m)\sin\gamma_m)^T$ is a wave

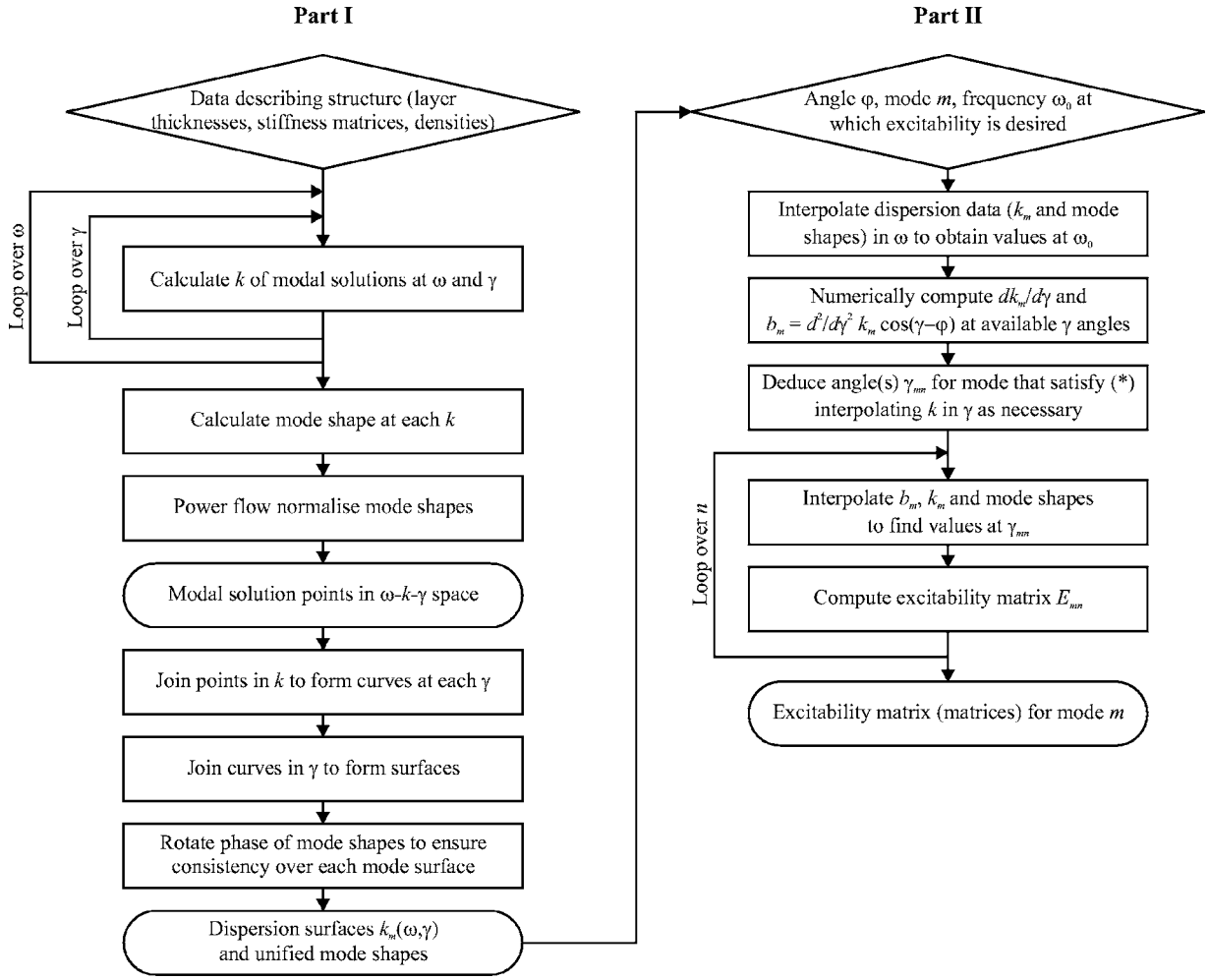


FIG. 2. Flow chart of excitability calculation.

vector and $\mathbf{n}=(\cos\varphi,\sin\varphi)^T$ is a unit vector in φ direction. Therefore, the angle(s) γ_m is the phase velocity direction(s) for the mth mode. Moreover, the relation (18) shows that vector $\mathbf{n}$ is normal to the slowness surface, $\mathbf{k}_m(\gamma)/\omega$, at the direction γ_m. It means that angle φ is direction of group velocity for the mth mode.[11]

Note that the coefficients b_m in expression (16) can be written in the form

$$b_m = -\frac{\omega}{c\sqrt{c^2+(c')^2}}(c''+c)\Bigg|_{\gamma=\gamma_M}. \tag{20}$$

It is supposed in the previous analysis that $b_m \neq 0$. From Eq. (18) we obtain that if $b_m=0$, then $d\varphi/d\gamma=0$. In this case the group velocity direction remains the same while the phase velocity direction varies and in such directions the wave field decays as $O(r^{-1/3})$.

The complete procedure for predicting the far-field displacement in the φ direction under point harmonic loading is as follows:

- Compute dispersion relationships $k_m(\gamma)$. For mth mode, find angle or angles γ_m.
- Solve 2D problem at angle γ_m.
- Compute the amplitude of 3D mode at angle φ from 2D solution at angle γ_m.

In the special case of an isotropic or transversely isotropic layered medium, the slowness profiles of all modes are circular, hence the phase and group velocity directions are identical. In this case, the relationship between 2D and 3D modal excitability matrices reduces to

$$\mathbf{E}_m^{(3)}(\varphi,z) = \sqrt{\frac{k_m}{2\pi}}e^{-i\pi/4}\mathbf{A}^{-1}(\varphi)\mathbf{E}_m^{(2)}(z)\mathbf{A}(\varphi). \tag{21}$$

In the case of material damping all wave numbers k_m become complex and the asymptotic of the solution given by Eq. (15) to the 3D problem is not valid. This topic requires further research.

III. IMPLEMENTATION OF ASYMPTOTIC MODEL

The asymptotic model described in the previous section has been implemented numerically using functions written in the MATLAB (The Mathworks Inc., Natick, Massachusetts) modeling environment. The overall operation of the numerical program is shown in the flow chart in Fig. 2, and can be separated into two distinct parts. The first part of the program generates dispersion data for the structure and the second part converts this into modal excitability matrices for guided wave propagation in a specified direction.

A. Dispersion curve generation

Dispersion curves are generated for different angles of propagation using a semianalytic finite element technique (implemented in MATLAB) similar to that described by Hayashi, Song, and Rose.[5] Other dispersion curve calculation techniques such as the global matrix or transfer matrix methods[2] could also be used. The semianalytic finite element method is used here to allow easy integration of the complete model in MATLAB. For each angle, γ, this technique yields a number of discrete points lying in ω-k space, where k is complex. Each point corresponds to a modal solution for the propagation of straight-crested guided waves in the γ direction. Associated with each point is a displacement mode shape, from which the stress mode shape and hence the power flow associated with any chosen amplitude of wave can be deduced. At this stage the mode shape at each point is power-flow normalized and points with imaginary k components are discarded as these correspond to nonpropagating modes. The result is a number of discrete points lying in $\omega-k-\gamma$ space and a power-flow normalized mode shape associated with each point.

The most challenging aspect of the numerical implementation is to link the discrete points in ω-k-γ space together into modes to create dispersion surfaces. This is required since it is necessary in the subsequent excitability calculation to interpolate modal data in both ω and γ and also to differentiate the wave number, k, of a mode with respect to γ [Eq. (17)]. First, points are linked at each γ angle by comparing the mode shapes at nearby points in ω-k space to form dispersion curves. Next, dispersion curves are linked between adjacent γ angles to form dispersion surfaces by comparing both mode shapes and dispersion curve shapes.

This aspect of the model is illustrated by the example in Fig. 3 which shows (a) the discrete points generated by the semianalytical finite element method and (b) the dispersion surfaces for three modes obtained by joining the points. For clarity the data in this figure have been plotted as phase-velocity rather then wave number versus frequency and angle. The dispersion data are for the cross-ply composite plate example discussed in the next section.

The final stage of the generation of dispersion data is to rotate the phase of mode shapes for each mode so that the phase is consistent at all points. This is necessary for subsequent interpolation between mode shapes of a mode in both ω and γ. The phase of mode shapes for all points on a mode are rotated so that the phase of the dominant surface component of the displacement mode shape is zero.

B. Excitability matrix calculation

Once the dispersion data are obtained for all modes, excitability matrices for any mode, m, frequency, ω_0, and propagation direction, φ, can be computed. First, the dispersion data (power flow normalized mode shape and wave number) are interpolated in ω to find its values at ω_0. The next stage of the calculation is to compute the angle or angles, γ_{mn}, which satisfy Eq. (17) for the mth mode in the φ direction. To perform this calculation, it is first necessary to numerically compute the derivatives $d/d\gamma\ [k_m(\omega_0,\gamma)]$ and

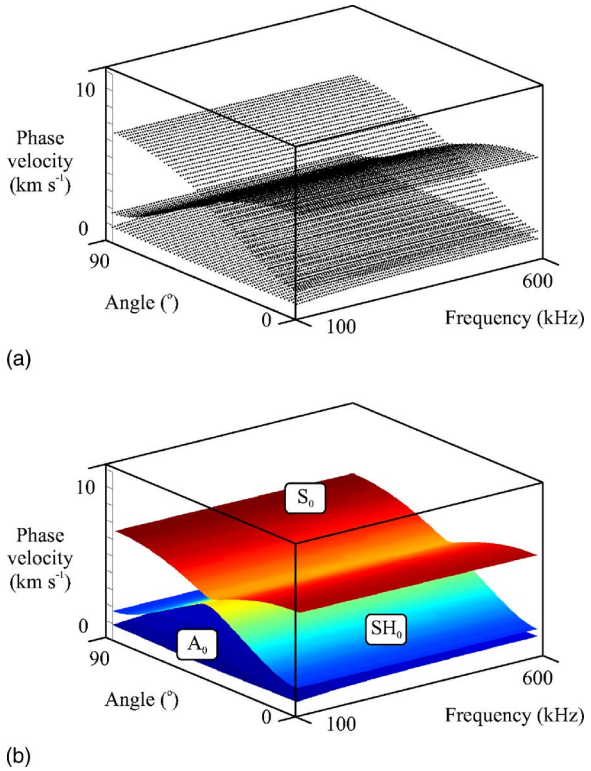


FIG. 3. (Color online) Example of dispersion data: (a) discrete points generated by the semianalytical finite element method and (b) the resulting dispersion surfaces. The data are for a cross-ply composite plate.

$d^2/d^2\gamma\ [k_m(\omega_0,\gamma)\cos(\gamma-\varphi)]$. The first of these two quantities is used to identify pairs of discrete γ angles at which the available dispersion data bracket solutions to Eq. (18). Approximations to solutions γ_{mn} are obtained by linear interpolation between these angles. If there is more than one γ_{mn}, then each must be treated separately, and in the final result this effect appears as extra modes. The dispersion data and $d^2/d^2\gamma[k_m(\omega_0,\gamma)\cos(\gamma-\varphi)]$ are then interpolated again, this time in γ, to obtain values at γ_{mn}. The 2D excitability matrix, $\mathbf{E}^{(2)}_{mn}(\omega_0,\gamma_{mn})$, is calculated by using expression (6). Finally, the 3D excitability matrix, $\mathbf{E}^{(3)}_{mn}(\omega_0,\varphi)$, is computed according to Eq. (19) and the effective wave number in the φ direction, Φ_{mn}, is calculated from Eq. (16).

The excitability matrix and effective wave number provide all information necessary to perform wave excitation simulation. Typically this may involve the simulation of either time-domain signals recorded at a particular point or the surface displacement around a source at a particular instant in time.

IV. FINITE ELEMENT MODELING

In order to validate the implementation of the three-dimensional (3D) excitability model described in the previous sections, a number of explicit time marching finite element (FE) simulations have been performed. These are also 3D and require considerable computational power. Symmetry conditions are exploited where possible, but in order to distinguish different modes and separate directly excited modes from edge reflections, a significant area of a plate-like structure must be modeled. The maximum size of FE model

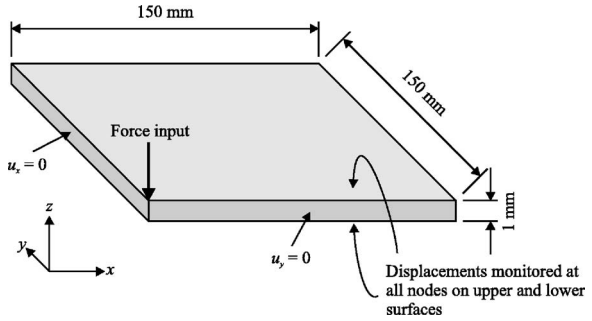


FIG. 4. Finite element (FE) model geometry.

that can be run on the computer resources available is limited to around 5×10^6 degrees of freedom and this means that a compromise must be made between the size of the structure modeled and the mesh density. For this reason, the mesh density used is somewhat less than ideal and this is manifested in an underestimate of guided wave velocity that is more pronounced for short wavelength modes. Notwithstanding these limitations, the FE results are sufficient to qualitatively show that the excitability algorithm has been correctly implemented and also to indicate the regions where the asymptotic assumptions break down.

The general FE model geometry, used for all cases considered here, is shown in Fig. 4. Cubic eight noded elements are used with side length 0.25 mm. Planes of symmetry are assigned as shown and only the response to out-of-plane forcing is considered. The force in all cases is a five cycle Hanning windowed toneburst with a center frequency of 300 kHz. All FE modeling was performed using the ABAQUS package (Version 6.5.2, ABAQUS Inc., Providence, RI) running on a Viglen CL2000, using a single Intel Xeon 32 GHz processor incorporating 64 bit PCI with a Linux operating system and 8 GB random access memory.

For comparison purposes, the displacement field due to each mode should ideally be analyzed separately. However, while the asymptotic model can be used on a mode by mode basis, the FE model implicitly includes the contributions from all guided wave modes. A rudimentary method for partially separating the contributions from different modes in the FE model, which has been employed here, is to monitor displacements at nodes on both upper and lower surfaces of the plate. This then allows the contributions to be separated into those due to symmetric and antisymmetric nodes. In the relatively low frequency thickness regime where the modeling has been performed in this paper, only three fundamental mode types exist corresponding to A_0 and S_0 Lamb-type modes and a symmetric shear-horizontal *SH*-type mode, referred to as SH_0. The modal separation technique employed therefore allows complete separation of the A_0 mode but not of the S_0 and SH_0 modes.

V. RESULTS

The results from a number of sample cases are presented here, beginning with the simplest case of guided waves in an isotropic plate excited by an out-of-plane point force. This demonstrates the procedure and also highlights the limited accuracy of the FE model. The excitation of guided waves by an out-of-plane point force in two highly anisotropic plates is then considered to illustrate the capability and limitations of the asymptotic model. The final case demonstrates the application of the asymptotic model to in-plane forcing of an isotropic plate which produces a nonaxisymmetric guided wave field. In all cases, a 1-mm-thick plate is considered and the excitation signal is a five cycle Hanning windowed toneburst with a center frequency of 300 kHz. At this frequency-thickness product the only propagating modes in all cases are the two fundamental Lamb modes (A_0 and S_0) and the fundamental shear-horizontal mode (SH_0). The reason for this choice of frequency-thickness product is to provide results that can be quantified and compared to FE results, it does not reflect any limitation of the asymptotic calculation. There is also no limitation other than increased mesh density and computation time in the semianalytic finite element method, although the subsequent connection of discrete dispersion points into dispersion surfaces becomes increasingly challenging if more modes are present.

TABLE I. Properties of aluminium plate

Property	Value	Units
Density	2700	kg m^{-3}
Young's modulus	70	MN mm^{-2}
Poisson's ratio	0.3	

A. Out-of-plane excitability of isotropic plate

The first case considered is an isotropic 1-mm-thick aluminium plate, the properties of which are shown in Table I. A_0, S_0 and SH_0 modes may exist in this system. However, the isotropy of the plate and orientation of the input force means that the problem is axisymmetric and hence only A_0 and S_0 are excited. Figures 5(a) and 5(b) show snapshots of the out-of-plane surface displacement 25 μs after the start of the input signal obtained from the FE model for the A_0 and S_0 modes. The nonaxisymmetric signal near the origin in Fig. 5(b) is due to the unwanted presence of higher order modes at the upper frequency limit of the input signal that cannot be correctly modeled by the mesh density used. Figures 5(c) and 5(d) show the equivalent results obtained from the asymptotic model. The gray scale in both images is the same, and it can be seen that the FE and asymptotic models are in excellent agreement. There is in fact a slight velocity discrepancy for the A_0 mode due to the relatively coarse mesh used in the FE model.

B. Cross-ply composite plate

A cross-ply composite plate has been modeled using equivalent homogenous properties which are listed in Table II. Again over the frequency range considered A_0, S_0 and SH_0-type modes may exist in this system. However, it should be noted that, other than in the 0° and 90° directions the mode shapes of S_0 and SH_0 both contain displacement components in directions parallel and perpendicular to the direction of propagation, hence the designation of the mode names in these directions is ambiguous. However, this at-

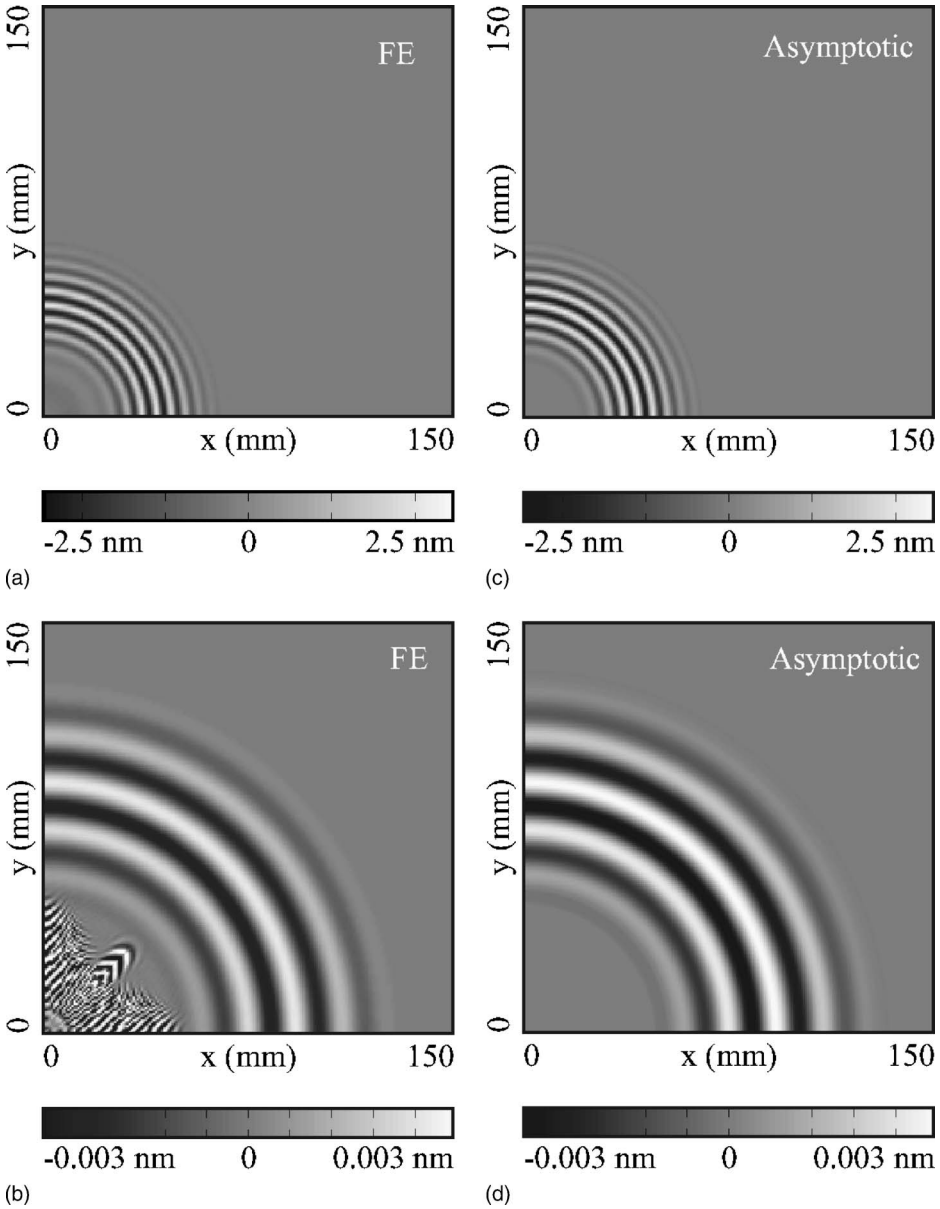


FIG. 5. Out-of-plane surface displacement of a 1-mm-thick aluminum plate, 25 μs after the start of a five cycle Hanning windowed toneburst of out-of-plane force applied at (0,0) with center frequency 300 kHz: FE model results showing contributions from (a) A_0 mode and (b) S_0 mode; asymptotic model results showing contributions from (c) A_0 mode and (d) S_0 mode.

tribute of the mode shapes means that in this system, all three modes are excited by the application of an out-of-plane point force in certain directions.

Figures 6(a) and 6(b) show snapshots of the out-of-plane surface displacement 25 μs after the start of the input signal obtained from the FE model for the anti-symmetric (A_0) and symmetric modes (S_0 and SH_0), respectively. Figures 6(c) and 6(d) show the equivalent results obtained from the asymptotic model. Figures 6(e) and 6(f) show the results from the asymptotic model separated into contributions from the S_0 and SH_0 modes.

It can be seen that the FE and asymptotic models are in reasonable agreement with regard to the overall pattern of the radiated wave field and its amplitude. Of particular interest here is the behavior of the SH_0 mode. The latter has a highly anisotropic velocity profile and over the range of

TABLE II. Properties of cross-ply composite plate.

Property	Value						Units
Density	1560						kg m^{-3}
Stiffness matrix							
$c_{11}, c_{12}, c_{13}, c_{14}, c_{15}, c_{16}$	64.24	5.6	7.73	0	0	0	MN mm^{-2}
$c_{21}, c_{22}, c_{23}, c_{24}, c_{25}, c_{26}$	5.6	70.78	8.39	0	0	0	MN mm^{-2}
$c_{31}, c_{32}, c_{33}, c_{34}, c_{35}, c_{36}$	7.73	8.39	13.3	0	0	0	MN mm^{-2}
$c_{41}, c_{42}, c_{43}, c_{44}, c_{45}, c_{46}$	0	0	0	2.97	0	0	MN mm^{-2}
$c_{51}, c_{52}, c_{53}, c_{54}, c_{55}, c_{56}$	0	0	0	0	3.06	0	MN mm^{-2}
$c_{61}, c_{62}, c_{63}, c_{64}, c_{65}, c_{66}$	0	0	0	0	0	4.7	MN mm^{-2}

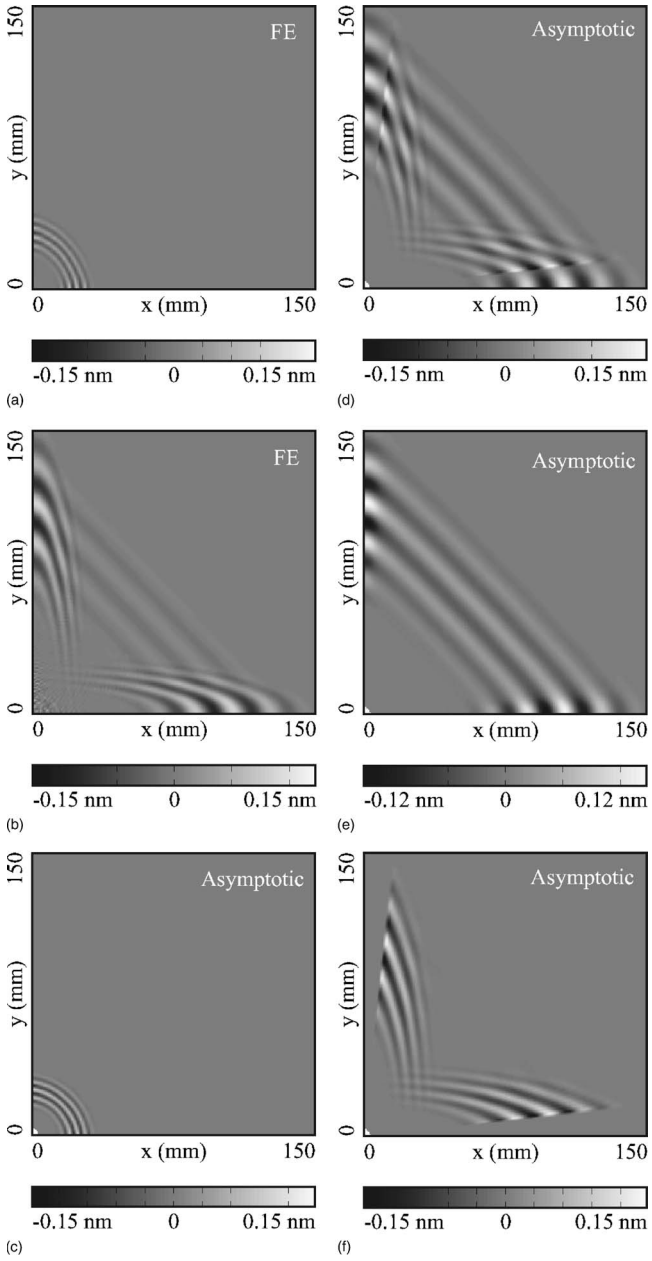


FIG. 6. Out-of-plane surface displacement of a 1-mm-thick cross-ply composite plate, 25 μs after the start of a five cycle Hanning windowed toneburst of out-of-plane force applied at $(0,0)$ with center frequency 300 kHz: FE model results showing contributions from (a) A_0 mode and (b) S_0+SH_0 modes; asymptotic model results showing contributions from (c) A_0 mode, (d) S_0+SH_0 modes, (e) S_0 mode and (f) SH_0 mode.

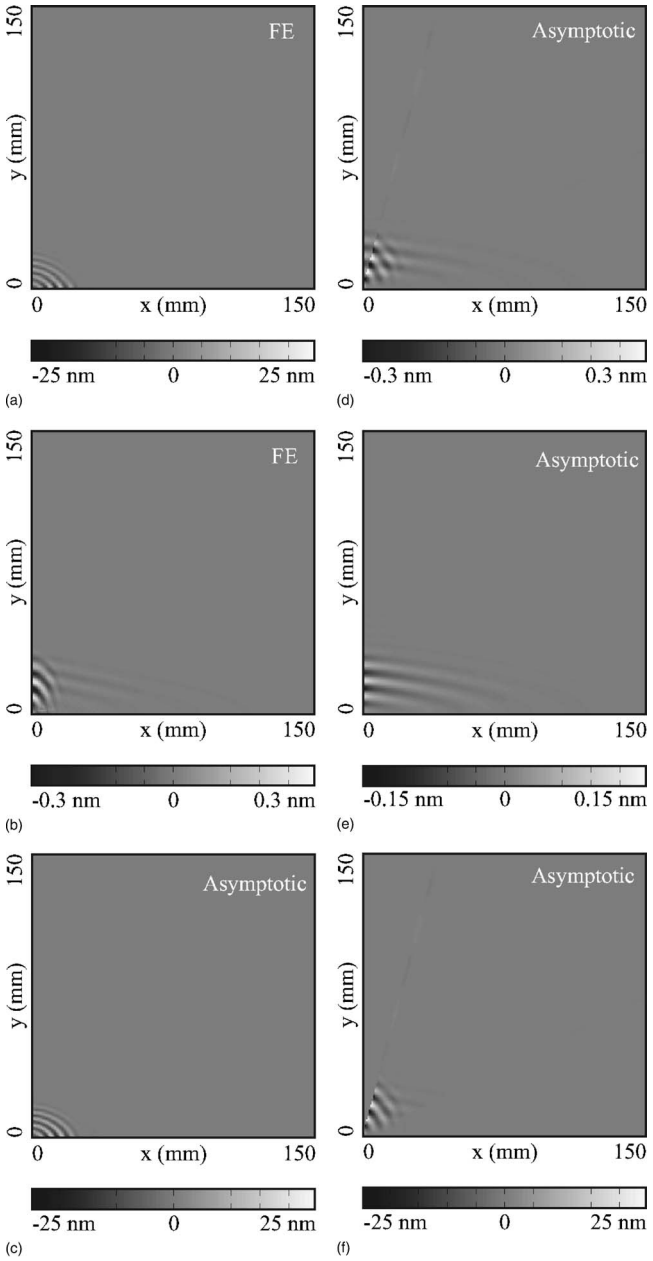


FIG. 7. Out-of-plane surface displacement of a 1-mm-thick uni-directional composite plate, 25 μs after the start of a five cycle Hanning windowed toneburst of out-of-plane force applied at $(0,0)$ with center frequency 300 kHz: FE model results showing contributions from (a) A_0 mode and (b) S_0+SH_0 modes; asymptotic model results showing contributions from (c) A_0 mode, (d) S_0+SH_0 modes, (e) S_0 mode and (f) SH_0 mode.

TABLE III. Properties of unidirectional composite plate

Property	Value						Units
Density	1560						kg m^{-3}
Stiffness matrix							
$c_{11},c_{12},c_{13},c_{14},c_{15},c_{16}$	143.8	6.2	6.2	0	0	0	MN mm^{-2}
$c_{21},c_{22},c_{23},c_{24},c_{25},c_{26}$	6.2	13.3	6.5	0	0	0	MN mm^{-2}
$c_{31},c_{32},c_{33},c_{34},c_{35},c_{36}$	6.2	6.5	13.3	0	0	0	MN mm^{-2}
$c_{41},c_{42},c_{43},c_{44},c_{45},c_{46}$	0	0	0	3.6	0	0	MN mm^{-2}
$c_{51},c_{52},c_{53},c_{54},c_{55},c_{56}$	0	0	0	0	3.6	0	MN mm^{-2}
$c_{61},c_{62},c_{63},c_{64},c_{65},c_{66}$	0	0	0	0	0	5.7	MN mm^{-2}

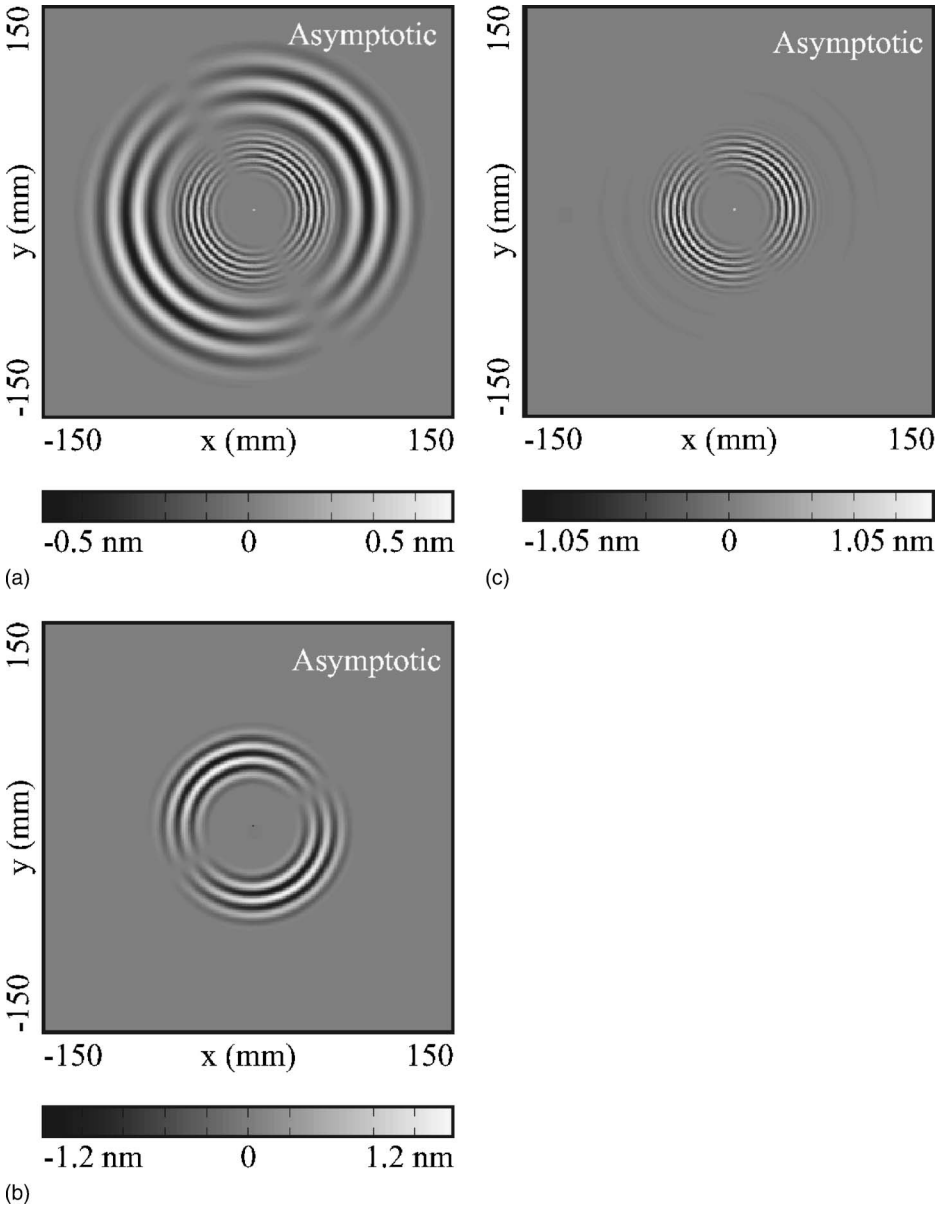


FIG. 8. Surface displacement of a 1-mm-thick aluminium plate, 25 μs after the start of a five cycle Hanning windowed toneburst of in-plane force applied at (0,0) at 30° to horizontal with center frequency 300 kHz: (a) radial displacement, (b) angular displacement and (c) out-of-plane displacement.

propagation angles from 6 to 84° has three possible values of γ_{mn}, resulting in three different wave packets (the third and fastest SH_0 wave packet propagates with a similar profile to the S_0 mode but is of low amplitude and is scarcely visible in the figure). The velocity discrepancy between the models is particularly apparent for the slower SH_0 components around 45°. At the extremities of the angular range where the SH_0 modes exist, it can be seen that there is an abrupt discontinuity in displacements. This represents the breakdown of the asymptotic approximation in Eq. (15), since at these points the coefficient b_m in expression (16) is equal to zero and the group velocity direction is stationary.

C. Unidirectional composite plate

A uni-directional composite plate has also been modeled as an example of a highly anisotropic plate. The equivalent bulk properties used for computing dispersion curves are listed in Table III. As for the cross-ply case, all three fundamental modes are excited in certain directions. Figures 7(a) and 7(b) show snapshots of the out-of-plane surface displacement obtained from the FE model 25 μs after the start of the input signal for the anti-symmetric (A_0) and symmetric modes (S_0 and SH_0), respectively. Figures 7(c) and 7(d) show the equivalent results obtained from the asymptotic model. The fiber direction in all cases is aligned with the x axis. The agreement in all cases is good with the exception of an obvious discontinuity in the symmetric modes predicted by the asymptotic model. Figures 7(e) and 7(f) show the symmetric modes predicted by the asymptotic model separated into the contributions from the S_0 and SH_0 modes, respectively. From Fig. 7(f) it is clear that the discontinuity is again due to the SH_0 mode and that in fact two discontinuities occur at the ends of the angular range over which the SH_0 mode is excited. Between these angles, the mode again has multiple components as in the case of the cross-ply plate.

D. In-plane excitation of isotropic plate

The final example is chosen to illustrate the nonaxisymmetric field excited by an in-plane point force applied to an

isotropic aluminium plate. The force is orientated at 30° to the x axis and the resulting in-plane radial, in-plane angular and out-of-plane surface displacements predicted by the asymptotic model 25 μs after the start of the input signal are shown in Figs. 8(a)–8(c), respectively. There is no accompanying FE validation in this case as the lack of symmetry requires a model with too many degrees of freedom to run on the available computer resources.

The in-plane radial and out-of-plane displacement components in this example are due entirely to the Lamb modes, A_0 and S_0, while the in-plane angular displacement is due entirely to the SH_0 mode. The modal amplitude as a function of angle with respect to the forcing direction is either sinusoidal for the SH_0 mode (i.e., maximum amplitude at right angles to forcing direction) or co-sinusoidal for the A_0 and S_0 Lamb wave modes (i.e., maximum amplitude in line with the forcing direction).

This example also illustrates a minor implementation challenge. The excitability matrices are only computed over an angular range from 0° to 90° to save time but excitability matrices may be required at any angle and a procedure is therefore required to map the available data to the desired angle. This procedure, while simple in principle, proved surprisingly awkward to implement correctly due to the need to preserve the correct sign of all nine elements in the excitability matrix in all four angular quadrants. The results in Fig. 8 show that the implementation is successful. There is continuity of all three displacement components between quadrants and the displacement components are of opposite sign on either side of the null direction for each mode.

VI. CONCLUSION

A mathematical basic of a far-field asymptotic technique for predicting the modal amplitude of the 3D guided wave field due to a harmonic point force applied to the surface of a planar multilayered anisotropic waveguide has been described. The amplitude of the displacement fields of each mode is related to the input force by modal excitability matrices which are functions of direction and frequency. A key attribute of the technique is that the excitability matrices in the 3D case are computed from the excitability matrices for the 2D case of straight-crested waves excited by line sources. The latter are readily obtained from modal dispersion data that can be computed by a number of existing methods. The numerical implementation of the technique has been discussed and practical challenges highlighted. Example results from a number of test cases have been presented which show generally good agreement with 3D time-marching finite element simulations. The points where the asymptotic assumptions are invalid are clearly visible in these results and relate to the points where the normal to the phase slowness surface of a mode (i.e., the group velocity direction) is stationary. In principle a more accurate asymptotic approximation at the vicinity of these points could be obtained but this has not yet been implemented.

[1] S. Lih and A. K. Mal, "On the accuracy of approximate plate theories for wave field calculations in composite laminates," Wave Motion **21**, 17–34 (1995).
[2] M. J. S. Lowe, "Matrix techniques for modeling ultrasonic waves in multilayered media," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **42**, 525–542 (1995).
[3] L. Wang and S. I. Rokhlin, "Stable reformulation of transfer matrix method for wave propagation in layered anisotropic media," Ultrasonics **39**, 413–424 (2001).
[4] L. Gavric, "Computation of propagative waves in free rail using a finite element technique," J. Sound Vib. **185**, 531–543 (1995).
[5] T. Hayashi, W.-J. Song, and J. L. Rose, "Guided wave dispersion curves for a bar with an arbitrary cross-section, a rod and rail example," Ultrasonics **41**, 175–183 (2003).
[6] B. Pavlakovic, M. Lowe, D. Alleyne, and P. Cawley, "Disperse: A general purpose program for creating dispersion curves," in *Review of Progress in Quantitative NDE*, edited by D. Thompson and D. Chimenti (Plenum, New York, 1997), Vol. **16**, pp. 185–192.
[7] E. Glushkov and N. Glushkova, "Blocking property of energy vortices in elastic waveguides," J. Acoust. Soc. Am. **102**, 1356–1360 (1997).
[8] W. A. Green and E. R. Green, "Elastic waves in fiber composite laminates," Ultrasonics **38**, 228–231 (2000).
[9] A. J. Niklasson and S. K. Datta, "Transient ultrasonic waves in multilayered superconducting plates," J. Appl. Mech. **69**, 811–818 (2002).
[10] M. Veidta, T. Liub, and S. Kitipornchai, "Modeling of lamb waves in composite laminated plates excited by interdigital transducers," NDT & E Int. **35**, 437–447 (2002).
[11] B. A. Auld, *Acoustic Fields and Waves in Solid*, 2nd ed. (Krieger, Malabar, FL, 1990).
[12] J. J. Ditri and J. L. Rose, "Excitation of guided waves in generally anisotropic layers using finite sources," J. Appl. Mech. **61**, 330–338 (1994).
[13] V. A. Babeshko, E. V. Glushkov, and N. V. Glushkova, "Energy vortices and backward fluxes in elastic waveguides," Wave Motion **16**, 183–192 (1992).
[14] A. Mal, "Elastic waves from localized sources in composite laminates," Int. J. Solids Struct. **39**, 5481–5494 (2002).
[15] J. D. Achenbach and Y. Xu, "Wave motion in an isotropic elastic layer generated by a time harmonic point load of arbitrary direction," J. Acoust. Soc. Am. **106**, 83–90 (1999).
[16] O. Diligent and M. J. S. Lowe, "Reflection of the S_0 lamb wave mode from a flat bottom circular hole," J. Acoust. Soc. Am. **118**, 2869–2879 (2005).
[17] D. E. Chimenti, "Guided waves in plates and their use in materials characterization," Appl. Mech. Rev. **50**, 247–284 (1997).
[18] J. J. Ditri, A. Pilarski, B. Pavlakovic, and J. L. Rose, "Generation of guided waves in a plate by axisymmetric normal surface loading," in *Review of Progress in Quantitative NDE*, edited by D. Thompson and D. Chimenti (1994), Vol. **13**, pp. 133–140.
[19] P. Wilcox, "Modeling the excitation of Lamb and SH waves by point and line sources," in *Review of Progress in Quantitative NDE*, edited by D. Thompson and D. Chimenti (AIP, New York, 2004), Vol. **23**, pp. 206–213.
[20] E. Moulin, S. Grondel, M. Baouahi, and J. Assaad "Pseudo-3D modeling of a surface-bonded Lamb wave source," J. Acoust. Soc. Am. **119**, 2575–2578 (2006).
[21] I. Núñez, R. K. Ing, C. Negreira, and M. Fink "Transfer and Green functions based on modal analysis for Lamb waves generation," J. Acoust. Soc. Am. **107**, 2370–2378 (2000).
[22] E. R. Green and W. A. Green, "A computational method for wave propagation from a point load in anisotropic material," Ultrasonics **38**, 262–266 (2000).
[23] S. Banerjee, W. Prosser, and A. Mal, "Calculation of the response of a composite plate to localized dynamic surface loads using a new wave number integral method," J. Appl. Mech. **72**, 18–24 (2005).

Coherent backscattering and far-field beamforming in acoustics

Alexandre Aubry[a] and Arnaud Derode[b]
Laboratoire Ondes et Acoustique, ESPCI, Université Denis Diderot, Paris VII, CNRS UMR 7587, 10 rue Vauquelin, 75005 Paris, France

Philippe Roux
Laboratoire de Géophysique Interne et Tectonophysique, Observatoire de Grenoble, Université Joseph Fourier and CNRS UMR 5559, 1381 rue de la Piscine BP 53 - 38041 Grenoble Cedex 9, France

Arnaud Tourin
Laboratoire Ondes et Acoustique, ESPCI, Université Denis Diderot, Paris VII, CNRS UMR 7587, 10 rue Vauquelin, 75005 Paris, France

(Received 20 July 2006; revised 23 October 2006; accepted 24 October 2006)

Coherent backscattering of waves by a random medium is spectacular evidence of interference effects despite disorder and multiple scattering. It manifests itself as a doubling of the wave intensity reflected exactly in the backward direction. This phenomenon has been observed experimentally in optics, acoustics, or seismology. While optical measurements are realized in far-field conditions with a plane wave illumination and a beamwidth much larger than the wavelength, ultrasonic experiments are carried out with wideband controllable arrays of (nearly) pointlike transducers that directly record the wave field, in amplitude and phase. Therefore it is possible to perform beamforming of the incoming and outgoing wave fields before computing the average backscattered intensity. In this paper, the advantages of plane wave beamforming applied to the study of the coherent backscattering effect are shown. Particularly, the angular resolution, the signal-to-noise ratio, as well as the estimation of the enhancement factor can be improved by beamforming. Experimental results are presented with ultrasonic pulses, in the 2.5–3.5 MHz range, propagating in random collections of scatterers. Since the coherent backscattering effect can be taken advantage of to measure diffusive parameters (transport mean free path, diffusion constant), plane-wave beamforming can be applied to the characterization of highly scattering media. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2400662]

PACS number(s): 43.20.Fn, 43.40.Hb [EJS] Pages: 70–77

I. INTRODUCTION

The coherent backscattering effect corresponds to an enhancement (by a factor of 2) in the intensity of waves scattered in the backward direction from a disordered medium. This phenomenon, also known as weak localization, originates from a constructive interference between a wave traveling a multiple-scattering path and its time-reversed counterpart: it appears when multiple scattering occurs and the reciprocity symmetry is preserved. For the last 25 years or so, enhanced backscattering has attracted a great deal of attention. In 1984, Ishimaru and Kuga[1] reported the first observation of coherent backscattering in optical experiments. Since then, a lot of theoretical and experimental efforts[2–16] have been undertaken to understand this effect in detail; while the first observations were carried out with optical waves, coherent backscattering was later observed with other types of waves, and applied to a variety of contexts ranging from light scattering on cold atoms to geophysics or ultrasound. By and large, the acoustic and seismic experimental procedures reproduced the optical experiments: The average backscattered intensity was directly measured as a function of the source-receiver distance, or backscattering angle. An important advantage of seismic or acoustic waves was the possibility to observe the time dependence of the enhanced backscattering spot.[10–12]

However, acoustic waves offer more than that: since ultrasonic transducers are controllable and can directly record time-dependent fluctuations in both amplitude and phase, one can "play with the wave field" in emission and reception, i.e., perform beamforming prior to calculating the intensity. The main purpose of this article is to illustrate the advantages of plane-wave beamforming in the study of the coherent backscattering effect. Particularly, we will show that it can improve the angular resolution as well as the signal-to-noise ratio.

Moreover, another important aspect is that optical measurements are realized in far-field conditions with an incoming beam whose width is much larger than the wavelength λ. On the contrary, in seismology, sources and detectors sit on the surface of the Earth, i.e., in the near field.[12–14] In such conditions, the width of the coherent backscattering peak depends only on λ and does not display any time dependence. Ultrasonic laboratory measurements correspond to an intermediate configuration between experiments in optics and seismology. Indeed, the typical experimental setup consists in an array of transducers placed in front of a random scat-

[a]Electronic mail: alexandre.aubry@espci.fr
[b]Electronic mail: arnaud.derode@espci.fr

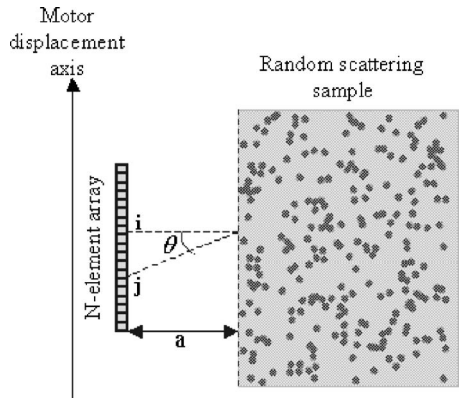


FIG. 1. Experimental setup used in our experiments: An N-element linear array is placed in front of a random scattering sample immersed in a water tank. The array is parallel to the scattering slab. A motor allows a lateral displacement of the array, in order to average the backscattering intensity over several configurations of disorder.

tering sample at a distance a[11] (see Fig. 1). Like in seismology, one element of the array is used as a pointlike source; the backscattered wave field is recorded and its intensity is calculated at every receiver (including the initial source element). Averaging can be performed by changing the source element and repeating the same experiment. In far-field conditions, the dynamics of the coherent backscattering effect can be taken advantage of to measure the diffusion constant D. Tourin *et al.*[11] showed that as long as $a \gg \sqrt{Dt}$ (far-field condition), the coherent backscattering spot width narrows as $a\lambda/\sqrt{Dt}$, as was also shown in optics.[5] Yet for small distances a or weakly scattering media (high D) or long times, the spatial extent of the coherent peak no longer displays such a time dependence and saturates at about $\lambda/2$, like in the near-field configuration. This saturation reduces the time interval during which the cone width evolution in $(Dt)^{-1/2}$ is valid: It prevents a quantitative measurement of D. We will show that this inconvenience can be overcome by plane-wave beamforming.

Beamforming is a classical technique employed, e.g., in medical imaging or underwater acoustics. Instead of firing just by one element, plane-wave beamforming consists in emitting a plane wave with a steering angle α with respect to the normal direction of the array. This angle is controlled by a set of delays that are applied to each array element, since ultrasonic array elements are controlled by independent electronic channels. In the reception mode, delays corresponding to an angle β are applied to the received signals before they are summed. Hence, the whole array can be used as a source and as a receptor. By adjusting the delays, it is possible to vary progressively the angles of emission α and reception β.

In this study, first we will show that plane-wave beamforming with ultrasonic arrays can be applied in the context of coherent backscattering. The advantages and drawbacks of beamforming compared to the classical approach[11] (direct intensity measurements, without beamforming) are emphasized. Particularly, a possible improvement in angular resolution is discussed. It would allow one to estimate D even though the far-field condition ($a \gg \sqrt{Dt}$) is not fulfilled. Moreover, in the presence of additive white noise, beamforming will be shown to increase the signal-to-noise ratio, due to residual spatial correlations between multiple scattered waves. In the last section of the paper, beamforming will be applied to the case of a weakly scattering medium, where the coherent backscattering enhancement is hardly visible, hence the enhancement factor is very low. In that case, we will show that beamforming can do better than direct intensity measurements, provided a correction is taken into account.

II. PRINCIPLE AND APPLICATIONS

A. Experimental setup and basic principle

The experiment takes place in a water tank. We use an N-element ultrasonic array with a 3 MHz central frequency and a 2.5–3.5 MHz bandwidth; each array element is 0.39 mm in size and the array pitch p is 0.417 mm. The sampling frequency is 20 MHz. The first step of the experiment consists in measuring the interelement matrix of the array (see Fig. 1). A 100-μs-long chirp is emitted from transducer i into the scattering sample. The backscattered wave is recorded with the N transducers of the same array. The operation is repeated for the N emitting transducers. The response from transducer i to transducer j is correlated with the emitted chirped signal, which gives the impulse response $h_{ij}(t)$. The $N \times N$ array response matrix $\mathbf{H}(t)$ whose elements are the N^2 impulse responses $h_{ij}(t)$ is finally obtained.

The usual way to detect the coherent backscattering phenomenon is to compute the average backscattered intensity directly from the time-dependent response matrix $\mathbf{H}(t)$.[15] The signals are time shifted to compensate for the difference of arrival times Δt_{ij} between receiving array elements, so that the first backscattered wave front arrives at $t=0$ on every element. The backscattered intensity is obtained by integrating the square of the signals on 10-μs-long overlapping time windows. The results are averaged for all transmitter/receiver couples that correspond to the same backscattering angle θ, in order to reduce the fluctuations of the intensity pattern. Finally, the averaged backscattered intensity is determined as

$$I^{\mathrm{DM}}(\theta,T) = \langle |h_{ij}(t-\Delta t_{ij})|^2 \rangle_{t-\Delta t_{ij} \in [T,T+\delta t],i}, \tag{1}$$

where $\delta t = 10\ \mu$s, $\theta = \arctan[[(i-j)p]/a]$ The symbol $\langle . \rangle$ denotes the average over the variables in the subscript. The superscript DM stands for "direct measurements." Figure 2(a) shows an experimental result for the averaged backscattered intensity $I^{\mathrm{DM}}(\theta,T)$ from a multiple scattering medium. Corresponding experimental conditions are described in Sec. II B.

Another way to process the data $\mathbf{H}(t)$ consists in plane-wave beamforming. Once the array response matrix is known, beamforming can be achieved in two ways. One way is to work directly in the time domain and to introduce a time delay between impulse responses $h_{ij}(t)$ at emission and reception, so that signals from the required direction are brought into phase and can be added together coherently. For a plane-wave emission with an angle α, the time delay Δt_i to apply to transducer i is

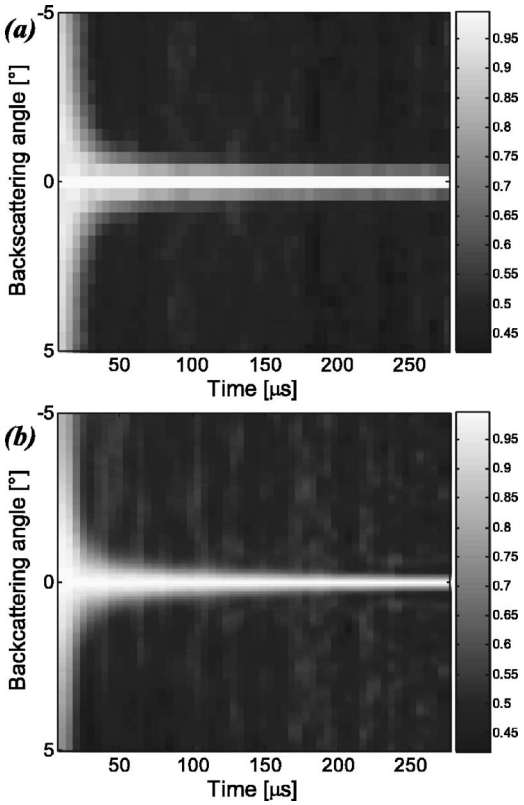


FIG. 2. Dynamic backscattered intensity obtained from direct measurements (a) and after beamforming (b). The scattering sample is made of 0.4-mm-radius steel rods with a concentration of 29.54 rods/cm^2. The distance a is 67 mm and the number of array elements is 284. At each time, the dynamic intensity has been normalized by its maximum.

$$\Delta t_i = \frac{(i-1)p \sin \alpha}{c}, i = 1,2, \ldots, N \tag{2}$$

if the first transducer is taken as reference ($\Delta t_1 = 0$). c is the wave velocity in water. Since small variations in the angle α may be required, the time delays Δt_i must be calculated with a precision lower than the sampling time ($0.05\ \mu$s). This requires an oversampling of the data well above Shannon's limit before processing, which is possible yet time-consuming. We chose to perform beamforming in the frequency domain instead. First, the time signals $h_{ij}(t)$ are truncated into 10-μs-long overlapping windows: $k_{ij}(T,t) = h_{ij}(t-T)w(t)$ with $w(t)=1$ for $t \in [0, 10\ \mu\text{s}], w(t)=0$ elsewhere. For each value of time T, the k_{ij} form a matrix $\mathbf{K}$. A short-time Fourier analysis is achieved by a fast Fourier transform and gives the response matrices $\mathbf{K}(T,f)$ in the frequency domain. For each frequency f in the bandwidth, the plane wave whose propagation direction makes an angle α with the axis perpendicular to the array corresponds to the spatial Fourier transform at the spatial frequency f_s given by

$$f_s = \frac{f \sin \alpha}{c}. \tag{3}$$

Next, the two-dimensional (2D) spatial Fourier transform of the response matrices gives the time-dependent plane wave decomposition $k_{\alpha\beta}(T,f)$ of $\mathbf{K}(T,f)$ where α and β correspond to angular directions at emission and reception, respectively. After that, the backscattered intensity $I^{\text{BF}}(\theta,T)$ (the superscript BF stands for beamforming) is obtained by averaging the squared norm of $k_{\alpha,\alpha+\theta}(T,f)$ over the emission angle α and the frequency bandwidth [see Eq. (5)]. As the evolution of the cone width in $(DT)^{-1/2}$ is valid only under the assumption of small incident angles, the average over the emission angle α is restricted to the interval $[-10°, 10°]$. Moreover, transducers' directivity $D(\alpha,f)$ is compensated for, using the theoretical expression under the far-field assumption:

$$\mathcal{D}(\alpha,f) = \frac{\sin(\pi b f \tan \alpha/c)}{\pi b f \tan \alpha/c}, \tag{4}$$

where b is the effective aperture of an array element. Finally, the backscattered intensity $I^{\text{BF}}(\theta,T)$ is inferred from

$$I^{\text{BF}}(\theta,T) = \left\langle \frac{|k_{\alpha,\alpha+\theta}(T,f)|^2}{\mathcal{D}(\alpha,f)\mathcal{D}(\alpha+\theta,f)} \right\rangle_{f,\alpha}, \tag{5}$$

where subscripts denote the averages over f and α. Figure 2(b) shows an experimental result for the averaged backscattered intensity obtained after beamforming from a multiple scattering medium. Corresponding experimental conditions are described in Sec. II B.

With or without beamforming, the averaged backscattered intensity $I(\theta,T)$ at a given time T, originates from single scattering (I_{SS}) and multiple scattering (I_{MS}) contributions: $I(\theta,T) = I_{\text{SS}}(\theta,T) + I_{\text{MS}}(\theta,T)$. The contribution due to multiple scattering can be split in two terms: an incoherent term (I_{inc}) that corresponds to the incoherent summation of multiple-scattering paths intensity, and a coherent term (I_{coh}) which results from the interference between multiple-scattering paths and their reciprocal counterparts.[2] At exact backscattering (i.e., $\theta=0$), these two terms are equal: $I_{\text{inc}}(0,T) = I_{\text{coh}}(0,T)$. Given the scattering angles considered in this study, the incoherent term as well as the single scattering contribution are almost flat: $I_{\text{SS}}(\theta,T) = I_{\text{SS}}(T)$ and $I_{\text{inc}}(\theta,T) = I_{\text{inc}}(T)$. In the case of plane waves, Akkermans *et al.*[5] calculated the time-dependent coherent intensity using plane wave decomposition in the long time limit $\sqrt{DT} \gg l_e$, where l_e is the mean free path, and for a small angle θ:

$$I^{\text{BF}}_{\text{coh}}(\theta,T) \simeq I^{\text{BF}}_{\text{inc}}(T) e^{-DTk^2\theta^2}. \tag{6}$$

A linear fit of the inverse square of the half width at half-maximum $\Delta\theta_{\text{BF}}$ gives access to the diffusion constant D, since

$$\Delta\theta^{-2}_{\text{BF}} = \frac{k^2 D}{\ln(2)} T. \tag{7}$$

When plane-wave beamforming is not applied, the expression of the coherent intensity as well as the influence of transducers' directivity are more complex. Under the far-field condition $a \gg \sqrt{DT}$, the inverse square of the half width at half maximum is then given by[17]

$$\Delta\theta_{\mathrm{DM}}^{-2} = \frac{k^2 D}{\Gamma} T, \tag{8}$$

where Γ is a factor which mainly depends on transducers' directivity and on the distance a. One advantage of plane-wave beamforming is that it allows one to compensate for transducers' directivity, with no need to compute the factor Γ.

B. Angular resolution

Another advantage of beamforming is a possible improvement in angular resolution. To illustrate this idea, we measure the response matrix from a highly scattering sample. It consists of steel rods (C_L=5.9 mm/μs, C_T=3.2 mm/μs, radius 0.4 mm, $\rho=7.85\times10^3$ kg/m^3) randomly distributed with a concentration n=29.54 rods/cm^2. The elastic mean free path l_e is typically 3.15±0.15 mm for this medium between 2.5 and 3.5 MHz.[18] The distance a is 67 mm. The measured response matrix $\mathbf{H}(t)$ is of dimension N=284. The averaged backscattered intensities, with or without preliminary beamforming, are then calculated as described in Sec. II A and are shown in Fig. 2. Because this sample is a strongly multiple scattering medium, the single scattering contribution $I_{\mathrm{SS}}(\theta,T)$ becomes rapidly negligible with time compared to the multiple scattering contribution $I_{\mathrm{MS}}(\theta,T)$ and the enhancement factor reaches 2 after about 20 μs. A comparison between experimental results with or without beamforming (Fig. 2) shows that the angular resolution is much finer when beamforming is performed. Indeed, the angular resolution $\delta\theta_r^{\mathrm{DM}}$ of a direct intensity measurement is given by

$$\delta\theta_r^{\mathrm{DM}} \simeq \frac{b}{a}. \tag{9}$$

The effective aperture b of transducers equals 0.55 mm and is different from the array pitch p=0.417 mm because of the mechanical coupling between transducers. The effective aperture b is calculated by measuring the far-field radiation diagram of a single element. A fit of the experimental directivity with its theoretical expression given in Eq. (4) provides an estimation for b. A finer resolution could be obtained by moving the array away from the scattering sample (thus increasing a) but the signal-to-noise ratio would decrease. Beamforming improves the angular resolution while working at a moderate distance a, which allows an acceptable signal-to-noise ratio. Indeed, the angular resolution $\delta\theta_r^{\mathrm{BF}}$ in the far-field is given by

$$\delta\theta_r^{\mathrm{BF}} \simeq \frac{\lambda}{Np}. \tag{10}$$

Increasing the number of array elements, i.e., widening the array, refines the angular resolution and allows one to go below b/a. In our experimental conditions, the angular resolution $\delta\theta_r^{\mathrm{DM}}$ of direct measurements is about 0.47° [Eq. (9)] whereas the angular resolution with beamforming $\delta\theta_r^{\mathrm{BF}}$ is 0.24° [Eq. (10)].

In order to compare both experimental results with the same precision, the angular intensity distributions are resampled with an angular step of 45 μrad using a low-pass interpolation. The result of this resampling is shown in Fig. 3, which compares the angular distribution of intensity with and without beamforming, at T=265 μs. The gain in resolution thanks to beamforming is obvious in Fig. 3: The peak is significantly narrower with beamforming. In fact, in this experiment, the width $\Delta\theta_{\mathrm{DM}}$ of the coherent backscattering peak obtained without beamforming reaches its saturation limit at T=75 μs. The evolution of $\Delta\theta^{-2}$ with time is plotted in Fig. 4. For direct measurements (i.e., no beamforming), the cone narrows with time; a linear evolution of $\Delta\theta_{\mathrm{DM}}^{-2}$ versus time is found, once the diffusive regime is reached (after ~15 μs) and before the saturation due to the angular resolution limit occurs (after ~75 μs). According to Eq. (8), the slope of the linear fit gives an estimation of the diffusion constant once the constant Γ is known. The angular resolution limit restricts the time interval where a linear fit of $\Delta\theta_{\mathrm{DM}}^{-2}$ can be achieved and thus the estimation of the diffusion constant is less reliable. It also prevents one from observing the evolution of the diffusive regime for long times. The situa-

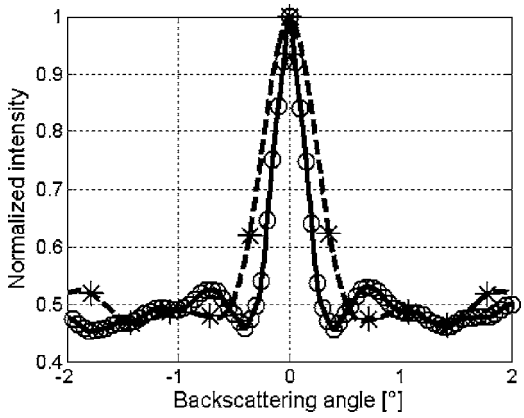


FIG. 3. Normalized intensity vs backscattering angle at T=265 μs, with (circles) or without (stars) beamforming. In order to measure the half width at half maximum, original data have been resampled at the same sample rate (angular step of 45 μrad). Resampled data with (continuous line) or without (dashed line) beamforming are superposed to original measurements.

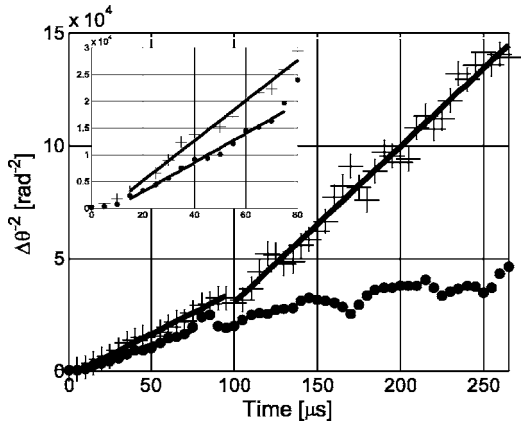


FIG. 4. Evolution of the inverse square of the half width at half maximum $\Delta\theta^{-2}$ with time. Dots: Direct measurements, i.e., no beamforming; crosses: With beamforming. For direct measurements, a linear fit has been made from T=15 μs to T=75 μs. In the case of beamforming, linear fits are represented from T=15 μs to T=95 μs and from T=100 μs to T=265 μs. A zoom on the first 80 μs is displayed on the top left corner.

tion is very different if beamforming is achieved (see Fig. 4); no saturation is observed. Interestingly, a change of slope is observed around $T=100$ μs that could not have been perceived without beamforming because saturation occurs around $T=75$ μs. This change of slope can be accounted for. For short times, the signal spectrum is wideband; the −3 dB bandwidth, $\delta f_{-3\ \mathrm{dB}}$, is larger than 1.2 MHz and as a consequence, the measured diffusion constant is averaged over the whole frequency bandwidth (2.5–3.5 MHz). Between T $=100$ μs and T$=265$ μs, the signal spectrum becomes narrowband ($\delta f_{-3\ \mathrm{dB}} \simeq 0.2$ MHz) and centered around 2.8 MHz, because of the individual resonance of rods at 2.8 MHz. Mamou[17] showed that the diffusion constant and the mean free path are higher at this frequency: for instance, the mean free path l_e is 4.8 mm at the resonance frequency,[18] i.e., about 1.5 times higher than the mean free path averaged over the frequency bandwidth. This is consistent with our measurements because a linear fit from $T=15$ μs to $T=95$ μs of $\Delta\theta_{\mathrm{BF}}^2$ gives a diffusion constant $D=1.9$ mm$^2/\mu$s, whereas from $T=100$ μs to $T=265$ μs, $D=3.4$ mm$^2/\mu$s.

It should also be noted that there are many techniques other than dynamic coherent backscattering to measure diffusive parameters; particularly, the diffusion constant can be estimated from measurements of the time-of-flight distribution,[19–23] or the spatial extent of the diffuse halo.[19] Yet the dynamic coherent backscattering enhancement, aside from being a very elegant physical illustration of reciprocity, is independent from intrinsic absorption. Moreover, the size of the arrays is usually not very large compared to $\sqrt{DT}$, so the transverse size of the diffuse halo cannot always be directly measured from the data available on the array.

C. Robustness to noise

Another advantage is that intensity measurements with beamforming are less noise-sensitive than direct measurements. In underwater acoustics or telecommunication, the *array gain* is the improvement in signal-to-noise ratio (SNR) provided by coherently combining the beamformer's N antenna signal. The addition of N identical signals increases the signal amplitude by N, whereas the total amplitude of N uncorrelated noises with identical power only increases as $\sqrt{N}$. Therefore, in the presence of additive uncorrelated noise, if the array elements were to receive perfectly identical signals, then an array gain of $\sqrt{N}$, relative to the SNR of a reference single-element system, could be expected. On the contrary, if the signals received by the array have no spatial correlation, then there is no array gain: The SNR is the same whatever the number of elements on the array. Here, for each scattering angle, the beamforming technique utilizes the whole array to compute the backscattered intensity. The question we address in this paragraph is: is there an array gain? In other words, in a coherent backscattering experiment, is the SNR improved by beamforming?

In the experiment we consider here, the wave forms received on the array are backscattered from a highly scattering random medium. In that case, the signals received on two array elements are far from being identical. Yet the correlation length of the backscattered waves is larger than the array pitch. There are two reasons for that. First, there is a mechanical coupling between neighboring array elements. Second, the wave recorded on the array can be seen as the radiation of an incoherent source with width W (the diffuse halo inside the medium) observed at a distance a. The Van Cittert-Zernike theorem[24,25] states that the typical coherence length of the wave field is $\lambda a/W$ (in other words, the waves radiated by a finite-size incoherent source see their coherence length increases as they propagate).

These two effects result in a short-range correlation between the scattered signals recorded on the array. In the experimental situation we investigated, these residual correlations are limited in range to adjacent elements, both in emission and reception. Experimentally, it has been observed that the two impulse responses $h_{i,j}$ and $h_{m,n}$ are correlated if $|i-m| \leq 1$ and $|j-n| \leq 1$. So, only the following correlation coefficients are not zero:

$$C_1 = \frac{\langle h_{i,j} h^*_{i\pm1,j} \rangle}{\langle |h_{i,j}|^2 \rangle} = \frac{\langle h_{j,i} h^*_{i\pm1,j} \rangle}{\langle |h_{i,j}|^2 \rangle},$$

$$C_2 = \frac{\langle h_{i,j} h^*_{i,j\pm1} \rangle}{\langle |h_{i,j}|^2 \rangle} = \frac{\langle h_{j,i} h^*_{i,j\pm1} \rangle}{\langle |h_{i,j}|^2 \rangle},$$

$$C_3 = \frac{\langle h_{i,j} h^*_{i\pm1,j\pm1} \rangle}{\langle |h_{i,j}|^2 \rangle} = \frac{\langle h_{j,i} h^*_{i\pm1,j\pm1} \rangle}{\langle |h_{i,j}|^2 \rangle},$$

$$C_4 = \frac{\langle h_{i,j} h^*_{i\pm1,j\mp1} \rangle}{\langle |h_{i,j}|^2 \rangle} = \frac{\langle h_{j,i} h^*_{i\pm1,j\mp1} \rangle}{\langle |h_{i,j}|^2 \rangle}.$$

The average, denoted by $\langle . \rangle$, is achieved over time and source/receiver pairs. In the experiment described in Sec. II B, the typical values we obtained are: $C_1 = C_2 \simeq 0.425$, $C_3 \simeq 0.15$, and $C_4 \simeq 0.26$. Though weak, these correlations imply an array gain in the presence of an additive white noise. This effect can be calculated more precisely, the details are given in the Appendix. The theoretical array gain G due to these correlations is found to be

$$G \simeq 1 + 4\sum_{i=1}^{4} C_i. \tag{11}$$

From the above-given C_i, we obtain an array gain of 6.0. Here, the correlations are mainly due to the mechanical coupling between neighbor transducers. The correlations linked to the Van Cittert-Zernike theorem cannot be perceived for long times because the diffusive halo is too large and the corresponding coherence length $\lambda a/W$ is smaller than the array pitch p. Experimentally, this improvement in robustness to noise has been checked by adding a numerical white noise on initial measurements before beamforming; an array gain G of 5.4($\simeq$7.3 dB) has been found. This experimental value is in reasonable agreement with the theoretical one, the remaining mismatch probably comes from the small angles' assumption used in our theoretical study (see the Appendix).

D. Weakly scattering media

In Secs. II A–II C, only the case of a strongly multiple scattering medium has been discussed, and the advantages of beamforming have been revealed. In this section, beamforming is applied to the case of a weakly scattering medium. It is interesting to point out an error that can occur when beamforming is applied to a situation in which single scattering dominates. However, once a correction is applied to $I_{SS}^{BF}(\theta,T)$, another advantage of beamforming will be shown. It should be noted that there is no such thing as a purely multiple scattering medium or a purely single scattering medium. In any inhomogeneous medium, there is a combination of both. Yet, according to the time and path lengths involved compared to the mean free time or mean free path, single scattering can sometimes predominate or, on the contrary, be neglected. Here we consider the application of beamforming to a weakly scattering medium, where single scattering dominates. Yet multiple scattering is not ruled out: a coherent backscattering peak can emerge, but the enhancement factor is so weak that it is difficult to distinguish.

In this experiment, the weakly scattering medium is a slab of agar gel whose thickness L is about 100 mm. The distance a is 50 mm and 128 elements have been used to measure the response matrix $\mathbf{H}(t)$. Because the medium is only weakly scattering ($l_e > 50$ mm), average over several disorder configurations is needed. To achieve such an average, 19 response matrices have been measured by moving the array with a motor (see Fig. 1). The averaged backscattered intensity obtained after beamforming is shown in Fig. 5(a). Since the medium scatters weakly, there should be almost no coherent backscattering enhancement and the intensity profile should be flat. The experimental results contradict this prediction [Fig. 5(a)]: A peak seems to emerge at late times. One may wrongly conclude that there is more multiple scattering than expected. Actually, this artifact is due to the finite dimension of the array. Assume an incident quasiplane wave which makes an angle α with the axis perpendicular to the array (see Fig. 6). The backscattered signals which are received by the array elements at time T come from reflections on scatterers located at a depth $d=(a+cT/2)$. As the array has a finite aperture, the transverse size of the area $\mathbf{A}$ insonified by the incident plane-wave is approximately $N \times p$, corresponding to the array width. Only scatterers contained in this area $\mathbf{A}$ scatter the incident plane-wave. Now, if the backscattered plane-wave with an angle $\alpha+\theta$ is considered, only a fraction of the backscattered beam can be intercepted by the array, as it is depicted in Fig. 6. Note h the tranverse size of the beam backscattered out of the array. For small incident and backscattering angles, at time T, h can be expressed as

$$h \simeq d|\theta| \simeq \left(a + \frac{cT}{2}\right)|\theta|. \tag{12}$$

Therefore, a fraction h/Np of the backscattered intensity is not recorded by the array. This fraction increases with time and with the backscattering angle θ: Even if single scattering dominates, it gives the illusion that there is a coherent backscattering peak [Fig. 5(a)]. So, when single scattering dominates, the measured intensity must be corrected into

$$I_{SS}^{BF}(\theta,T)\left[1 - \frac{(a+cT/2)|\theta|}{Np}\right]^{-1}. \tag{13}$$

The corrected single scattering intensity calculated from the measured intensity of Fig. 5(a) is shown in Fig. 5(b). The intensity profile is now flat except for a weak enhancement at exact backscattering that can now be perceived for long times. Fluctuations of the intensity pattern are observed in

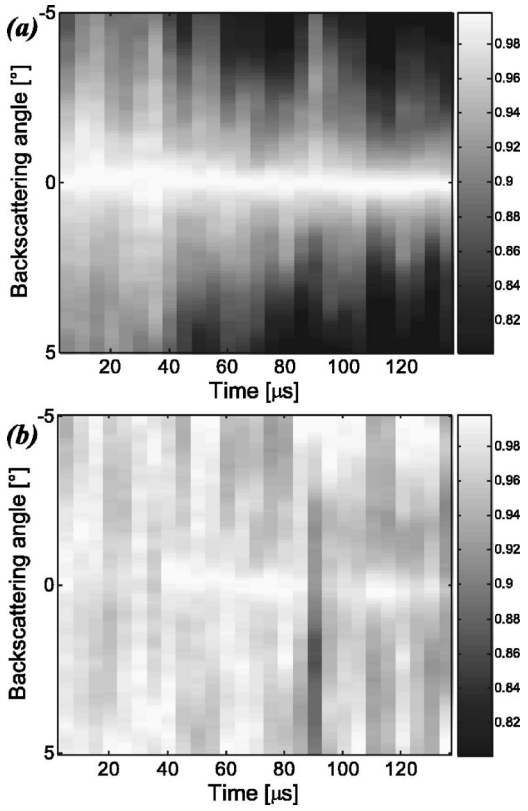


FIG. 5. Dynamic backscattered intensity obtained for a weakly scattering sample made of agar gel. The distance a is 50 mm. At each time, the dynamic intensity has been normalized by its maximum. The dynamic intensity obtained after beamforming is shown before (a) and after (b) applying the correction of Eq. (13)

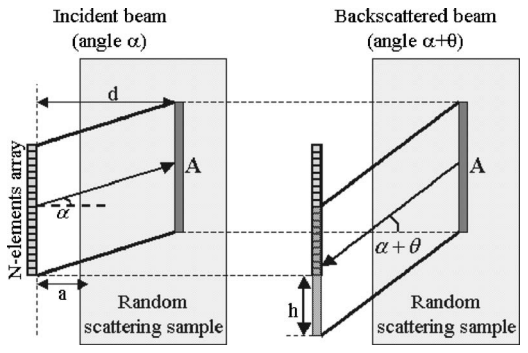


FIG. 6. Effect of the array finite aperture on the single scattering intensity obtained after beamforming. On the left is depicted the area $\mathbf{A}$ sounded by an incident beam corresponding to a plane-wave with an incident angle α, at a distance $d(=a+cT/2)$ from the array. On the right, the backscattered beam at an angle $\alpha+\theta$ comes from $\mathbf{A}$ and intercepts then only a fraction of the array. h is the size of the beam scattered out of the array.

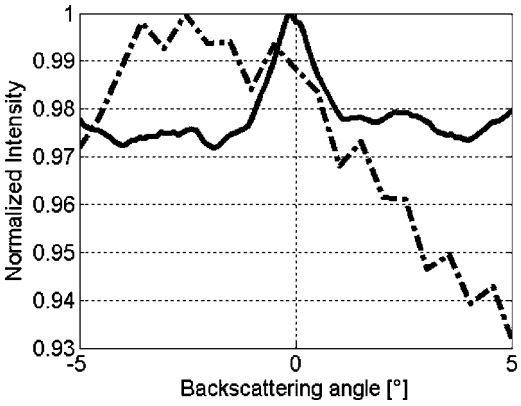


FIG. 7. Dynamic backscattered intensity obtained with the agar gel, integrated from $T=35$ μs to $T=135$ μs before beamforming (dash-dotted line) and after beamforming and correction (continuous line).

Fig. 5(b) and are due to a lack of average on disorder configurations. The intensity has been integrated from $T=35$ μs to $T=135$ μs to reduce these variations. The integrated intensity is shown in Fig. 7 before and after beamforming. The intensity obtained from direct measurements does not display a flat baseline, which cannot provide a quantitative measurement of a low enhancement factor. On the contrary, the intensity obtained after beamforming exhibits a very regular plateau which enables one to estimate a very weak enhancement factor of about 1.025. The fluctuations in the single scattering intensity profile obtained from direct measurements originates from the variation in sensitivity of transducers. These fluctuations in sensitivity cannot be considered as random and display long-range correlations: it appears that one side of the array is more sensitive than the other one and as a consequence, intensity for negative backscattering angles is a bit higher than for positive angles (see Fig. 7). With beamforming, fluctuations of transducers' sensitivity are averaged out and a flat single scattering intensity profile is obtained.

III. CONCLUSION

In this study, advantages and drawbacks of far-field beamforming for ultrasonic measurements of the coherent backscattering effect have been investigated. The main disadvantage is the computation time that is quite long compared to a direct intensity measurement from the array response matrix. Nevertheless, beamforming can be of a great interest to measure more properly dynamic transport properties of a multiply scattered acoustic wave from the coherent backscattering effect. Indeed, beamforming can provide an improvement in angular resolution which is very useful for a quantitative measurement of the diffusion constant. Furthermore, beamforming processing is more robust to additional white noise on measurements because it takes advantage of correlations between neighboring transducers. Concerning weakly scattering media, a correction factor has to be applied to the single scattering intensity because of the array finite aperture. Once this correction has been made, the measurement of a low enhancement factor can be obtained.

ACKNOWLEDGMENTS

The authors wish to acknowledge the groupe de recherches IMCODE of CNRS (GDR 2253), Patricia Daenens for her technical help, and Victor Mamou who made the steel rod samples.

APPENDIX

Assuming an additive white noise on each measurement, the measured impulse response is written $h_{i_1,i_2}(t)+n_{i_1,i_2}(t)$, where h_{i_1,i_2} is the non-noisy impulse response and $n_{i_1,i_2}(t)$ an additive white noise. The averaged backscattered intensity I_n^{DM} obtained without beamforming is

$$I_n^{\mathrm{DM}}(\theta,T) = I^{\mathrm{DM}}(\theta,T) + \eta^2 = I_{\mathrm{inc}}^{\mathrm{DM}}(T) + I_{\mathrm{coh}}^{\mathrm{DM}}(\theta,T) + \eta^2, \tag{A1}$$

where $I^{\mathrm{DM}}(\theta,T)=\langle|h_{i_1,i_2}(t)|^2\rangle_{i_1,t\in[T,T+\delta t]}$ is the averaged backscattered intensity without noise. θ corresponds to the angle associated with the elements couple (i_1,i_2) and η is the standard deviation of the additive noise. The error ϵ^{DM} on the backscatterred intensity without beamforming is

$$\epsilon^{\mathrm{DM}}(\theta,T) = \frac{I_m^{\mathrm{DM}}(\theta,T) - I^{\mathrm{DM}}(\theta,t)}{I^{\mathrm{DM}}(\theta,t)} = \frac{\eta^2}{I_{\mathrm{inc}}^{\mathrm{DM}}(T) + I_{\mathrm{coh}}^{\mathrm{DM}}(\theta,T)}. \tag{A2}$$

When beamforming is applied, the original signals are first truncated in time, then a temporal Fourier transform is achieved on each time window. Finally, we have to deal with the 2D spatial Fourier transform of random signals. Without noise, the backscattered intensity is

$$I^{\mathrm{BF}}(\theta,T) = I_{\mathrm{coh}}^{\mathrm{BF}}(\theta,T) + I_{\mathrm{inc}}^{\mathrm{BF}}(T). \tag{A3}$$

If the transducers directivity is not taken into account in Eq. (5), the backscattered intensity directly equals $\langle|k_{\alpha,\alpha+\theta}(T,f)|^2\rangle_{f,\alpha}$. To achieve beamforming, the 2D spatial Fourier transform is done numerically using the discrete Fourier transform algorithm. If the assumption of small incident and backscattering angles is made, the backscattered intensity $I^{\mathrm{BF}}(\theta,T)$, obtained after beamforming and without noise, can be expressed as

$$I^{\mathrm{BF}}(\theta,T) \simeq \frac{1}{N^2}\sum_{i_1=1}^{N}\sum_{l_1=1}^{N}\sum_{i_2=1}^{N}\sum_{l_2=1}^{N}\langle k_{i_1,i_2}(T,f)k_{l_1,l_2}^{*}(T,f) \times e^{-j2\pi p(f/c)[\alpha(i_1-l_1)+(\alpha+\theta)(i_2-l_2)]}\rangle_{f,\alpha}. \tag{A4}$$

If the responses $k_{i_1,i_2}(T,f)$ were totally decorrelated one from another, only the incoherent terms $\langle|k_{i_1,i_2}^{T}(T,f)|^2\rangle$ and coherent terms $\langle|k_{i_1,i_2}(T,f)|^2 e^{j2\pi p(f/c)\theta(i_1-i_2)}\rangle$ would survive the average. But, as backscattered signals are correlated, correlation coefficients C_1,C_2,C_3, and C_4 have to be taken into account. Still under the assumption of small backscattering angle and $N\gg 1$, the coherent and incoherent intensity can be expressed as

$$I_{\mathrm{inc}}^{\mathrm{BF}}(T) \simeq \left(1+4\sum_{i=1}^{4}C_i\right)\langle|k_{i_1,i_2}(T,f)|^2\rangle_{i_1,i_2,f}, \tag{A5}$$

$$I_{\mathrm{coh}}^{\mathrm{BF}}(\theta,T) \simeq \left(1+4\sum_{i=1}^{4} C_i\right) \times \langle |k_{i_1,i_2}(T,f)|^2 e^{j2\pi p(f/c)\theta(i_1-i_2)} \rangle_{i_1,i_2,f}. \quad \text{(A6)}$$

As the noise is totally decorrelated on each element of the array, its contribution to the mean intensity is η^2. Finally, the error ϵ^{BF} on the measured backscattered intensity with beamforming is

$$\epsilon^{\mathrm{BF}}(\theta,T) = \frac{\eta^2}{I_{\mathrm{inc}}^{\mathrm{BF}}(T) + I_{\mathrm{coh}}^{\mathrm{BF}}(\theta,t)}. \quad \text{(A7)}$$

As the coherent intensity $I_{\mathrm{coh}}^{\mathrm{DM}}$ obtained without beamforming is not zero only close to the source, the coherent intensity contribution is negligible in $\langle |k_{i_1,i_2}(T,f)|^2 \rangle_{i_1,i_2,f} = \langle |h_{i_1,i_2}(t)|^2 \rangle_{i_1,i_2,t\in[T,T+\delta t]}$. It comes out that $I_{\mathrm{inc}}^{\mathrm{DM}}(T) \simeq \langle |k_{i_1,i_2}(T,f)|^2 \rangle_{i_1,i_2,f}$, and Eq. (A5) simplifies in

$$I_{\mathrm{inc}}^{\mathrm{BF}}(T) \simeq \left(1+4\sum_{i=1}^{4} C_i\right) I_{\mathrm{inc}}^{\mathrm{DM}}(T). \quad \text{(A8)}$$

Let θ_{lim} be the angle from which $I_{\mathrm{coh}}^{\mathrm{BF}}$ and $I_{\mathrm{coh}}^{\mathrm{DM}}$ vanish. For $\theta > \theta_{\mathrm{lim}}$, the array gain G which corresponds to the ratio between ϵ^{DM} and ϵ^{BF} is found to be

$$G = \frac{\epsilon^{\mathrm{DM}}}{\epsilon^{\mathrm{BF}}} \simeq 1+4\sum_{i=1}^{4} C_i. \quad \text{(A9)}$$

Note that Eq. (A9) is also valid at $\theta=0$ because the coherent intensity is strictly equal to the incoherent one with or without beamforming in that case. Between $\theta=0$ and $\theta=\theta_{\mathrm{lim}}$, the exact computation of the array gain G is tedious.

[1] Y. Kuga and A. Ishimaru, "Retroreflectance from a dense distribution of spherical particles," J. Opt. Soc. Am. A **1**, 831–835 (1984).
[2] P.-E. Wolf and G. Maret, "Weak localization and coherent backscattering of photons in disordered media," Phys. Rev. Lett. **55**, 2696–2699 (1985).
[3] M. van Albada and A. Lagendijk, "Observation of weak localization of light in a random medium," Phys. Rev. Lett. **55**, 2692–2695 (1985).
[4] E. Akkermans, P.-E. Wolf, and R. Maynard, "Coherent backscattering of light by disordered media: Analysis of the peak line shape," Phys. Rev. Lett. **56**, 1471–1474 (1986).
[5] E. Akkermans, P.-E. Wolf, R. Maynard, and G. Maret, "Theoretical study of the coherent backscattering of light by disordered media," J. Phys. (France) **49**, 77–98 (1988).
[6] D. Wiersma, M. van Albada, B. van Tiggelen, and A. Lagendijk, "Experimental evidence for recurrent multiple scattering events of light in disordered media," Phys. Rev. Lett. **74**, 4193–4196 (1995).
[7] G. Labeyrie, F. de Tomasi, J.-C. Bernard, C. Müller, and R. Kaiser, "Coherent backscattering of light by cold atoms," Phys. Rev. Lett. **83**, 5266–5269 (1999).
[8] G. Labeyrie, C. Müller, D. Wiersma, C. Miniatura, and R. Kaiser, "Observation of coherent backscattering of light by cold atoms," J. Opt. B: Quantum Semiclassical Opt. **2**, 672–685 (2000).
[9] D. Kupriyanov, I. Solokov, C. Sukenik, and M. Havey, "Coherent backscattering of light from ultracold and optically dense atomic ensembles," Laser Phys. Lett. **3**, 223–243 (2006).
[10] G. Bayer and T. Niederdränk, "Weak localization of acoustic waves in stongly scattering media," Phys. Rev. Lett. **70**, 3884–3887 (1993).
[11] A. Tourin, A. Derode, P. Roux, B. van Tiggelen, and M. Fink, "Time-dependent backscattering of acoustic waves," Phys. Rev. Lett. **79**, 3637–3639 (1997).
[12] E. Larose, L. Margerin, B. van Tiggelen, and M. Campillo, "Weak localization of seismic waves," Phys. Rev. Lett. **93**, 048501 (2004).
[13] B. van Tiggelen, L. Margerin, and M. Campillo, "Coherent backscattering of elastic waves: Specific role of source, polarization, and near field," J. Acoust. Soc. Am. **110**, 1291–1298 (2001).
[14] L. Margerin, M. Campillo, and B. van Tiggelen, "Coherent backscattering of acoustic waves in the near field," Geophys. J. Int. **145**, 593–603 (2001).
[15] A. Derode, V. Mamou, F. Padilla, F. Jenson, and P. Laugier, "Dynamic coherent backscattering in a heterogeneous medium: Application to human trabecular bone characterization," Appl. Phys. Lett. **87**, 114101 (2005).
[16] K. Sakai, K. Yamamoto, and K. Takagi, "Observation of acoustic coherent backscattering," Phys. Rev. B **56**, 10930–10933 (1997).
[17] V. Mamou, "Caractérisation ultrasonore d'échantillons hétérogènes multiplement diffuseurs," Ph.D. thesis, Université Paris 7 - Denis Diderot, 2005.
[18] A. Derode, V. Mamou, and A. Tourin, "Influence of correlations between scatterers on the attenuation of the coherent wave in a random medium," Phys. Rev. E **74**, 036606 (2006).
[19] J. Page, H. Shriemer, A. Bailey, and D. Weitz, "Experimental test of the diffusion approximation for multiply scattered sound," Phys. Rev. E **52**, 3106–3114 (1995).
[20] S. Skipetrov and B. van Tiggelen, "Dynamics of weakly localized waves," Phys. Rev. Lett. **92**, 113901 (2004).
[21] M. Storzer, P. Gross, C. Aegerter, and G. Maret, "Observation of the critical regime near Anderson localization of light," Phys. Rev. Lett. **96**, 063904 (2006).
[22] L. Margerin, M. Campillo, and B. van Tiggelen, "Radiative transfer and diffusion of waves in a layered medium: New insight into coda Q," Geophys. J. Int. **134**, 596–612 (1998).
[23] M. Haney, K. van Wijk, and R. Snieder, "Radiative transfer in layered media and its connection to the O'Doherty-Anstey formula," Geophysics **70**, 1–11 (2005).
[24] A. Derode and M. Fink, "Partial coherence of transient ultrasonic fields in anisotropic random media: Application to coherent echo detection," J. Acoust. Soc. Am. **101**, 690–704 (1997).
[25] J. Goodman, *Statistical Optics* (Wiley, New York, 1985), Chap. 5.

Excitation of soft porous frame resonances and evaluation of rigidity coefficients

Jean F. Allard and Bruno Brouard
Laboratoire d'Acoustique de l'Université du Maine, UMR CNRS 6613, Avenue Olivier Messiaen, 72085 Le Mans Cedex 9, France

Noureddine Atalla and Sebastian Ghinet
Mechanical Engineering, Université de Sherbrooke, Sherbrooke, QC J1K2R1, Canada

(Received 7 June 2006; revised 21 September 2006; accepted 11 October 2006)

Recent works have been performed concerning the guided elastic waves in soft porous elastic frames and the related normal displacement of the surface generated by a mechanical excitation. In this work, the mechanical excitation is replaced by a monopole acoustic field in air above the porous structure. Two cases are considered: a frame bonded on a rigid impervious layer and a free frame. The normal and the radial velocity induced by the monopole field at the surface of the frame is predicted with the Biot theory and the Sommerfeld representation of the monopole field. Measurements are performed with a laser vibrometer on a urethane foam. It is shown that the characterization of the excited modes can lead to an evaluation of the rigidity coefficients of the frames at acoustical frequencies. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2387127]

PACS number(s): 43.20.Gp, 43.20.Jr, 43.20.Ye [LPF] Pages: 78–84

I. INTRODUCTION

The rigidity parameters of porous frames are important parameters for the prediction of the acoustical properties of materials including porous and elastic layers. Many methods of measuring these parameters have been proposed.[1–9] Most of them are related to a quasistatic characterization of the rigidity parameters or of a characterization of the bulk waves from simple deformation fields. Other methods are based on the properties of the Rayleigh wave, and of different vibration modes of the porous structures.[10–12] In all these methods, a mechanical excitation of the frame is performed. The density of the porous frames is generally much larger than the density of air. As a consequence, in the context of the Biot theory,[13] for air saturated media, the shear Biot wave, and one of the two Biot compressional waves, are similar to the shear wave, and to the compressional wave of the frame in vacuum, respectively.[14] The vibration modes of a porous structure in air are not very different from the modes in vacuum, and the porous structure can be replaced in a first approximation by an elastic homogeneous medium. In the previous works by Boeckx *et al.*,[11,12] an analysis of the vertical displacement of the frame measured as a function of the distance from the excitation was performed to separate the contributions of the different modes.[11,12] In this work, the mechanical excitation is replaced by a monopole field created by a point source in air above the structure, and the analysis is more simpler. Predictions of porous structure displacement induced by a point source have been previously performed by Attenborough and Richard.[15] The predictions are performed like in this previous work with the Biot theory and the Sommerfeld representation of the monopole field. Two configurations are considered, the porous layer free at both faces, and bonded to a rigid impervious backing. The source-receiver geometry is represented in Fig. 1. The radial distance from the source S to M where the velocity of the frame is evaluated is r, and h is the height of the source. The vertical and radial displacements of the porous surface are predicted to set in evidence characteristic variations as a function of frequency. The simplest cases are symbolically represented in Fig. 2. The experimental identification of these variations will allow an evaluation of the rigidity coefficients. It may be noticed that an air mode, related to the Biot compressional wave that mainly propagates in air, has been previously studied for thin porous layers.[16] For media with a sufficiently small loss angle, the air mode can be modified by the shear resonance,[17] but the effect is generally difficult to set in evidence.

II. FRAME DISPLACEMENT INDUCED BY A SOUND SOURCE IN AIR

A. Frame displacement induced by a plane wave in air

The Biot theory is used to predict the horizontal and the vertical displacements u_x and u_z of the frame. Let k_1 and k_2 be the wave numbers of the two Biot compressional waves and k_3 is the wave number of the shear wave. Each plane wave is related to a displacement $\mathbf{u}^s$ of the frame and $\mathbf{u}^f$ of air, with $\mathbf{u}^f=\mu_j\mathbf{u}^s$, j=1, 3. Let P, Q, and R be the Biot rigidity coefficients. The quantities k_1, k_2, k_3, μ_1, μ_2, μ_3,P, Q, and R, are defined in Refs. 13, 14, and 18. An incident plane wave in air above the porous layer of Fig. 1 with a horizontal wave number component $\xi=k\sin\theta$, θ being the angle of incidence and k being the wave number in air, is related in the layers to Biot waves propagating from the upper to the lower face and from the lower to the upper face with the same horizontal wave number component. Let $\varphi_j^{\pm}$, $j=1,2$ the velocity potentials of the compressional waves, with the related displacement $\mathbf{u}^s=i\nabla\boldsymbol{\varphi}/\boldsymbol{\omega}$, and $\boldsymbol{\Psi}^{\pm}=\mathbf{n}\varphi_3^{\pm}$ the velocity potential of the shear waves, $\mathbf{n}$ being the unit vector on the y

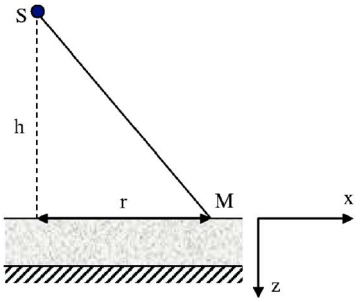


FIG. 1. Symbolic sketch of the monopole source S at a height h and a point at M at the surface at the radial distance r.

axis with the related frame displacement $\mathbf{u}_s = i\nabla \wedge \Psi/\omega$. The scalar functions $\varphi_i^{\pm}$ can be written, the time dependence $\exp(-i\omega t)$ being discarded

$$\varphi_j^{\pm} = a_j^{\pm}\exp(\pm i\alpha_j z + i\xi x), \tag{1}$$

$$\alpha_j = (k_j^2 - \xi^2)^{1/2}. \tag{2}$$

The velocity components for the frame are given by

$$\dot{u}_z^s = i\alpha_1(\varphi_1^+ - \varphi_1^-) + i\alpha_2(\varphi_2^+ - \varphi_2^-) + i\xi(\varphi_3^+ + \varphi_3^-), \tag{3}$$

$$\dot{u}_x^s = i\xi(\varphi_1^+ + \varphi_1^- + \varphi_2^+ + \varphi_2^-) - i\alpha_3(\varphi_3^+ - \varphi_3^-), \tag{4}$$

The velocity components for the air inside the porous medium are obtained by replacing the $\varphi_j^{\pm}$ by $\mu_j\varphi_j^{\pm}$ in Eqs. (3) and (4). Let $\sigma_{\alpha\beta}^f$ and $\sigma_{\alpha\beta}^s$ be the components of the Biot stress tensors, for air and for the frame, respectively. They are related to forces acting on the frame, and on air, for a unit surface of porous medium. The stress components in the Biot theory are given by

$$\sigma_{ij}^s = [(P - 2N)\theta^s + Q\theta^f]\delta_{ij} + 2Ne_{ij}^s, \tag{5}$$

$$\sigma_{ij}^f = (Q\theta^s + R\theta^f)\delta_{ij}, \tag{6}$$

where θ^f and θ^s are the dilatation of the air, and of the frame, respectively, and the coefficients e_{ij}^s are the strain components of the frame. The stress components are given by

$$\sigma_{zz}^s = B_1(\varphi_1^+ + \varphi_1^-) + B_2(\varphi_2^+ + \varphi_2^-) + C_3(\varphi_3^+ - \varphi_3^-), \tag{7}$$

$$\sigma_{xz}^s = C_1(\varphi_1^+ - \varphi_1^-) + C_2(\varphi_2^+ - \varphi_2^-) - H(\varphi_3^+ + \varphi_3^+), \tag{8}$$

$$\sigma_{zz}^f = A_1(\varphi_1^+ + \varphi_1^-) + A_2(\varphi_2^+ + \varphi_2^-), \tag{9}$$

$$A_j = (-i/\omega)(Q + R\mu_j)k_j^2, \tag{10}$$

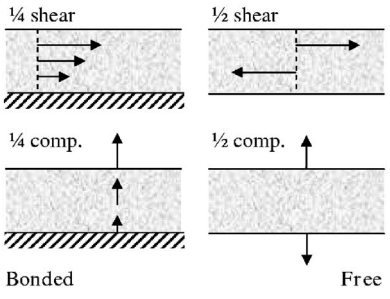


FIG. 2. Resonances for a bonded and a free elastic layer.

$$B_j = -(i/\omega)[(P + Q\mu_j)k_j^2 - 2N\xi^2], \tag{11}$$

$$C_j = -(i/\omega)2N\alpha_j\xi, \tag{12}$$

$$H = (iN/\omega)(\alpha^3 - \xi^2). \tag{13}$$

The coefficient N is the shear modulus of the frame. The six unknown $a_j^{\pm}$ can be obtained from a set of six equations. The 6×6 matrices are avoided in the following way. In a first step, the nine coefficients $r_{j,1}=a_1^-$, $r_{j,2}=a_2^-$, $r_{j,3}=a_3^-$, $j=1,3$ which are associated, respectively, to the three fields defined by one coefficient $a_j^+=1$, and the two other $a_k^+=0$, $k\neq j$ (i.e., only one incident Biot wave in the layer impinging on the lower face), are evaluated to satisfy the boundary conditions at the lower face. The boundary conditions at the lower face are for the bonded layer

$$u_z^s = 0, \tag{14}$$

$$u_z^f = 0, \tag{15}$$

$$u_x^s = 0, \tag{16}$$

and for the layer free on both faces

$$\frac{\sigma_{zz}^f}{\phi} = \frac{\sigma_{zz}^s}{1-\phi}, \tag{17}$$

where ϕ is the porosity

$$\frac{Z}{\cos\theta} = -\frac{p}{v_z}, \tag{18}$$

where p and v_z are the pressure and the z component of the air velocity out of the porous layer, Z is the characteristic impedance of the free air, and

$$\sigma_{x,y}^s = 0. \tag{19}$$

The pressure $p=-\sigma_{zz}^f/\phi$, and the velocity component v_z is given by

$$v_z = -i\omega(1-\phi)u_s^s - i\omega\phi u_z^f. \tag{20}$$

Let $\mathbf{b}_1, \mathbf{b}_2$, and $\mathbf{b}_3$ the frame displacement fields defined by

$$\mathbf{b}_1 = \frac{i}{\omega}[\nabla(\exp(i\alpha_1 z) + r_{1,1}\exp(-i\alpha_1 z) + r_{1,2}\exp(-i\alpha_2 z)) + \nabla \wedge \mathbf{n}r_{1,3}\exp(-i\alpha_3 z)], \tag{21}$$

$$\mathbf{b}_2 = \frac{i}{\omega}[\nabla(r_{2,1}\exp(-i\alpha_1 z) + \exp(i\alpha_2 z) + r_{2,2}\exp(-i\alpha_2 z)) + \nabla \wedge \mathbf{n}r_{2,3}\exp(-i\alpha_3 z)], \tag{22}$$

$$\mathbf{b}_3 = \frac{i}{\omega}[\nabla(r_{3,1}\exp(-i\alpha_1 z) + r_{3,2}\exp(-i\alpha_2 z)) + \nabla \wedge \mathbf{n}(\exp(i\alpha_3 z) + r_{3,3}\exp(-i\alpha_3 z))]. \tag{23}$$

Any linear combination $\Sigma\lambda_j\mathbf{b}_j$ of $\mathbf{b}_1, \mathbf{b}_2$, and $\mathbf{b}_3$ satisfies the boundary conditions at the lower face. Let v_z be the

velocity of air, and p be the pressure, close to the porous surface. At the upper face, the boundary conditions can be written

$$-i\omega(\phi u_z^f + (1-\phi)u_z^s) = v_z, \tag{24}$$

$$\sigma_{zz}^f = -\phi p, \tag{25}$$

$$\sigma_{zz}^s = -(1-\phi)p, \tag{26}$$

$$\sigma_{xz}^s = 0. \tag{27}$$

At the upper face, the frame displacement components u_z and u_x related to $\Sigma\lambda_j \mathbf{b}_j$ are given by

$$u_z(x) = \frac{-1}{\omega}[(\alpha_1(1-r_{1,1}) - \alpha_2 r_{1,2} + \xi r_{1,3})\lambda_1 + (\alpha_2(1 - r_{2,2}) - \alpha_1 r_{2,1} + \xi r_{2,3})\lambda_2 + (-\alpha_1 r_{3,1} - \alpha_2 r_{3,2} + \xi(1+r_{33}))\lambda_3]\exp(ikx\sin\theta), \tag{28}$$

$$u_x(x) = \frac{-1}{\omega}[(\xi(1+r_{1,1}+r_{1,2}) + \alpha_3 r_{1,3})\lambda_1 + (\xi(1+r_{2,2} + r_{2,1}) + \alpha_3 r_{2,3})\lambda_2 + (\xi(r_{3,1}+r_{3,2}) - (1-r_{3,3}))\lambda_3]\exp(ikx\sin\theta). \tag{29}$$

The three coefficients λ_j of the displacement field $\Sigma\lambda_j \mathbf{b}_j$ that satisfies Eqs. (24)–(27) are solutions of the following equations where the coefficient $\exp(ikx\sin\theta)$ has been removed on both sides

$$i\alpha_1(\phi\mu_1 + (1-\phi))[\lambda_1(1-r_{1,1}) - \lambda_2 r_{2,1} - \lambda_3 r_{3,1}] + i\alpha_2(\phi\mu_2 + (1-\phi))[-\lambda_1 r_{1,2} + \lambda_2(1-r_{2,2}) - \lambda_3 r_{3,2}] + i\xi(\phi\mu_3 + (1-\phi))[\lambda_1 r_{1,3} + \lambda_2 r_{2,3} + \lambda_3(1 + r_{3,3})] = v_z, \tag{30}$$

$$\{A_1[\lambda_1(1+r_{1,1}) + \lambda_2 r_{2,1} + \lambda_3 r_{3,1}] + A_2[\lambda_1 r_{1,2} + \lambda_2(1 + r_{2,2}) + \lambda_3 r_{3,2}]\} = -\phi p, \tag{31}$$

$$\{B_1[\lambda_1(1+r_{1,1}) + \lambda_2 r_{2,1} + \lambda_3 r_{3,1}] + B_2[\lambda_1 r_{1,2} + \lambda_2(1 + r_{2,2}) + \lambda_3 r_{3,2}] + C_3[-\lambda_1 r_{1,3} - \lambda_2 r_{2,3} + \lambda_3(1 - r_{3,3})]\} = -(1-\phi)p, \tag{32}$$

$$C_1[\lambda_1(1-r_{1,1}) - \lambda_2 r_{2,1} - \lambda_3 r_{3,1}] + C_2[-\lambda_1 r_{1,2} + \lambda_2(1 - r_{2,2}) - \lambda_3 r_{3,2}] - H[\lambda_1 r_{1,3} + \lambda_2 r_{2,3} + \lambda_3(1+r_{3,3})] = 0. \tag{33}$$

The three last equations provide the ratios $\lambda_1/p, \lambda_2/p$, and λ_3/p, which are given in the Appendix. Reporting these ratios in Eq. (30) provides the surface impedance $Z_s(\theta) = p/v_z$ and the reflection coefficient that is given by

$$R(\theta) = (Z_s(\theta) - Z/\cos\theta)/(Z_s(\theta) + Z/\cos\theta). \tag{34}$$

B. Frame displacement induced by a monopole field

The Sommerfeld representation of the monopole field[19] is used. The unit monopole pressure field $p_m = \exp(ikR_2)/R_2$ can be written, for $z > -h$,

$$p_m(r,z) = i\int_0^\infty \frac{kqdq}{\mu} J_0(kqr)\exp(i\mu k(h+z)), \tag{35}$$

where k is the wave number in air and μ is given by

$$\mu = (1-q^2)^{1/2}, \tag{36}$$

with the conditions $\mu > 0$ if $q < 1$, $\mathrm{Im}\,\mu > 0$ if $q > 1$. The total pressure field at the surface of the layer is given by

$$p_T(r) = i\int_0^\infty \frac{kqdq}{\mu} J_0(kqr)\exp(i\mu kh)(1 + R(\theta)), \tag{37}$$

where R is the plane wave reflection coefficient and θ is defined by $\cos\theta = \mu$, $\sin\theta = q$. The z component and the radial component of the frame velocity at the surface of the layer are given by

$$v_z^s = i\int_0^\infty \frac{kqdq}{\mu} J_0(kqr)\exp(i\mu kh)(1 + R(\theta))\dot{U}_z^s, \tag{38}$$

$$v_r^s = i\int_0^\infty \frac{kqdq}{\mu} J_1(kqr)\exp(i\mu kh)(1 + R(\theta))i\dot{U}_x^s, \tag{39}$$

where $\dot{U}_z^s$ and $\dot{U}_x^s$ are given by Eqs. (A22) and (A23). Predictions and measurements are performed for $h \leqslant 6$ cm and $r \leqslant 20$ cm. Small r and h are chosen to increase the measured velocity of the frame and to minimize the finite dimension effects (with the chosen source-laser beam geometries, as indicated at the beginning of Sec. III, measurements are performed on samples of area around 1 m^2, but this area cannot be strongly reduced). A precise evaluation of the integrals in Eqs. (37)–(39) using the trapezoidal rule can be obtained under two conditions: The integration must be performed from 0 to a limit q_M where the argument of the integral is sufficiency small, and the interval Δq must be sufficiently small. For $h \geqslant 2$ cm and $r \leqslant 1$ m, a limit $q_M = 10$ and an interval $\Delta q = 10^{-4}$ can be used. A larger limit q_M should be used for smaller h because the decreasing of the argument is mainly due to $\exp(i\mu kh)$. Predictions have been performed for a material with a small Poisson ratio $\nu = 0.3$, and a material with a large Poisson ratio $\nu = 0.44$. For the material with the low Poisson ratio, predictions for the normal velocity and the radial velocity of the frame at the surface of a bonded porous layer of thickness $l = 1$ cm are given in Fig. 3 for a height of the source $h = 2$ cm and a radial distance from the source to the point where the velocities are predicted $r = 6$ cm. In Fig. 4, $h = 6$ cm and $r = 2$ cm. The parameters that characterize the layer are given in Table I for material 1. The Biot theory was used for the predictions, with the model of Johnson *et al.*[20] for the effective density and of Lafarge[21] for the incompressibility of air. The predicted velocity dependences as a function of frequency present large peaks related to resonances of the frame. The speeds $c_i = \omega/\mathrm{Re}\,k_i$ of the compressional and the shear wave are given in Table II with the frequencies $f_i = c_i/(4l)$ of the $\lambda/4$ resonances for $\nu = 0.3$.

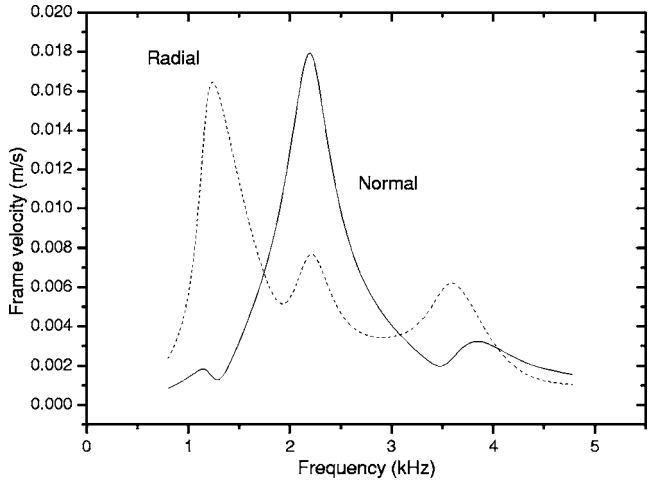


FIG. 3. Radial and normal frame velocity, material 1 bonded to a rigid backing, $h=2$ cm, $r=6$ cm.

The compression speed slightly depends on frequency and is given close to the compression resonance. The frequency for the compression resonance is 2.317 kHz and the maximum of the normal velocity is located at 2.22 kHz. The frequency for the shear resonance is 1.165 kHz and the maximum for the radial velocity is located at 1.24 kHz. The peak for the radial velocity only appears when the ratio r/h is sufficiently large and is not present in Fig. 4. The maxima of the radial and the normal velocity are related to the $\lambda/4$ shear and compressional resonances. Similar distributions can be predicted for the free layer. Peaks appear for the normal velocity, and the radial velocity, close to the half wavelength resonance frequencies. The identification of the peaks can lead to an evaluation of the rigidity parameters of the porous frame.

The material with the large Poisson ratio is characterized by the same parameters as material 1 except the Poisson ratio $\nu=0.44$. The peaks for the compressional waves must be located at frequencies 3.06 as large as the related peaks for the shear waves. The predicted peaks satisfy the relation with a reasonable approximation, but for the bonded layer, a new peak appears at a frequency lower than for the compressional resonance. The normal velocity is represented in Fig. 5. The

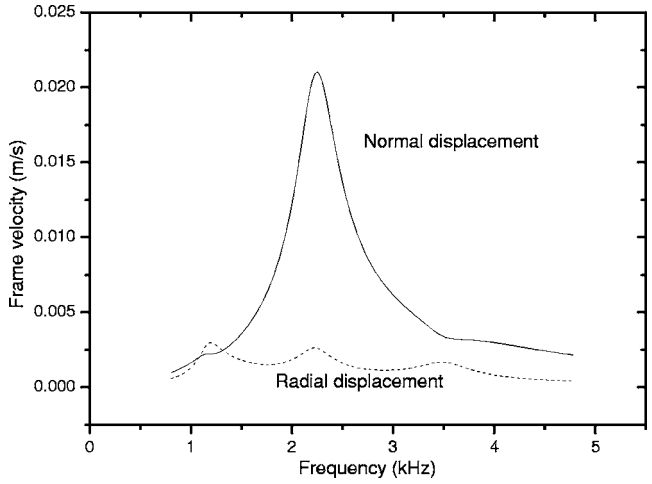


FIG. 4. Radial and normal frame velocity, material 1 bonded to a rigid backing, $h=6$ cm, $r=2$ cm.

TABLE I. Parameters for the porous materials.

Parameters	Material 1	Material 2
Porosity ϕ	0.99	0.97
Flow resistivity σ (Nm^{-4} s)	130000	22100
Thermal permeability k_0' (m^2)	5×10^{-10}	0.83×10^{-9}
Viscous dimension Λ (μm)	60	39
Thermal dimension Λ' (μm)	180	275
Tortuosity α_∞	1.1	2.2
Density ρ_s (kg/m^3)	59	24.5
Shear modulus N (kPa)	$130-i13$	$80-i12$
Poisson ratio ν	0.3	0.44
Thickness l (cm)	1	2

new peak is centered close to 3.2 kHz, and is larger than the compressional peak at its right hand side. The presence of this new peak when the Poisson ratio increases beyond 0.4 has been observed for all the different soft porous layers we have considered. The shape of this new peak is different from the one of the compressional peak for $\nu=0.3$, it is not symmetrical, but confusions could arise when the porous layer has not a sufficient lateral extension, that can lead to modifications of the shape of the peaks. The frequency of the $\lambda/4$ compressional resonance increases with the Poisson ratio and the occurrence of a peak detectable with the same source-laser beam geometry at a lower frequency is not surprising. Nevertheless, it has not been possible to find a simple physical description of its origin.

III. MEASUREMENTS

Measurements have been performed on a layer of urethane foam of thickness $l=2$ cm and area 0.75×1.5 m^2. Measurements have also been performed on samples of smaller area close to 0.4×0.4 m^2. With the samples of small area, the finite dimension effects become noticeable and the measurements cannot be related to the predictions obtained from the model of Sec. II. Measurements of the normal velocity for the bonded layer have been performed with $r=h=3.5$ cm, and measurements of the radial velocity have been performed for $r=16.5$ cm and $h=3.5$ cm. Measurements of the normal velocity for the free layer have been performed with $r=0.5$ cm and $h=3.5$ cm, and measurements of the radial velocity have been performed for $r=20$ cm and $h=2$ cm. The source is a pipe fed by a compression driver and a microphone is located at the end of the tube to provide a

TABLE II. Phase speeds and resonance frequencies

	Material 1	Material 1	Material 2
	$\nu=0.3$	$\nu=0.44$	
Shear wave speed (m/s)	46.6	46.6	56.4
$\lambda/4$ shear resonance frequency (kHz)	1.165	1.165	0.706
Compression wave speed (m/s)	92.7	147.8	186
$\lambda/4$ compression resonance frequency (kHz)	2.317	3.695	2.325

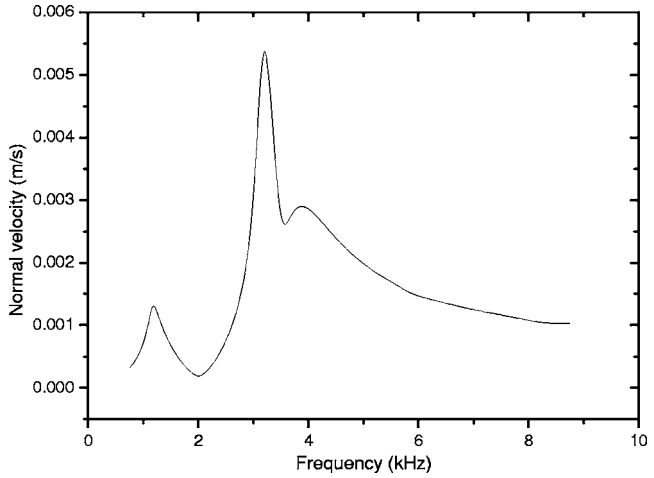


FIG. 5. Normal velocity, material 1 bonded, new Poisson coefficient $\nu=0.44, h=5$ cm, $r=0.5$ cm.

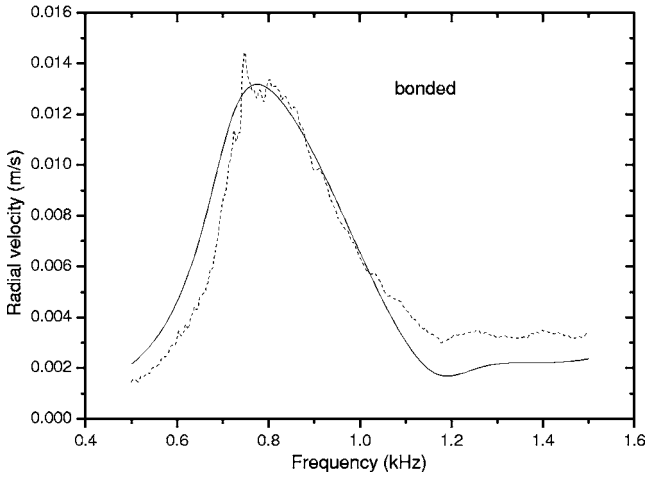


FIG. 7. Radial velocity, material 2 bonded, $r=16.5$ cm, $h=3.5$ cm, prediction (dashed line), measured transfer function between the laser velocity signal and the pressure signal at the end of the tube (solid line).

reference pressure signal in the frequency domain. The internal diameter of the pipe is equal to 16 mm. The vertical velocity is measured with a laser beam perpendicular to the surface of the layer, and the radial velocity with an angle of incidence of 66°. The parameters used for the predictions are given in Table I for material 2. The flow resistivity σ and the porosity ϕ are measured, and the set of acoustical parameters (σ, ϕ, viscous and thermal dimensions Λ and Λ', tortuosity α_∞ and thermal permeability k_0') is chosen to give predictions of the absorption coefficient with the equivalent fluid model in a good agreement with measurements performed in a Kundt tube following the double thickness algorithm presented in Ref. 22. With the equivalent fluid model, the frame is motionless and the air in the porous medium is replaced by a fluid with a density described with the model of Johnson *et al.*[20] and an incompressibility described with the model by Lafarge.[21] The three mechanical parameters are the density of the frame ρ_s, the Poisson ratio ν, and the shear modulus N. The density ρ_s is measured and ν and N are chosen to adjust the measured velocity peaks and the predicted normal and radial velocities for the bonded and the free layer. For the bonded layer, the predicted vertical velocity related to the unit source is presented in Fig. 6 and the radial velocity in Fig. 7. For the free layer, the predicted vertical velocity related to the unit source is presented in Fig. 8 and the radial velocity in Fig. 9. The predicted velocities are compared to the ratios of the velocity signals to the reference pressure signals in the frequency domain, normalized by a constant factor in order to make the comparison of the frequency dependences easier. All the theoretical and experimental velocity distributions present one main peak and the presented measurements are restricted to a restricted frequency range around the maximum in order to optimize its description. The location of the predicted peaks for the radial velocity mainly depends on the shear modulus and the location of the peaks for the radial velocity depends on the shear modulus and the Poisson coefficient. Under the hypothesis that the shear modulus does not noticeably vary between both peaks, it is possible to determine both coefficients from the measurement of the radial and the vertical velocities. A good agreement for the studied material is obtained for both normal velocities and for the radial velocity of the bonded layer for $N=80-i12$ kPa and $\nu=0.44$. The measured peak for the

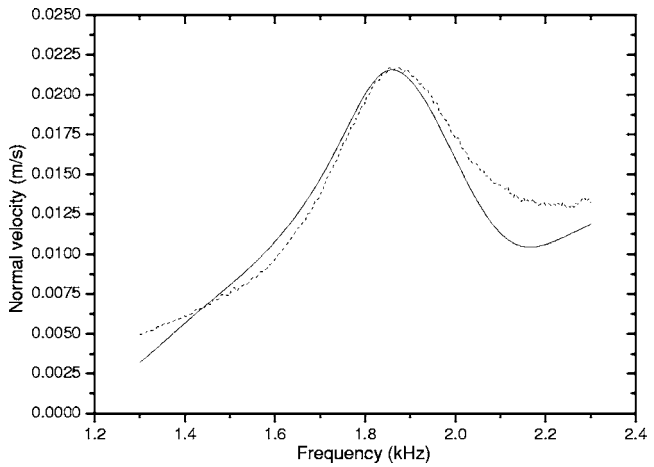


FIG. 6. Normal velocity, material 2 bonded, $r=h=3.5$ cm, prediction (dashed line), measured transfer function between the laser velocity signal and the pressure signal at the end of the tube (solid line).

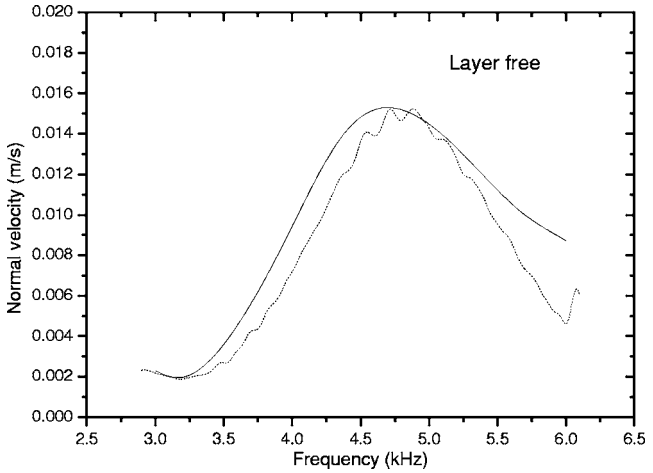


FIG. 8. Vertical velocity, material 2 free, $r=0.5$ cm, $h=3.5$ cm, prediction (dashed line), measured transfer function between the laser velocity signal and the pressure signal at the end of the tube (solid line).

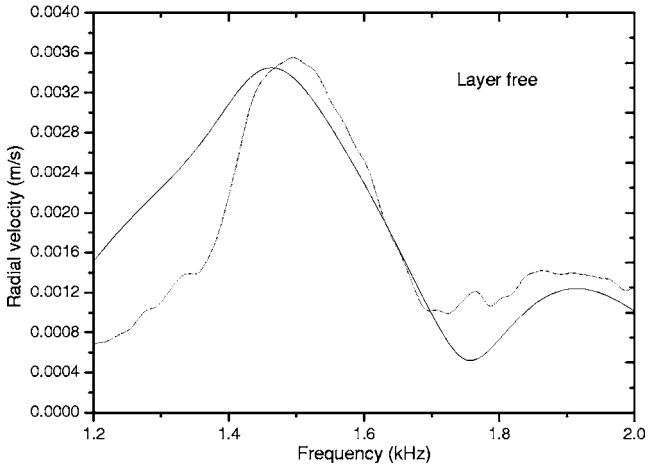


FIG. 9. Radial velocity, material 2 free, $r=20$ cm, $h=2$ cm, prediction (dashed line) measured transfer function between the laser velocity signal and the pressure signal at the end of the tube (solid line).

radial velocity of the free layer has not the same shape as the predicted one, probably because the finite dimension of the sample. With this material, the main peak for the normal velocity of the bonded material is not related to the quarter compressional wavelength resonance but to the main peak of Fig. 5 but an unambiguous determination of the rigidity parameters remains possible. Note that for this material, quasistatic measurements give a Poisson ratio $\nu=0.43$ close to the measurement at audible frequencies, but the quasistatic shear modulus $N=20.5(1-i0.08)$ kPa is much smaller than the shear modulus at audible frequencies.

IV. CONCLUSION

The excitation by a point source in air of different modes of vibration of a porous air saturated layer bonded to a rigid backing or free on both faces has been studied. The normal and the radial velocity created by the sound source at the surface of the structure has been predicted. The distance from the point source to the point of observation of the velocity is smaller or of the same order of magnitude than the wavelength in air and the Sommerfeld representation of the monopole field can provide precise predictions of the velocity. The manifestations of the shear and the quarter wavelength resonance, and of the shear and the compressional half wavelength resonance, can be observed separately without any processing of the velocity signals. The compressional resonances appear at a much larger frequency than the shear resonances, and mainly as a peak in the normal velocity. The shear resonances appear mainly as a peak in the radial velocity only if the radial coordinate of the point of observation is sufficiently large. The presence of an extra peak of the normal velocity for bonded materials with a large Poisson coefficient at a frequency lower than the one of the compressions resonance can complicate the evaluation of the rigidity coefficients from the observation of the peaks.

APPENDIX

In what follows we use capital letters to denote the quantities σ_{xy}^s, iu_z^s, iu_z^f, σ_{zz}^f, σ_{zz}^s, and iu_z^s when the coefficient $\exp(ikx\sin\theta)$ is removed. The new quantities can be written

$$\Sigma_{xy}^s = M_{11}\lambda_1 + M_{12}\lambda_2 + M_{13}\lambda_{33}, \tag{A1}$$

$$\dot{U}_z^s = M_{21}\lambda_1 + M_{22}\lambda_2 + M_{23}\lambda_3, \tag{A2}$$

$$\dot{U}_z^f = M_{31}\lambda_1 + M_{32}\lambda_2 + M_{33}\lambda_3, \tag{A3}$$

$$\Sigma_{zz}^f = M_{41}\lambda_1 + M_{42}\lambda_2 + M_{43}\lambda_3, \tag{A4}$$

$$\Sigma_{zz}^s = M_{51}\lambda_1 + M_{52}\lambda_2 + M_{53}\lambda_3, \tag{A5}$$

$$\dot{U}_x^s = M_{61}\lambda_1 + M_{62}\lambda_2 + M_{63}\lambda_3. \tag{A6}$$

The coefficients M_{ij} can be obtained from Eqs. (24)–(29). The condition $\sigma_{xy}^s=0$ leads to

$$\lambda_3 = -\frac{M_{11}}{M_{13}}\lambda_1 - \frac{M_{12}}{M_{13}}\lambda_2, \tag{A7}$$

and $\dot{U}_z^s$, $\dot{U}_z^f$, Σ_{zz}^f, Σ_{zz}^s, and $\dot{U}_x^s$ can be expressed as functions of λ_1 and λ_2,

$$\dot{U}_z^s = \left(M_{21} - \frac{M_{11}}{M_{13}}M_{23}\right)\lambda_1 + \left(M_{22} - \frac{M_{12}}{M_{13}}M_{23}\right)\lambda_2, \tag{A8}$$

$$\dot{U}_z^f = \left(M_{31} - \frac{M_{11}}{M_{13}}M_{33}\right)\lambda_1 + \left(M_{32} - \frac{M_{12}}{M_{13}}M_{33}\right)\lambda_2, \tag{A9}$$

$$\Sigma_{zz}^f = \left(M_{41} - \frac{M_{11}}{M_{13}}M_{43}\right)\lambda_1 + \left(M_{42} - \frac{M_{12}}{M_{13}}M_{43}\right)\lambda_2, \tag{A10}$$

$$\Sigma_{zz}^s = \left(M_{51} - \frac{M_{11}}{M_{13}}M_{53}\right)\lambda_1 + \left(M_{52} - \frac{M_{12}}{M_{13}}M_{53}\right)\lambda_2, \tag{A11}$$

$$\dot{U}_x^s = \left(M_{61} - \frac{M_{11}}{M_{13}}M_{63}\right)\lambda_1 + \left(M_{62} - \frac{M_{12}}{M_{13}}M_{63}\right)\lambda_2. \tag{A12}$$

The symbols T_{ij} are defined by the relations

$$\dot{U}_z^s = T_{11}\lambda_1 + T_{12}\lambda_2, \tag{A13}$$

$$\dot{U}_z^f = T_{21}\lambda_1 + T_{22}\lambda_2, \tag{A14}$$

$$\Sigma_{zz}^f = T_{31}\lambda_1 + T_{32}\lambda_2, \tag{A15}$$

$$\Sigma_{zz}^s = T_{41}\lambda_1 + T_{42}\lambda_2, \tag{A16}$$

$$\dot{U}_x^s = T_{51}\lambda_1 + T_{52}\lambda_2. \tag{A17}$$

Using the relation $\phi\sigma_{zz}^s=(1-\phi)\sigma_{zz}^f$ gives the ratio $s_1=\lambda_2/\lambda_1$,

$$s_1 = -\frac{\phi T_{41} - (1-\phi)T_{31}}{\phi T_{42} - (1-\phi)T_{32}}. \tag{A18}$$

The surface impedance $Z_s=-\sigma_{zz}^f/(\phi v_z)$ is given by

$$Z_s = -\frac{1}{\phi}\frac{(T_{31} + T_{32}s_1)}{D}, \tag{A19}$$

where D is given by

$$D = \phi T_{21} + (1-\phi)T_{11} + s_1(\phi T_{22} + (1-\phi)T_{12}). \tag{A20}$$

The coefficient λ_1 related to a unit pressure is given by

$$\lambda_1 = \frac{-\phi}{T_{31} + s_1 T_{32}}. \tag{A21}$$

The z component and the x component of the velocity per unit pressure are given by, respectively,

$$\dot{U}_z^s = -\phi\frac{T_{11} + T_{12}s_1}{T_{31} + T_{32}s_1}, \tag{A22}$$

$$\dot{U}_x^s = -\phi\frac{T_{51} + T_{52}s_1}{T_{31} + T_{32}s_1}. \tag{A23}$$

These velocity components are related to a unit pressure, and they must be multiplied by $(1+R)$ to obtain the velocity components V_z and V_x per unit incident pressure

$$1 + R = \frac{-2(T_{31} + T_{32}s_1)/\phi}{-(T_{31} + T_{32}s_1)/\phi + DZ/\cos\theta}. \tag{A24}$$

The component V_z is given by

$$V_z = \frac{2(T_{11} + T_{12}s_1)}{-(T_{31} + T_{32}s_1)/\phi + ZD/\cos\theta}. \tag{A25}$$

The component V_x is given by

$$V_x = \frac{2(T_{51} + T_{52}s_1)}{-(T_{31} + T_{32}s_1)/\phi + ZD/\cos\theta}. \tag{A26}$$

[1]T. Pritz, "Transfer function method for investigating the complex modulus of acoustic materials: spring-like specimen," J. Sound Vib. **72**, 317–341 (1980).

[2]T. Pritz, "Transfer function method for investigating the complex modulus of acoustic materials: rod-like specimen," J. Sound Vib. **81**, 359–376 (1982).

[3]T. Pritz, "Frequency dependence of frame dynamic characteristics of mineral and glass wool materials," J. Sound Vib. **106**, 161–169 (1986).

[4]T. Pritz, "Dynamic Young's modulus and loss factor of plastic foams for impact sound isolation," J. Sound Vib. **178**, 315–322 (1994).

[5]A. Sfaoui, "On the viscoelasticity of the polyurethane foam," J. Acoust. Soc. Am. **97**, 1046–1052 (1995).

[6]T. Pritz, "Measurement methods of complex Poisson's ratio of viscoelastic materials," Appl. Acoust. **60**, 279–292 (2000).

[7]M. Melon, M. Mariez, C. Ayrault, and S. Sahraoui, "Acoustical and mechanical characterization of anisotropic open-cell foams," J. Acoust. Soc. Am. **104**, 2622–2627 (1998).

[8]J. Park, "Measurement of the frame acoustic properties of porous and granular materials," J. Acoust. Soc. Am. **118**, 3483–3490 (2005).

[9]V. Tarnow, "Dynamic measurements of the elastic constants of glass wool," J. Acoust. Soc. Am. **118**, 3672–3678 (2005).

[10]J. F. Allard, G. Jansens, G. Vermeir, and W. Lauriks, "Frame-borne surface waves in air-saturated porous media," J. Acoust. Soc. Am. **111**, 690–696 (2002).

[11]L. Boeckx, P. Leclaire, P. Khurana, C. Glorieux, W. Lauriks, and J. F. Allard, "Investigation of the phase velocity of guided acoustic waves in soft porous layers," J. Acoust. Soc. Am. **117**, 545–554 (2005).

[12]L. Boeckx, P. Leclaire, P. Khurana, C. Glorieux, W. Lauriks, and J. F. Allard, "Guided waves at audible frequencies in poroelastic layers in the Lamb conditions," J. Appl. Phys. **97**, 094911 (2005).

[13]M. A. Biot, "Theory of propagation of elastic waves in a fluid-saturated porous solid," J. Acoust. Soc. Am. **28**, 168–191 (1956).

[14]J. F. Allard, *Propagation of Sound in Porous Media, Modelling Sound Absorbing Materials* (Elsevier, London, 1993).

[15]K. Attenborough and T. L. Richards, "Solid particle motion induced by a point source above a poroelastic half-space," J. Acoust. Soc. Am. **86**, 1085–1091 (1989).

[16]J. F. Allard, M. Henry, V. Gareton, G. Jansens, and W. Lauriks, "Impedance measurement around grazing incidence for nonlocally reacting thin porous layers," J. Acoust. Soc. Am. **113**, 1210–1215 (2003).

[17]J. F. Allard, M. Henry, L. Boeckx, P. Leclaire, and W. Lauriks, "Acoustical measurement of the shear modulus of thin porous layers," J. Acoust. Soc. Am. **117**, 1737–1744 (2005).

[18]M. A. Biot and D. G. Willis, "The elastic coefficients of the theory of consolidation," J. Appl. Mech. **24**, 594–601 (1957).

[19]L. M. Brekhovskikh and O. A. Godin, *Acoustics of Layered Media II, Point Source and Bounded Beams*, Springer Series on Wave Phenomena (Springer, New York, 1992).

[20]D. L. Johnson, J. Koplik, and R. Dashen, "Theory of dynamic permeability and tortuosity in fluid-saturated porous media," J. Fluid Mech. **176**, 379–402 (1987).

[21]D. Lafarge, P. Lemarinier, J. F. Allard, and V. Tarnow, "Dynamic compressibility of air in porous structures at audible frequencies," J. Acoust. Soc. Am. **102**, 1995–2006 (1997).

[22]Y. Atalla and R. Panneton, "Inverse acoustical characterization of open cell porous media using impedance tube measurements," Can. Acoust. **33**, 11–24 (2005).

Measurements of sound propagation in a littoral environment using a vertical synthetic array

Harry J. Simpson
Physical Acoustics Branch Code 7136, Naval Research Laboratory, Washington, DC 20375

Carl K. Frederickson
Department of Physics and Astronomy, University of Central Arkansas, Conway, Arkansas 72035

Erik C. Porse and Brian H. Houston
Physical Acoustics Branch Code 7136, Naval Research Laboratory, Washington, DC 20375

Larry A. Kraus, Steve W. Liskey, Alain R. Berdoz, and Philip A. Frank
SFA Inc., Largo, Maryland 20774

Steve Stanic
Acoustic Simulations, Measurements and Tactics Code 7180, Naval Research Laboratory, Stennis Space Center, Mississippi 39529

(Received 15 June 2006; revised 25 September 2006; accepted 25 September 2006)

Shallow-water propagation experiments were carried out in St. Andrews Bay, Florida. These investigations used a vertical one-dimensional synthetic array measurement system where two hydrophones incrementally mapped the acoustic pressure from 2 m below through 2 m above the sand-water interface. A broadband (1 to 12 kHz) chirp was used to insonify the water channel. The source to receiver distance was varied and included measurements at 20, 30, 50, 70, and 90 m. These measurements quantified the acoustic pressure above and below a sandy bottom and the results are presented along with frequency analysis, temporal impulse analysis, and wave number analysis. To obtain a better understanding of the results, the measurements are compared to two numerical models. The first model is a temporal ray path prediction of sound propagation in the water channel. The second model, range-dependent acoustic model (RAM), based on a parabolic equation, predicts the sound propagating in a water channel with a sandy bottom. The experimental results agreed well with both numerical predictions in the water column. However, the measured acoustic energy in the sandy bottom was different from the prediction by RAM, which assumed a smooth air-water and water-bottom interface and a point source. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2372449]

PACS number(s): 43.30.Cq, 43.20.Dk, 43.20.El, 43.30.Zk [RAS] Pages: 85–97

I. INTRODUCTION

We present a series of acoustic propagation measurements designed to study and quantify the pressure just above and below a sand-water interface in a littoral environment. The primary motivation for this work is to develop techniques that detect and identify proud and buried objects in sandy, shallow-water regions. To facilitate this development, it is essential to understand the physics of sound propagation within a shallow-water channel and how the acoustic pressure changes with increasing range above and below the sand-water interface.

To date there are numerous investigations of sound propagation in shallow water for frequencies above 20 kHz,[1–7] and there are several studies of shallow-water propagations for frequencies below 1 kHz.[8–21] However, there are only a few investigations of sound propagation in the 1 to 12-kHz band.[22–27] This frequency band offers several distinct advantages over the higher and the lower frequency ranges. Most targets of interest in littoral environments are on the size of meters. Therefore, for frequencies above 1 kHz, the target is in the Kirchhoff scattering region, $ka > 1$. For lower frequencies, where $ka \ll 1$, Rayleigh scattering physics makes it difficult to detect targets. For higher frequencies, above 20 kHz, those traditionally used for acoustic imaging, the propagation range is limited. The 1 to 12-kHz band provides for good backscattering from targets with dimensions on the order of meters, while still facilitating long-range propagation.

Previous work in St. Andrews Bay[24,25] showed the effectiveness of a synthetic array technique in quantifying sound speeds and attenuations in the first 2 m of a sandy bottom. This experiment quantified the total acoustic pressure above and below a sandy bottom that propagates in the shallow-water channel. To study the propagation of the acoustic pressure in a shallow-water channel, a system was built that would minimally disturb the measurement area. The advantage of this system[24,25] is the minimal impact on the immediate area surrounding the insertion of a single hydrophone into the bottom. Also, with only one small hydrophone inserted into the bottom, there is no self-noise due to scattering from nearby hydrophones or scattering from the support hardware, such as other electronic cables and preamplifiers. Although surface roughness is still an issue for scat-

tering, St. Andrews Bay was chosen because there is very little current and wave action and thus very little ripple formation on the sandy bottom. In this environment, the experiment can address the feasibility of a synthetic array system for measuring propagation in a shallow-water channel as well as measure the sound energy in the bottom as a function of distance from the source. This synthetic array system was also developed to provide acoustic information without extensive measurements of all the surrounding environmental parameters that are usually needed to predict the acoustic pressure in the bottom. Instead, the synthetic array system can be used to quickly measure the sound pressure at any location, for example next to a proud or buried target, with a minimum of disturbance to the immediate area. As a secondary issue, these experiments were designed to measure slower compressional waves, propagating in the sandy bottom, that are predicted to occur for subcritical insonification of sandy bottoms. Previous work by Chotiros[28] suggested a second slower compressional wave exists in a sandy bottom, but was not measured in sound-speed measurements made by Simpson[24,25] in previous St. Andrews Bay experiments. Some conjecture has been put forth that the slower compressional wave will only be seen for subcritical insonification of the bottom. The results of the subcritical compressional acoustic pressure in this bottom will be discussed in detail later in the paper.

The investigations presented in this paper are focused on quantifying the pressure above and below the sand-water interface as sound propagates in a shallow-water channel. The horizontal source-to-receiver distance was varied from 20 m up to 90 m. The measurements were conducted in St. Andrews Bay, Florida, in a ~6-m-deep water channel above a sandy bottom. The sandy bottom had no ripple structure and the surface topography was consistent with random biological activities. The measurement site was chosen in an area where the depth varied less than a meter over the 100-m measurement range. The overall water depth varied over the course of the day due to tides and ranged from 6 to 7 m. Some environmental parameters were measured at the site and are referenced in a publication by Simpson *et al.*[25] that detailed a set of bottom penetration experiments that quantified sound speeds and attenuations in this sandy bottom for insonification normal to the bottom.

The shallow-water acoustic propagation is measured above and below the sandy bottom using a synthetic array. The data are analyzed in the time domain, frequency domain, and wave number domain. The experimental data are compared to two different numerical models. The first is a simple ray model of sound propagating though the water column. The second uses the parabolic equation to predict the sound propagating in both the water channel and the sediment.

This paper is focused on reporting the experimental results. Some idealized numerical models, along with other analysis techniques described later in the paper, are used to gain insight into the experimental results. However, the numerical models are not the focus of this paper and the idealized environmental conditions assumed for the calculations were chosen only to aid in the understanding of the measured results. A more robust set of environmental conditions including interface roughness, bottom inhomogeneities, water column variations, and surface wave motion would need to be included to conduct a more precise numerical study. However, extensive measurements of these physical parameters were cost prohibitive and outside the objectives of the measurement exercise. Thus, this level of numerical investigation is also outside the scope of this experimental paper.

The remainder of the paper will address the experimental system and setup in the next section. The following section will discuss the measurements and the last section will be a discussion of the experimental results. This will included time, frequency, and wave number analysis along with comparisons to a ray model and to range-dependent acoustic model (RAM).

II. MEASUREMENT SYSTEM AND SETUP

The measurement system consists of a low-frequency (1 to 12 kHz) source [Fig. 1(a)] and two hydrophones mounted to a robotics system [Fig. 1(b)] that are used to synthesize a vertical array. The measurement system was set on a sandy bottom in St. Andrews Bay, Florida, [Fig. 1(c)]. The water depth was approximately 6.5 m and the source was mounted 2.6 m above the sandy bottom. This system was used to map the pressure at 20, 30, 50, 70, and 90 m from the source. Two of the hydrophones were mounted to a thin (1.3 mm) stainless-steel cable. These hydrophones were mounted 2.0 m apart and the cable was attached to a vertical linear robot. The system was designed to have one hydrophone buried in the sandy bottom (initially 2 m below the sand-water interface) and the second hydrophone mounted 2 m above the buried hydrophone (initially at the sand-water interface). Using the vertical robot, both hydrophones were pulled up 2 m in 10-cm increments for the 20-, 30-, and 50 -m measurements and 5-cm increments for the 70- and 90 -m measurements. In this way, a 4-m synthetic array was formed to map the pressure normal to the sand-water interface. A third hydrophone was mounted to the bottom of the robotics stage and remained stationary throughout the synthetic array measurements to provide measurement stability data.

The overall measurement system consists of a vertical robotics stage and a frequency source mounted above the sea floor, as shown in Fig. 1(c). The receiver system consists of a 5.5-m-long aluminum I-beam that is mounted 3.7 m above the bottom on "A-frame" legs shown in Fig. 1(b). A 3-m vertical robot is mounted to the A-frame and is used to pull a buried hydrophone up through the bottom.[25] The source is mounted atop a tower with a pitch and yaw robotics system that is used to control the direction of the transducer, shown in Fig. 1(a). The angles range from 7.4° to 1.7°, measured up from the sand-water interface to a line drawn from the source to insertion point of the hydrophone array, and the horizontal source-to-receiver distances range from 20 to 90 m. The critical angle for this sand-water interface is approximately 25°, for a water speed of 1535 m/s and a bottom speed of 1700 m/s; thus, all these measurements are far below the critical angle for the direct ray path from the source to the insertion point of the receiver.

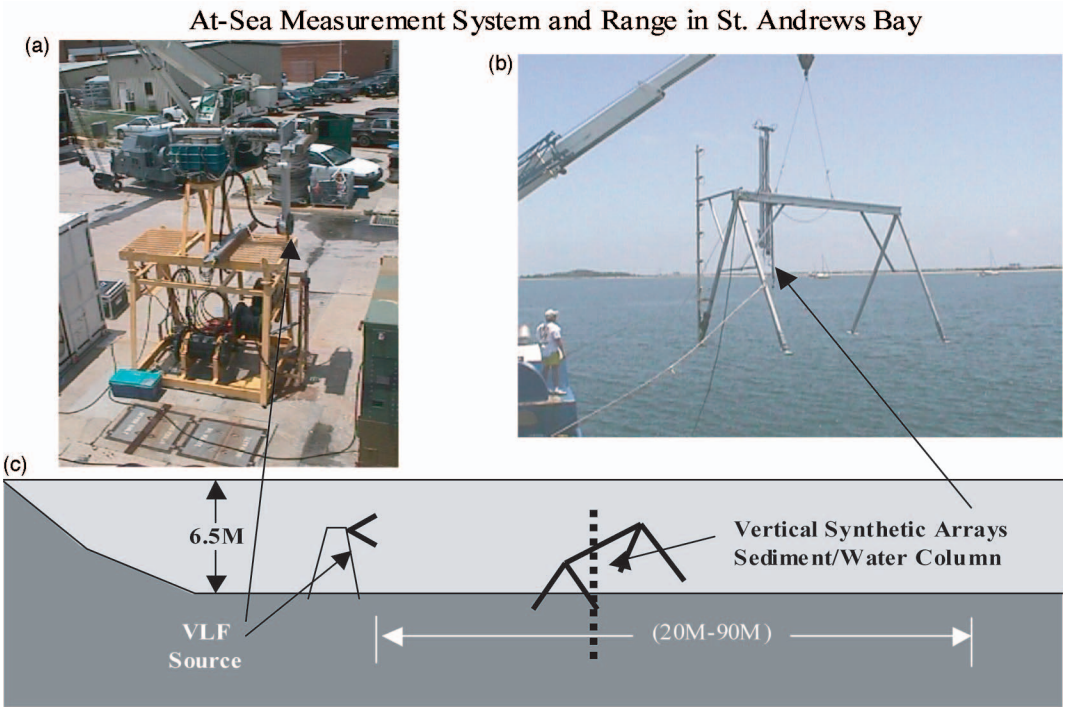


FIG. 1. (Color online) The measurement system and configuration used in St. Andrews Bay. (a) is a picture of the source tower and the source that was used for the measurements; (b) is a picture of the receiver system including the aluminum backbone and the vertical robot. (c) is a diagram of the measurement configuration set on the sandy bottom of the bay.

Note that this 25° critical angle is different than the ~27° angle reported by Simpson *et al.*[25] This difference is a result of the dispersion in the wave speed measured in the sandy bottom. For frequencies above 20 kHz the sound speed[25] is nominally 1720 m/s and for frequencies below 20 kHz the sound speed[25] is nominally 1700 m/s.

Divers are used to set the measurement system on the seafloor and bury the hydrophone with a water jet. The water jet consists of a stainless-steel tube with a 1.9-cm outer diameter that is 3 m long. Typically, at each new site, the water jet is first inserted into the bottom without the hydrophone to clear a pathway for the hydrophone. Care was taken in this process to insure that the water jet system is free of air bubbles. Once the path is cleared to approximately 2.5 m below the sand-water interface, the hydrophone is inserted with the water jet. The water jet is then removed and the sand-water interface is shaped to its original contour. This burial technique is identical to that used for previous measurements,[24,25] and minimally disturbs the sediment, which solidifies immediately after the water jet is turned off. Laboratory and at-sea tests show that acoustically the affected sediment, from the water jet, is indistinguishable from the surrounding media within the few minutes it takes the divers to clear the area and for the measurements to start.

Cables from the deployed measurement system are used to connect the underwater electronics with shipboard control and data acquisition systems.[25] In the measurements reported here, the computer outputs a shaped chirp that is bandpass filtered using an Ithaco 4302 set at 250 Hz high pass and 12.5 kHz low pass. The filtered signal is amplified using an Instruments Inc. model L6 power amplifier that drives the source.

The hydrophone (a Celesco LC10) was connected to a preamplifier with a gain of 10 (20 dB). The signal is further amplified with an Ithaco 1201 that provides gains of 20 to 200. The signal was bandpass filtered using an Ithaco 4302 filter set to 250 Hz high pass and 12.5 kHz low pass and then digitized. For the measurements reported here, 36 ensembles were recorded for each hydrophone position in the synthetic array.

The data acquisition system and robotics system are identical to the system used by Simpson *et al.*[25] The source (a USRD type G34 transducer) was configured to produce a 4-ms-long chirp with a bandwidth of 1 to 12 kHz. Figure 2(a) shows the temporal response and Fig. 2(b) shows the frequency response of the source-receiver system normalized to 1 m. This reference measurement was made in a pool laboratory at NRL.

III. MEASUREMENTS

Using the measurement system, a synthetic vertical array mapped the acoustic pressure from 2 m below through 2 m above the sand-water interface. At each receiver position 36 pings were recorded before the hydrophone was pulled up through the sand or raised in the water column. The ensemble data were acquired over ~36 s and analyzed for temporal stability and coherence. The ping-to-ping variability for each point was typically less than a few percent. As a result of this good temporal stability, the ensemble was averaged with no further processing. The entire data acquisition for a single synthetic array required between ~20 min for the 20-, 30-, and 50-m measurements and ~40 min for the 70- and 90-m measurements. Figure 3 shows the frequency response

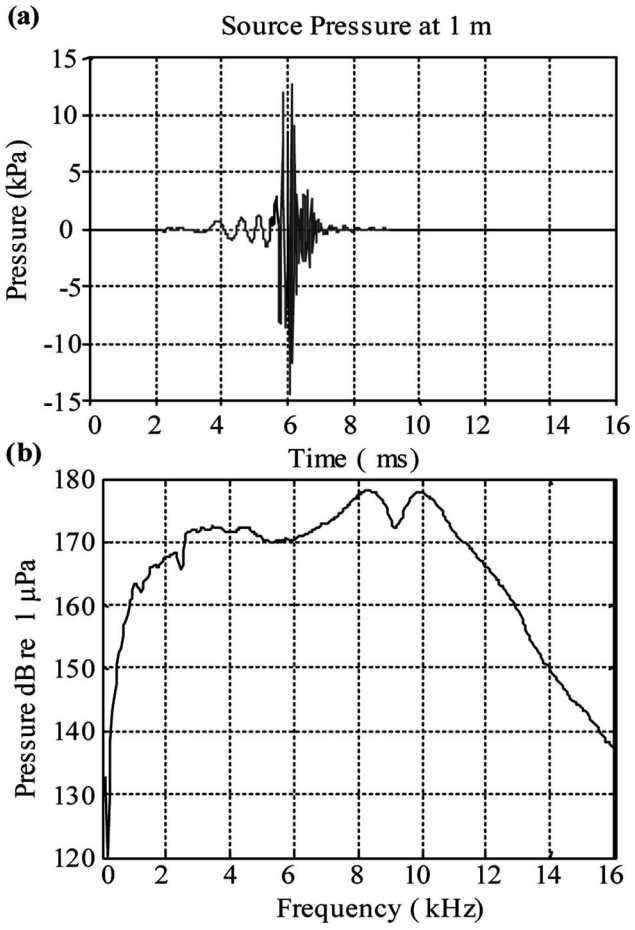


FIG. 2. The free-field temporal and frequency response of the source-receiver system normalized to a 1-m separation. The temporal response shown in (a) is the measurement of a 4-ms-long chirp designed to provide a response from 1 to 12 kHz. (b) is the frequency response of the source-receiver system to the chirp measured at an NRL pool laboratory.

of the pressure 50 m from the source. The horizontal axis of Fig. 3 is frequency and the vertical axis is the distance above and below the sand-water interface. The bottom was relatively undisturbed except for a small area where the hydrophone was water-jetted into the bottom. Figure 3(a) clearly shows a measurable amount of energy in the first 2 m of the bottom for frequencies below 3 kHz and very little measurable energy in the bottom for frequencies above 3 kHz. The insonification angle of the direct ray, at the insertion point, is 3.0°, well below the 25° critical angle.

In the first 2 m of the water column [Fig. 3(a)] there is energy throughout the 12-kHz band of the measurement. The interference pattern in the water column will be discussed in more detail later and is a result of the multiple paths the sound travels in this littoral waveguide, as depicted in Fig. 3(b).

The 50-m measurement is shown as a typical measurement because it lies in the middle of the 20 to 90-m range. Some results from the other measurements will be discussed later in the paper and will be used to emphasize specific points. Notable differences in the experimental procedure used for these measurements include the fact that the record lengths were increased with increased range, and the number of array points was doubled for the 70 and 90-m ranges.

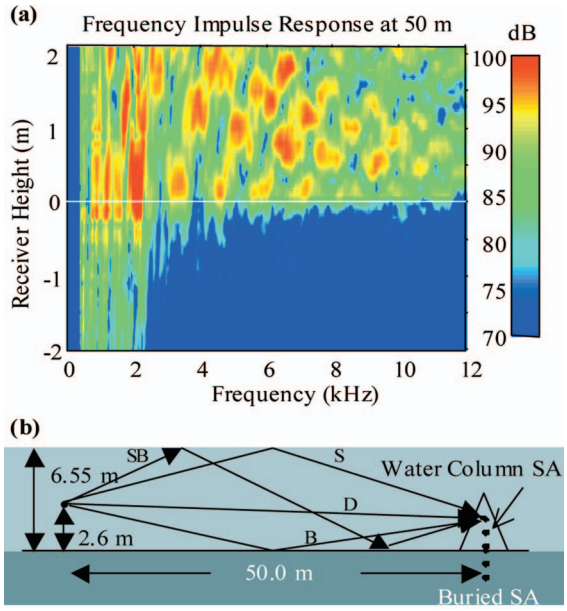


FIG. 3. (Color online) (a) is the frequency response from 2 m above through 2 m below the sandy bottom for the receivers set 50 m horizontal range from the source. The bottom diagram (b) shows the measurement configuration on the seafloor for these 50-m data and shows the first three ray paths to the receiver array.

IV. DISCUSSION

A. Temporal measurement analysis

These measurements were designed to quantify the 1 to 12-kHz pressure near the sand-water interface. We studied the pressure, relatively close to the source (under 100 m), to quantify the energy both in the bottom and in the water column that would scatter from buried, half buried and proud objects.[26,29]

Figure 4(a) shows the bandlimited (1 to 12 kHz) impulse frequency response of the pressure measured 50 m from the source and 30 cm above the bottom. Here, the measured frequency response is normalized by the free-field source-receiver frequency response measured in the NRL pool laboratory. This transfer function [Fig. 4(a)] is the bandlimited impulse frequency response of the water channel. We also show the temporal bandlimited impulse response of the water channel [Fig. 4(b)], where the impulse frequency response is inverse Fourier transformed to time and filtered using an FIR to remove out-of-band noise. This bandlimited impulse response has the temporal resolution to distinguish individual pulses for each ray path in the water column. The physics of this will be discussed in the next two sections, where the experimental results will be compared to two models. The first is a simple physics-based ray model of a homogeneous water column. The ray model provides direct insight into the nature of this bandlimited water channel response. A second, more sophisticated, parabolic equation model is also compared to these measurements.

B. Ray model comparison

The first model to be compared to the experimental data is based on the simple concept of ray paths. The first three paths from the source to the receiver are shown in Fig. 3(b).

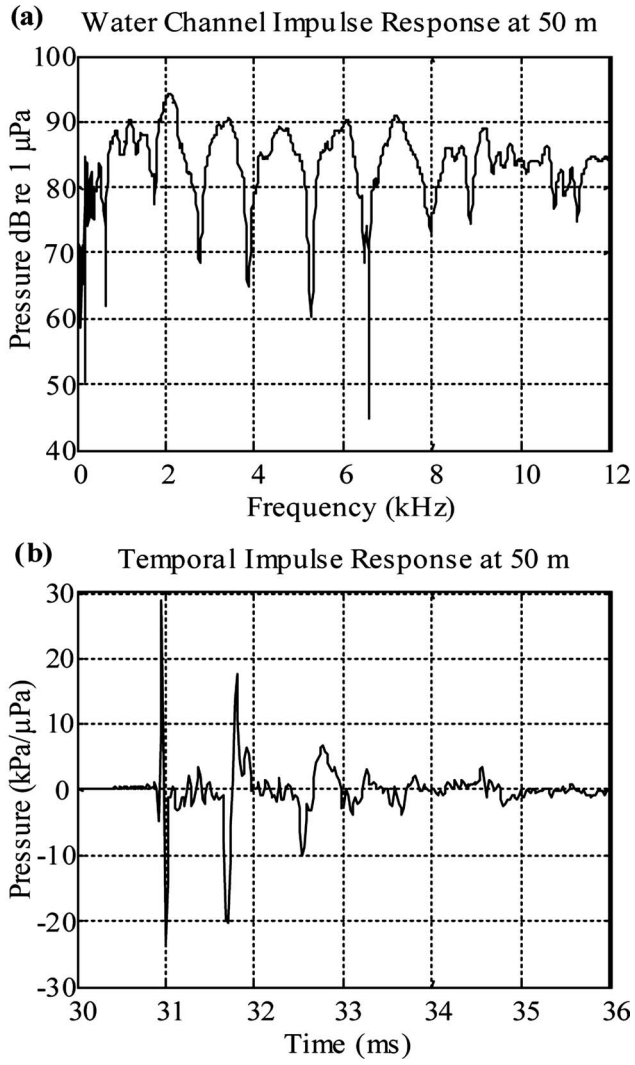


FIG. 4. (a) shows the water channel frequency response for a bandlimited impulse 50-m horizontal range from the source and 30 cm above the bottom. (b) shows an FIR filtered bandlimited temporal impulse response for the same 50-m source-receiver location.

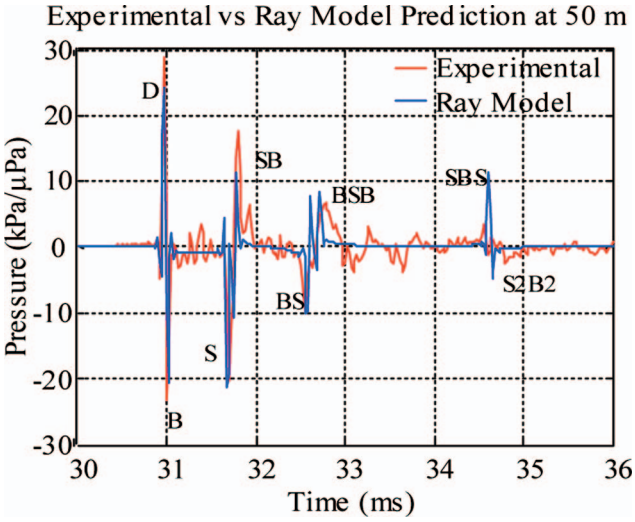


FIG. 5. (Color online) A comparison between the measured temporal impulse response and a ray model prediction of the same 50-m source-receiver location discussed in Fig. 4. The red line is the measured temporal response and the blue line is the ray model prediction.

The path lengths to the receiver increase with each bounce off an interface; thus, the temporal pulses associated with each path should arrive at different times. In addition there is an associated phase shift with each reflection off the air-water interface, and a similar phase shift for reflections off the sand-water interface that is calculated using Brekhovskikh.[30] Spherical spreading is assumed for each ray path and the amplitudes for each path are adjusted for the total distance to the receiver. Both the sandy bottom and the air-water interface are assumed to be flat for the calculations.

Figure 5 shows the same temporal response shown in Fig. 4(b) with an overlay of the superposition of impulses generated from the ray model. In Fig. 5, the first time arrival, just before 31 ms and designated with a *D* (direct), is associated with the direct path between the source and the receiver, shown in Fig. 3(b). The next time arrival, *B* (bottom), is associated with the path that includes a single bounce off the bottom, shown in Fig. 3(b). Here, the phase is nearly 180° from the direct path due to the subcritical reflection off the sand-water interface.[30] The next ray path, *S* (surface), is the single reflection off the air-water interface, shown in Fig. 3(b). There is also an 180° phase shift due to the pressure release boundary conditions of the air-water interface. In Fig. 5, the air-water bounce is then immediately followed by an additional bounce off the sand-water interface back to the receiver, *SB* (surface-bottom). Since this measurement point is only 30 cm above the bottom, there is very little temporal separation between these two paths; however, there is a phase shift due to the additional interaction with the bottom.

The next two time arrivals, between 32 and 33 ms, are associated with multiple bounces. The fifth arrival, *BS* (bottom-surface), corresponds to one bounce off the bottom and one bounce off the air-water surface before reaching the receiver. This is immediately followed by a second arrival from a bounce off the sand-water interface back to the receiver, *BSB* (or *B*2*S*1). The last time arrivals modeled in Fig. 5, between 34 and 35 ms, are for a ray path that reflects off the surface, then off the bottom, then off the surface and then to the array, *SBS* (or *S*2*B*1), and then the next ray that reflects off the bottom, *SBSB*, which is designated in Fig. 5 as *S*2*B*2 for two surface and two bottom interactions. This simple ray model, shown in Fig. 5, predicts the first eight ray paths. The data indicate that subsequent paths with more interface interactions result in very little additional energy at the 50-m range. As the number of bounces from the two interfaces increases, the associated ray paths form a steeper angle with the bottom. As this angle surpasses the critical angle, part of the energy from the ray is refracted into the bottom.[25] These ray paths with bottom insonification angles greater than the critical angle quickly attenuate in the water column due to this refracted energy. Also note that as the grazing angle increases, the phase shift due to the bottom interaction decreases. Thus, the first ray reflecting off the bottom, *B*, has a nearly 180°degree phase shift, as do the next two rays, *S* and *SB*. However, the *BS* and *BSB* rays have a grazing angle just below the critical angle that results in a phase shift due to the reflection off the bottom that is much

TABLE I. The physical parameters used for the ray model prediction of the 50-m data.

Physical parameter	Value
Water sound speed	1537 m/s
Water density	1026 kg/m^3
Bottom sound speed	1720 m/s
Bottom density	2070 kg/m^3
Bottom attenuation	5.46 dB/λ
Water column depth	6.75 m
Range	47.55 m
Source height	2.6 m
Hydrophone height	0.3 m

less that 180°. These relative phase shifts are accounted for in the ray model and agree well with the measured data.

This type of calculation allows the study of each ray path separately with a limited knowledge of environmental physical parameters. These parameters are shown in Table I. The numerical prediction was adjusted to match the experimental measurement by varying only two parameters, the source-to-receiver distance and the water depth. The source-to-receiver distance was varied until the direct pulse, *D*, of the numerical and experimental data matched. The water depth was then changed until the third pulse, *S*, resulting from a single reflection off the air-water interface, was aligned. Both parameters were within the error associated with the initial alignment of the system and the variation of water depth due to tides.

This simple model, which included the phase shift due to the reflections off either the bottom or the surface of the water channel, provides a good predictor of the acoustic pressure in the relatively stable environment of St. Andrews Bay. It has worked especially well for the relatively short range of these measurements, 20 m through 90 m, and predicted the amplitude and phase of the initial pulses within 2 dB. The difference between the ray model and the experiment can be accounted for with experimental error and suggests that the experimental error is within 2 dB. Also note that the later time arrivals, *S*2*B*2 in Fig. 5, have the largest differences between the ray model and the experiment. In the experiment, the air-water surface was moving due to wave actions, and the measurement shown in Fig. 5 is a coherent average of 36 pings. Although the water column was stable, within a few percent for the direct ray, *D*, the coherence from ping to ping degrades with each interface interaction, especially with the air-water interface. Thus, the assumption of a smooth flat surface and bottom becomes less valid for the later time arrivals, which are a result of more interface interactions. However, the model does provide great insight into the basic physics of the propagation in the water channel of St. Andrews Bay and validates the experimental approach being implemented with this experimental technique. This version of the ray model does not calculate the energy in the sandy bottom. To model the energy in the bottom, RAM will be used and the results will be discussed in the next section.

C. Range-dependent acoustic model comparison

To better understand the physics of this littoral environment, especially the acoustics at and below the sand-water interface, a comparison with a parabolic equation (PE)-based model will be made. The PE algorithm was developed by Collins[33] and is based on a split-step Padé solution.[31,32] The code is programed in FORTRAN and is named the range-dependent acoustic model (RAM). The details of the algorithm are referenced in the literature[31–37] and the inputs to the code used for these measurements are listed in Table II.

TABLE II. The physical parameters used for the RAM calculation.

Physical parameter	Value
Frequency range	0.5–10 by 0.050 kHz
Source depth	4.3 m
Receiver depth	4.6–9.2 m
Maximum range	50 or 90 m
Range step	0.75 m
Maximum water depth	24 m
Depth grid spacing	0.10 m
Reference sound speed	1537 m/s
Bathymetry range	0.0 m to 90 m
Bathymetry depth	6.9 m
Depth of profile points/ sound speed	1.0 m/1535.8 m/s
	2.0 m/1535.8 m/s
	3.0 m/1537.0 m/s
	4.0 m/1538.0 m/s
	5.0 m/1539.5 m/s
	5.8 m/1539.8 m/s
	6.1 m/1539.5 m/s
Sediment sound speed	1702 m/s
Sediment density	2035 kg/m^3
Sediment attenuation	Variable
	0.8 dB/λ at 1 kHz to 0.5 dB/λ at 10 kHz
Pade terms	8

The 50-m data are modeled first, and shown in Figs. 6(b) and 7(b). The 90-m data were also modeled and the results are shown in Figs. 6(d) and 7(d). There are several advantages to the RAM prediction over the ray model predictions. These advantages include modeling both the water column and the sandy bottom, incorporation of sound-speed variations in the water column, and the capacity to model inhomogeneties in the bottom and along the sand-water interface. The only drawback of RAM is the larger amount of computation time required for each data set. However, for the assumptions used for these calculations, of a smooth air-water surface and a smooth and level sandy bottom, the computational times are not significant. The comparison of the RAM prediction to the experimental data is shown in Figs. 6 and 7.

Figure 6 shows the phase of the pressure with respect to frequency and depth using a logarithmic color scale for the amplitude. These figures are made by plotting the sign of the real part of the data multiplied by $20\log_{10}$ of the normalized magnitude of the data. This plot provides a large dynamic range, 30 dB, while still giving phase information. The data displayed in Fig. 6 show that the wavefronts in the water column are nominally a superposition of a few waves, from slightly different insonification angles. The direct ray from the source to the receiver is dominant and is followed closely by the first bounce off the bottom [Fig. 3(b)]. However, the angles for both these rays are less than 3°, and thus appear

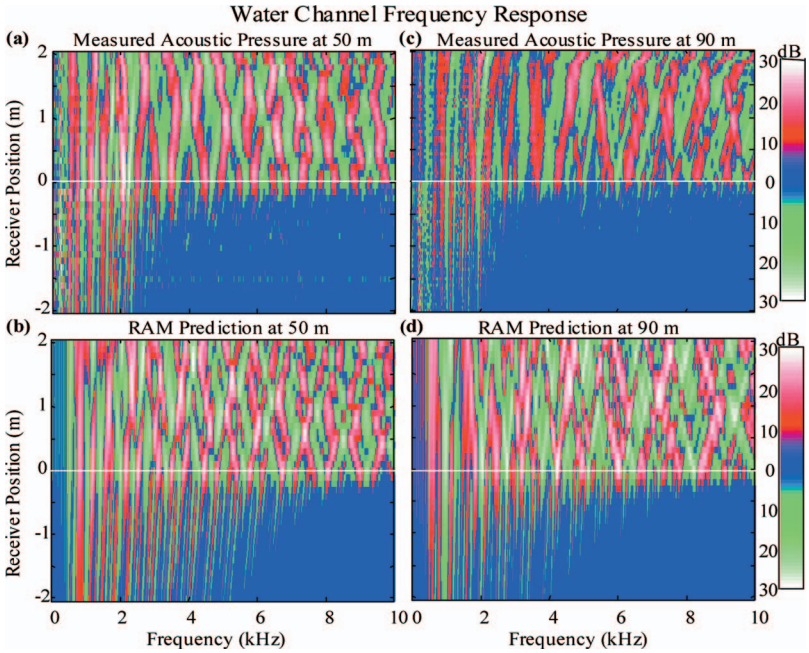


FIG. 6. (Color online) Logarithmic phase plots of the measured data and RAM predictions for both 50 and 90-m measurement locations. (a) is the logarithmic phase plot for the measured acoustic pressure 2 m above and below the bottom and 50 m from the source. (b) is the RAM prediction of the pressure at the 50-m location. (c) is the logarithmic phase plot for the measured acoustic pressure 90 m from the source and 2 m above and below the sand-water interface. (d) is the RAM prediction of the pressure at the 90-m location.

planar over the 2-m aperture of the array at the lower frequencies, below 2 kHz. As the frequency increases, the slight angles of these first two rays along with the energy from the first air-water interface and possibly the next few ray paths manifest themselves in an interference pattern seen in the phase above 2 kHz.

This interference pattern of the first few ray paths in the water column can be contrasted with the phase seen in the bottom. The phase fronts in the bottom are nearly vertical below 3 kHz, propagating perpendicular to the sand-water interface, and have more of a slope above 3 kHz, especially in the RAM predictions; however, the phase displays none of the interference pattern associated with multiple ray paths having nearly equal amplitudes. For the energy in the bottom below 3 kHz, the vertical phase fronts indicate that the region is dominated by evanescent waves. For the energy above 3 kHz, the slanted wavefronts indicate the traveling waves are dominant in the bottom. Also note that for this measurement, there is a sharp transition of the measured energy at 3 kHz in the bottom while the RAM prediction has a smoother decrease of energy penetrating into the bottom as the frequency increases. Both the measurements and the predictions indicate deeper penetration of energy into the bottom at lower frequencies versus higher frequencies, with 3 kHz being the nominal upper frequency for energy penetration into the bottom greater than 0.5 m. This is important for insonification of nearly flush-buried objects on the size of meters. The frequency analysis is remarkably similar for both the 50 and the 90-m sites. This apparent transition of

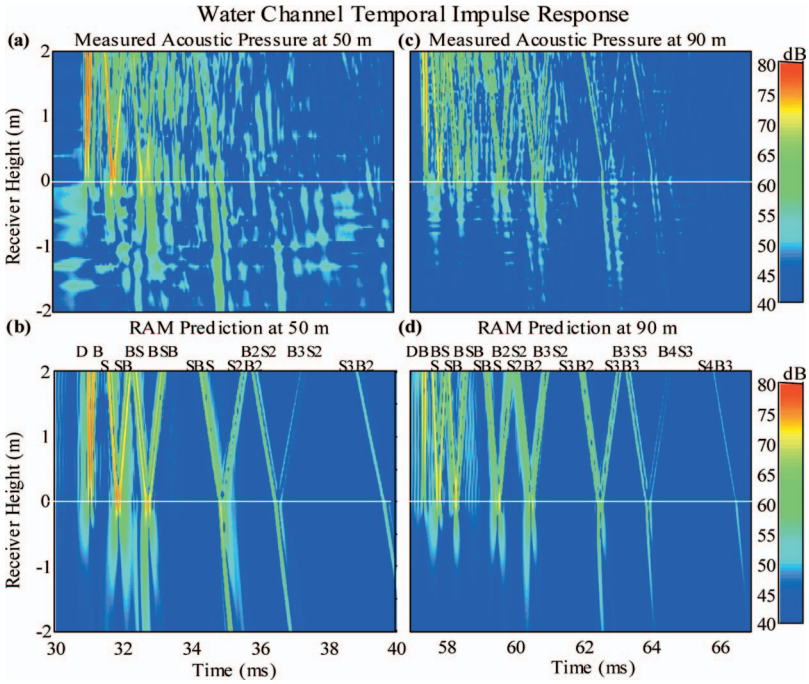


FIG. 7. (Color online) The temporal plots of the measured data and RAM prediction for both 50 and 90-m horizontal range from the source. (a) is the time plot for the measured acoustic pressure 50 m from the source and 2 m above and below the bottom. (b) is the RAM prediction of the pressure at the 50-m location. (c) is the time plot for the measured acoustic pressure along 90 m from the source. (d) is the RAM prediction of the pressure at the 90-m location.

energy in the bottom around 3 kHz will be discussed in more detail after a brief discussion of the temporal response, both in the water column and in the bottom.

Temporal analysis of the same four data sets is shown in Fig. 7. The frequency data, shown in Fig. 6, are inverse Fourier transformed to the time domain and the temporal data are filtered with an FIR filter to remove out-of-band noise. Figure 7 shows the bandlimited temporal impulse response in the water channel and in the sandy bottom for the 50 and the 90-m sites. The RAM predictions show, from left to right, the direct incident wave, *D*, quickly followed by the first bottom bounce, *B*, at around 31 ms for the 50 m case [Figs. 7(a) and 7(b)] and around 57 ms for the 90-m case [Figs. 7(b) and 7(d)]. The first four ray paths are shown and labeled in Fig. 3(b). The next impulse is the air-water reflection, followed closely by the subsequent reflection off the bottom and back to the hydrophone, *S* and *SB*, respectively. The subsequent bounces, after 32 ms for the 50-m case and 58 ms for the 90-m case, include multiple bounces off both the air-water surface and the sandy bottom, *BS* and *BSB*, *SBS* and *S*2*B*2, *B*2*S*2 and *B*3*S*2, and *S*3*B*2 for Figs. 7(a) and 7(b) and additionally, *S*3*B*3, *B*3*S*3, *B*4*S*3, and *S*4*B*3 for the 90 -m case shown in Figs. 7(b) and 7(d).

Generally the RAM predictions agree well with the measured data at the 50 and 90-m sites. Direct and detailed comparison between the 50-m experiment and RAM prediction of the first three rays that interact with the bottom (*D*/*B*), (*S*/*SB*), and (*BS*/*BSB*), shown in Figs. 7(a) and 7(b), provide excellent agreement, as did the ray model in the water column. This agreement between the experiment and RAM extends into the bottom for these first three bottom interactions, and detailed analysis of the amplitudes of these rays show that the difference is less than 2 dB. The early time arrivals provide the bulk of the energy in the water column and thus dominate the later time arrivals, which have less energy, especially in the experimental data.

Direct comparison of the later time arrivals (*SBS*/*S*2*B*2, *B*2*S*2/*B*3*S*2, etc.), in the 50-m data, [Figs. 7(a) and 7(b)], show that the experimentally measured amplitudes are significantly smaller than the RAM prediction. This is also true in the ray model comparison shown in Fig. 5. There are two contributing factors to this result. First, the source is a piston source that is omnidirectional at 1 kHz, but becomes directional, with a ±20-degree beam pattern (3 dB down) at 10 kHz. The effect of this beam pattern is that less energy is put into the higher angle rays than into the direct rays. The later time arrival rays such as *SBS* in the 50-m data or *S*3*B*2 in the 90-m data have insonification angles greater than 25°, which suggests that the amplitude of these rays will be less due to the directionality of the source. The RAM prediction assumed a point source and thus overestimates the energy of the experiment for these higher order, later time arrival rays. Although the RAM model could be modified to include the source directionality, it is not the primary purpose of this paper to conduct a detailed numerical study of the experiment, but rather to understand the experimental results, for which RAM has been an invaluable tool.

The second factor contributing to the difference between the experiment and RAM prediction is a result of coherently averaging the 36 pings that are measured at each hydrophone location. The ping-to-ping temporal variability is less than 2% for these measurements. Since the majority of the energy reaching the hydrophone is in these early time arrivals, this is a valid way to analyze the data. However, the later time arrivals, *SBS* and later for the 50-m measurements and *S*3*B*2 and later for the 90-m measurements, have less energy than the earlier time arrivals and thus did not impact the initial temporal stability analysis conducted for this data. However, the air-water surface was not stationary during these experiments. Wave heights were estimated to be less than a 10 -cm rms for these measurements. This random rough surface would affect the coherence of the rays with each interaction of the air-water surface. Thus, these later time arrivals would have less coherence and the averaging of the 36 pings will reduce the amplitude as frequency and number of surface interaction increase. Both the incoherence and directionality combine to reduce the energy of each time arrival in the experiment as compared with the RAM model. It is easily seen in Fig. 5 and Figs. 7(a) and 7(b), for the 50-cm data, where the predicted amplitude of the *SBS* and *S*2*B*2 ray is substantially larger than the experimental data and in the 90-m data, Figs. 7(c) and 7(d), where the predicted *S*3*B*2 and later rays are considerably larger than the experimental data, both in the water column and in the sandy bottom.

This lower amplitude of the later time arrivals, *SBS* and later in the 50-m data, and *S*3*B*2 and later in the 90-m data, clearly manifests itself in the sandy-bottom frequency domain data, shown in Fig. 6. The sharp transition in the sandy bottom at 3 kHz for the experimental data is a result of the later time arrivals having lower amplitudes in the water column and thus less energy penetrates into the bottom than predicted by RAM.

To better understand and to better compare and contrast the models and the measurements, it is useful to do a detailed analysis of the predicted energy versus the measured energy in the sandy bottom.

D. Bottom analysis

Of interest is the amount of energy in the bottom associated with each bounce. Note that for the first three impulses that interact with the bottom in the 50-m data, shown in Figs. 7(a) and 7(b), the energy in the bottom extends down perpendicularly to the sand-water interface. Also note, for the first three acoustic bottom interactions of the 50-m data, the energy in the bottom increased for each subsequent interaction. Each bounce is associated with an increasing number of interface interactions, and each higher number interaction results in an increase in the incidence and reflection angle, measured from the interface. Brekhovskikh[30] predicts that the amplitude of the evanescent field increases with the incidence angle for below critical angle insonifications. Note that for the fourth interaction (*SBS*/*S*2*B*2), of the acoustic wave with the bottom, around 35 ms in the 50-m data [Figs. 7(a) and 7(b)], the energy in the bottom is no longer perpendicular to the sand-water interface. This suggests that the insonification angle at the sand-water interface is now above the critical angle and that energy is refracted through the inter-

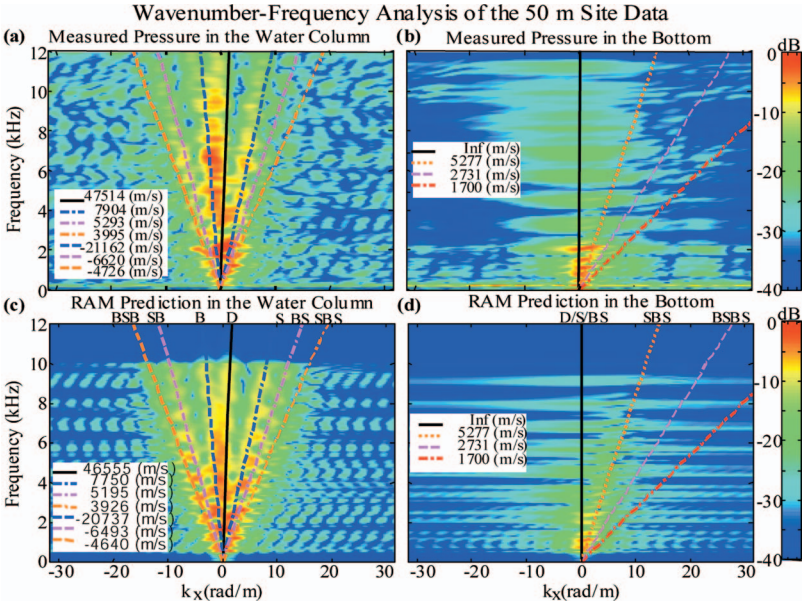


FIG. 8. (Color online) The measured and RAM predicted wave number–frequency response at 50-m horizontal range from the source. (a) (measured) and (c) (predicted) are the water channel band limited impulse responses. (b) (measured) and (d) (predicted) are the bottom bandlimited impulse responses. The positive wave numbers indicate energy moving down, or into the bottom. The negative wave numbers indicate energy moving up, or out from the bottom.

face creating a traveling wave in the bottom. The angle of incidence for each ray path is easily obtained using the ray model. The fourth and fifth bottom interactions [at 35 ms (*SBS*/*S*2*B*2) and 37 ms (*B*2*S*2/*B*3*S*2), respectively], have calculated incidence angles of 27° and 32°, respectively. The critical angle is 25.6° with a water speed of 1535 m/s and a bottom sound speed of 1702 m/s. From the RAM prediction of the 50-m data, Fig. 7(b), energy is refracted into the bottom for all acoustic interactions that arrive after 34 ms.

This same transition can be observed in the 90-m data, shown in Figs. 7(c) and 7(d). However, for the 90-m data the first five interactions with the bottom are subcritical, before 62 ms (*D*-*B*3*S*2), while the subsequent interactions, after 62 ms (*S*3*B*2-*S*4*B*3), are above critical angle interactions. Note that in both data sets, the energy remains trapped in the water column as long as the insonification angle with the bottom is below the critical angle. However, once the insonification is above the critical angle, the energy in the subsequent bounces reduces quickly as energy is refracted into the bottom.

Also note the differences between the predicted 90-m result and the measured 90-m results between 62 and 65 ms. The sixth and seventh bottom interactions (*S*3*B*2/*S*3*B*3 and *B*3*S*3/*B*4*S*3) predicted by the RAM model, Fig. 7(d), arrive sooner than the measured impulses. The experimental result shows that the sixth bounce (*S*3*B*2/*S*3*B*3) [Fig. 7(c)], arrives half a millisecond after the predicted arrival, Fig. 7(d), and using the slope (the impulse in the bottom lags in time as the depth increases) and attenuation (the amplitude does not decrease exponentially with depth) of the associated energy in the bottom, the insonification angle is above the critical angle. The seventh arrival (*B*3*S*3/*B*4*S*3), in the experiment [Fig. 7(c)], is also later than the predicted value [Fig. 7(d)], and again both the predicted and measured results suggest that the insonification angle is above the critical angle. The RAM prediction assumes a perfectly flat and level sand-water interface and air-water interface, neither of which existed at the measurement site. A slight increase in water depth over the 90-m range could account for the longer times seen in the experiment. However, the RAM prediction is still reasonable and gives us greater insight into the physics of these shallow-water environments.

Although the RAM prediction agrees well with the measured data, especially the strong temporal responses from the previously discussed ray paths in the water column, it does not predict some of the lower level energy observed in Figs. 7(a) and 7(c). The experimental and numerical data were processed and filtered identically, so the differences are related to physical mechanisms rather than data processing anomalies. These physical mechanisms include the wave action at the air-water interface, making the acoustical waves less coherent, as well as the rough bottom scattering of the acoustic wave with each subsequent surface interaction. Finally, velocity variability and inhomogeneities in the water column and the sandy bottom could also produce scattering. Thus, there is measured energy in the bottom [Figs. 7(a) and 7(c)] not specifically associated with a ray interacting with the bottom. This energy is not predicted by RAM [Figs. 7(b) and 7(d)]. Some of this additional measured energy is possibly due to scattering from interface roughness and inhomogeneity in the bottom. To better understand these experimental results, especially some of the lower level energy, wave number analysis will be conducted and discussed.

E. Wave number analysis

The measurements consist of equally spaced array elements that can be spatially Fourier transformed into the wave number–frequency domain. Figure 8 shows the wave number analysis for both the water column and buried array of the 50-m impulse frequency response data shown in Figs. 6(a) and 6(b). The horizontal axes represent wave numbers, in radians/meter, with the positive wave numbers representing energy moving down, toward the sand-water interface in the water column or away from the sand-water interface in the bottom. The negative wave numbers indicate energy moving up the array. The vertical axes represent frequency, in kHz.

Figures 8(a) and 8(c) represent the water-column array and Figs. 8(b) and 8(d) represent the buried array.

For the water column array [Figs. 8(a) and 8(c)], the trace velocities associated with several ray paths are indicated by lines overlaid on the figures. The black line, with the highest trace velocity, indicates the ray path, *D*, directly from the source to the center of the receiver array. The dashed blue line, *B*, just to the left of the black solid line in Figs. 8(a) and 8(c), is the first bottom bounce to the center of the array. The length of the array is 2 m, which limits the wave number resolution of this plot. However, at the 50-m site, enough wave number resolution exists to distinguish all the ray paths to the array. The ray path, *S*, from a single reflection off the air-water interface is indicated by the dot-dashed blue line just to the right of the black line in Figs. 8(a) and 8(b). The dashed peach line on the left indicates the trace velocity of the ray path, *SB*, associated with a single air-water reflection and a single bottom reflection. This is followed by the dot-dashed peach line on the right of Figs. 8(a) and 8(c) that is a ray path, *BS*, with first a bottom bounce then an air-water reflection. Finally, there is the dashed orange line on the left of Figs. 8(a) and 8(c) representing a ray path, *BSB*, that reflects off the bottom, off the air-water interface, and then off the bottom. The dot-dashed orange line on the right of Figs. 8(a) and 8(c) indicates the trace velocity of the ray path, *SBS*, that reflects of the surface, off the bottom, and then off the surface.

The 50-m water column measurement clearly shows several waves insonifying the array from different angles. The insonification angles associated for each line overlaid on Figs. 8(a) and 8(c) were calculated from the ray model. The trace velocity seen by the array is directly related to the angle of the wave on the array by $ct=cw/\cos(\theta)$, where $cw=1535$ m/s and θ is measured from the sand-water interface.

The comparison of Figs. 8(a) and 8(c) shows that the ray calculation matches the RAM prediction very well. Note that the RAM prediction was only calculated up to 10 kHz, where the measurement frequency range is 1 to 12 kHz. Both the measured data and the RAM prediction show that the vast majority of the energy in the water column, at 50 m from the source, is in the first few rays propagating through the water channel. The higher order paths associated with more interface reflections contribute a small percentage of the total energy.

Of equal interest to the energy in the water column is the energy measured in the bottom. Recall that the first three bottom bounces (D,S,BS), for the 50-m site, Figs. 7(a) and 7(b), are subcritical angle insonifications, with the subsequent interactions having above critical angle insonification. Figure 8(b) is the experimental wave number–frequency response in the bottom and Fig. 8(d) is the RAM prediction in the bottom. Note that evanescent energy should manifest along the $k=0$ region of the plot, in contrast to traveling waves that will have an associated trace velocity. Figure 8(b) shows that the majority of energy, below 3 kHz, measured in the bottom is along the $k=0$ line, as indicated by the black solid line. However, the RAM predicts less energy along the $k=0$ line below 3 kHz.

In addition to this $k=0$ energy, especially above 3 kHz, is energy that has an associated trace velocity, as indicated by the three dashed lines to the right of the solid black lines in Figs. 8(b) and 8(d). The dashed orange line is associated with the trace velocities of the fourth ray, *SBS*, that interacts with the bottom between 34 and 36 ms, shown in Figs. 7(a) and 7(b). The RAM prediction clearly shows that the majority of energy in the bottom above 3 kHz is associated with this above-critical-angle insonification. However, the 50-m measurement shows energy ranging from this above-critical-angle insonification to a trace velocity of 1700 m/s. As discussed by Simpson *et al.*,[24,25] this energy is consistent with scattering from the roughened interface into the bottom. The possibility of a slower compressional wave will be discussed briefly in the next section. Note that this lower trace velocity energy aliases around 9 kHz, due to the 10-cm array spacing, and results in the appearance of energy in the upper left of Fig. 8(b). With no energy propagating up the array, this aliasing is distinct and should not be confused as upward-propagating energy.

Also note that the dashed peach line is associated with the trace velocity of the fifth ray, *BSBS*, which also insonifies the sand-water interface above the critical angle [seen in Figs. 7(a) and 7(b)], and refracts into the bottom. The energy of this later refraction is very small with respect to the more dominant energy of the earlier refracted energy. The temporal plots of Figs. 7(a) and 7(b) can be windowed to display only the energy around 35 or 37 ms. This processing technique shows that the dotted peach line overlays this later time arrival very well.

The 1700-m/s line is included to demarcate the energy of a 1700-m/s compressional wave that could enter the bottom through scattering mechanisms. A compressional wave speed of 1700 m/s represents the lowest trace velocity that an array could measure for energy propagating along the array. Any energy propagating at any other angle to the array, with a 1700-m/s velocity, would have a higher trace velocity. The energy seen in Fig. 8(b) above the 1700-m/s line and to the right of the $k=0$ line is consistent with energy refracted into the bottom from various ray paths in the water channel and with energy scattered into the bottom from the roughened interfaces. Note that in Fig. 8, higher energy regions result in a broad horizontal strip of energy across the wave number spectrum. This is a result of the boxcar-like window that was used for the spatial processing of the data. Much of this horizontal stripping could be reduced by applying a spatial window to the data. However, most of the energy of interest is next to the sand-water interface, which precludes the use of spatial windows.

To further understand the dynamics of the littoral waveguide, wave number analysis for the 30 and the 90-m measurement data are presented. Figure 9 shows wave number analysis of the 30-m measurement, with Fig. 9(a) showing the wave number–frequency analysis in the water channel and Fig. 9(b) showing the wave number–frequency analysis in the bottom. Since the 30-m measurement is closer to the source than the 50-m data, the angles of the various ray paths to the measurement sites should be greater. Thus, the separation of the ray paths in the wave number–frequency domain provides for better resolution of each path.

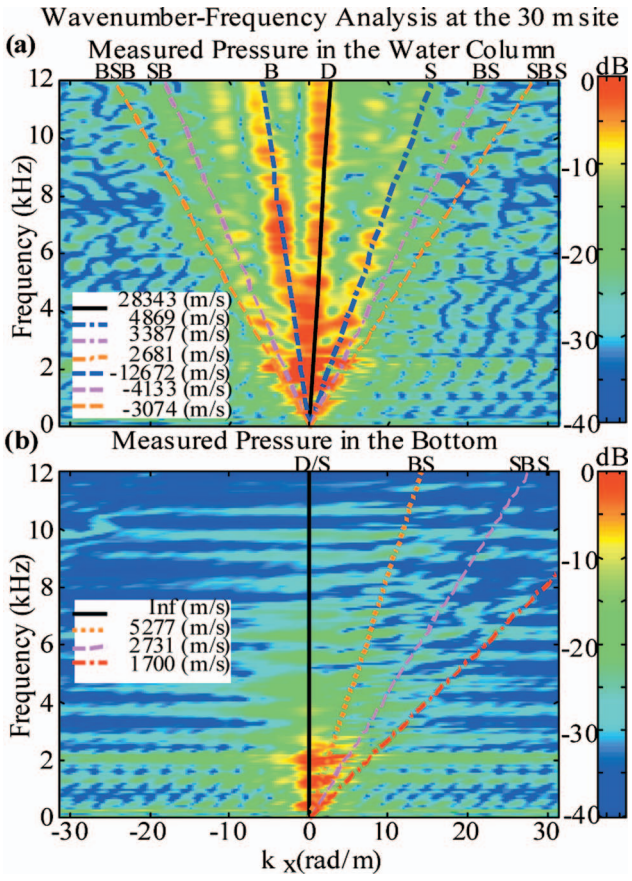


FIG. 9. (Color online) The measured wave number–frequency response at 30 m horizontal from the source. (a) is the water channel bandlimited impulse response and (b) is the sandy-bottom bandlimited impulse response.

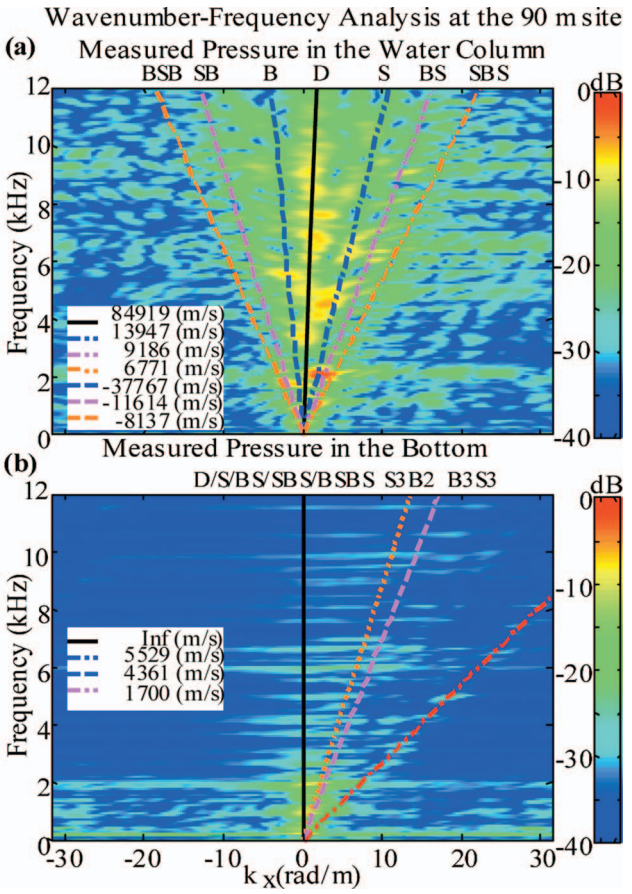


FIG. 10. (Color online) The measured wave number–frequency response at 90-m horizontal range from the source. (a) is the water channel bandlimited impulse response and (b) is the sandy-bottom bandlimited impulse response.

Overlaid on Figs. 9(a) and 9(b) are several trace velocities that are predicted using the ray model and are associated with different paths to the receiver array. These overlays have a different trace velocity than the 50-m data, since they have different incident angles for each ray. The predicted trace velocities in the water column [Fig. 9(a)] agree with the measurement. The first seven ray paths discussed in the 50 -m data are calculated for the new geometry of the 30-m site and overlaid on Fig. 9(a). Figure 9(b) is the bottom measurement at 30 m and again shows the majority of energy in the bottom is along the $k=0$ line and is below 3 kHz. Also, like the 50-m data, there is refractive energy entering the bottom. Unlike the 50-m data, the ray model predicts that only the first two bottom interactions, *D* and *S*, are below the critical angle. Thus, the third and the fourth bottom interactions, *BS* and *SBS*, refract energy in the bottom and are indicated by the orange and peach dashed lines in Fig. 9(b). Also, note the 1700-m/s line is overlaid to demarcate the lowest possible trace velocity from a 1700-m/s compressional wave propagating in the bottom, moving parallel to the array. Similar to the 50-m bottom data, the 30-m bottom data show most of the $k=0$ evanescent energy is below 3 kHz. More energy is seen above 3 kHz along the $k=0$ line than in the 50-m bottom data [Fig. 8(a)]. Also, the first two refracted ray paths agree with the predicted trace velocities in the bottom. There is evidence of other energy in the bottom between the $k=0$ line and the 1700-m/s line. Again, this is consistent with scattering of energy into the bottom from the roughened interfaces. However, there is little evidence of energy below the 1700-m/s line, even with the smearing of the energy from the boxcar-like spatial window used for the wave number space processing.

Figure 10 is the wave number–frequency analysis for the 90-m measurement; the measurement data are also shown in Figs. 6(c), 6(d), 7(c), and 7(d). Figure 10(a) shows the water-column response and overlaid are the same first few ray paths indicated in the 50 and 30-m data. The angular separation of the rays has decreased due to the increased distance. The overlays from the ray model prediction are consistent with the measured data, although the wave number resolution is marginal for clear delineation of each ray. Figure 10(b) is the wave number–frequency analysis of the energy in the bottom. Even at 90 m the majority of the energy in the bottom is along the $k=0$ line and is below 3 kHz. This is consistent with the previous analysis of the 90-m data and the other measurement sites. Energy is also seen between the $k=0$ line and the 1700-m/s line, with little energy seen below the 1700-m/s line. Most of the horizontal strips shown in the wave number plots are again associated with the boxcar window that was used when processing the array data into the wave number domain. This spreading in wave number space is well understood and has been addressed by the authors previously.[24,25]

F. Other compressional waves

Previous discussions by Simpson *et al.*[24,25] have addressed the subject of other compressional waves propagating in these types of sandy bottoms. These "slower" compressional waves were proposed by Chotiros,[28] who suggested that a Biot model of similar sandy bottoms can predict a 1200-m/s compressional wave. We will briefly discuss how a 1200-m/s slower compressional wave would manifest itself.

For the measurements reported here, the wave number plots do not clearly delineate the difference between a 1200-m/s compressional wave in the bottom and energy scattered into the bottom due to interface roughness. For the 50-m wave number plots shown in Fig. 8(b), the trace velocities across a vertical synthetic array would manifest in different lines associated with the different insonification angles. For the first five insonification angles associated with the individual bounces off the bottom, shown in Fig. 7(a), the trace velocities for a 1200-m/s wave would range from 1920 m/s for the 2.98° direct pulse, *D*, insonification to 1614 m/s for a 31.13° fifth bottom insonification shown in Fig. 7(a). This would result in energy both above and below the 1700-m/s line shown in Fig. 8(b). However, Fig. 8(b) shows a clear delineation between the higher energy levels seen above 2 kHz and above the 1700-m/s line and the lower energy seen below the 1700-m/s line but above 2 kHz. As was discussed by Simpson *et al.*,[25] some of the energy above the 1700-m/s line is associated with scattering from the roughened interface. As was discussed earlier, some of the energy seen in Fig. 8(b) is associated with the refraction into the fast compressional wave (1700 m/s), for insonification angles above the critical angle associated with the later time arrivals (*SBS* and later).

Although the evidence for a slower compressional wave is obscured in the wave number domain by energy from other scattering mechanisms, strong evidence for the existence of a slower 1200-m/s wave should be visible in the temporal data shown in Figs. 7(a) and 7(c). Note that for the later time arrivals (*SBS* and later), after 34 ms for the 50-m data and (*S*3*B*2 and later) 62 ms for the 90-m data, the insonification angles are greater than the critical angle and acoustic energy is refracted into the bottom. A slower compressional wave of 1200 m/s would not have a critical angle, so refractive energy would always enter the bottom for any insonification angle. In the 50-m data, the first temporal signal is associated with the direct energy from the source to the receiver. The insonification angle is 2.97°. Thus, there should be a refracted ray entering the bottom at ~39°. This refracted ray should manifest itself temporally in the same manner as the refracted energy (*SBS*), seen after 34 ms in the 50-m data. For a 1200-m/s compressional wave, a slanted line of energy should extend from this direct pulse, *D*, at 31 ms at the sand-water interface, downward 2 m below the sand-water interface and about 1.5 to 2 ms later. However, no refracted energy is observed for this first bottom interaction, *D*, in the 50-m data.

Clearer evidence of no refractive energy for below critical angle insonifications is seen in the temporal 90-m data [Fig. 7(c)]. For these 90-m temporal data, there is no energy that extends the full 2 m into the bottom for the first four bottom interactions. Later time arrivals (*B*2*S*2 and later), after 60 ms, have energy that extends the full 2 m into the bottom. The acoustic energy in the bottom [Fig. 7(c)] is predicted by RAM [Fig. 7(d)]; the prediction only includes 1700-m/s compressional waves and evanescent waves in the bottom. For the first four bottom interactions, a refracted 1200-m/s wave should clearly be seen since the evanescent field is small in the earliest time arrivals of the 90-m data. However, there is no refractive energy observed in the 90-m data before 62 ms, and this suggests that there is no slower compressional wave propagating in this sandy bottom. The 90-m site serendipitously had very little roughened interface scattering into the bottom, presumably since the bottom was locally fairly smooth. This makes the 90-m data set ideal for investigation of other wave types in this bottom. This null result for evidence of a slower compressional wave is consistent with other work reported by Simpson *et al.*[25] for this measurement site and this water-saturated sandy bottom.

V. CONCLUSIONS

A synthetic vertical hydrophone array has been successfully used to measure the acoustic pressure both above and below a sand-water interface in a shallow-water environment. The system has been designed to minimally impact the local area around the hydrophone array with the source located from 20 to 90 m away. The experiments have quantified the acoustic pressure levels in the 1 to 12-kHz frequency band both above and below the sand-water interface.

The measurements were conducted in St. Andrews Bay, where the environment was relatively stable and the environmental parameters, including the sound-speed profile in both the water column and sand, were easily measured and changed slowly over the time of the measurements. This stable environment resulted in a good comparison between the experimental data and two numerical models. A simple ray model provided a tangible understanding of the measurements and showed agreement, both in amplitude and phase. The comparison of the experimental data with a RAM prediction in the water column was good. However, RAM overestimated the energy in the later time arrivals, the rays at or above the critical angle. These rays have more surface interactions, which make them less coherent ping to ping, and the source directivity above 3 kHz results in the averaged measured energy being less than predicted by RAM. The lower measured amplitudes of the later time arrivals in the water column, the rays that interacted with the interfaces several times and were near the critical angle, resulted in less energy penetrating into the sandy bottom. However, there was additional measured acoustic energy in the bottom, probably due to scattering of energy from the sand-water and air-water interface, that was not predicted by the numerical model, RAM, which assumed a smooth air-water and sand-water interface. The synthetic array technique has been demonstrated successfully for quantifying the acoustic pressure both above and below the sand-water interface. Although this measurement system was utilized to characterize the acoustic

pressure at relatively close ranges, 20 to 90 m, the measurement system should be applicable for much longer ranges and for other sandy-bottom environments. The measurement system also has the advantage of being able to measure the sound-pressure levels above and below any type of bottom where a proud or buried target could be located without having to know all the environmental parameters and conditions of the local area. This is especially important for long-range shallow-water propagation, where measuring the environmental parameters can become daunting.

For the sandy bottom measured in St. Andrews Bay, these short-range propagation measurements provided great insight into the energy propagating in the water column and the mechanisms for the energy in the water column to penetrate into the sandy bottom. For the five measurement locations taken during this experimental exercise, no slower compressional waves were measured in this sandy bottom. Future investigations in other types of sandy bottoms will have to be conducted to explore the existence of a slower compressional wave. This newly developed procedure should provide very accurate measurements of the energy penetrating into the bottom with a minimum of disturbance to the natural bottom.

ACKNOWLEDGMENTS

This work was supported by the Office of Naval Research. The authors wish to thank Dr. Timothy Yoder for the use of his ray model and Dr. Michael Collins for the use of his PE model.

[1]E. I. Thorsos, K. L. Williams, N. P. Chotiros, J. T. Christoff, K. W. Commander, C. F. Greenlaw, D. V. Holliday, D. R. Jackson, J. L. Lopes, D. E. McGehee, J. E. Piper, M. D. Richardson, and D. Tang, "An overview of SAX99: Acoustic measurements," IEEE J. Ocean. Eng. **26**, 4–25 (2001).
[2]S. Stanic, B. E. Eckstein, R. L. Williams, P. Fleischer, and K. B. Briggs, "A high-frequency shallow-water acoustic measurement system," IEEE J. Ocean. Eng. **13**, 4–25 (1998).
[3]R. R. Goodman, J. W. Caruthers, S. J. Stanic, and M. A. Wilson, "Observations of high-frequency sound propagation in shallow water with bubbles due to storm and surf," IEEE J. Ocean. Eng. **25**, 501–506 (2000).
[4]S. J. Stanic, R. R. Goodman, R. W. Meredith, and E. Kennedy, "Measurements of high-frequency shallow-water acoustic phase fluctuations," IEEE J. Ocean. Eng. **25**, 507–515 (2000).
[5]K. L. Williams, D. R. Jackson, E. I. Thorsos, D. Tang, and K. B. Briggs, "Acoustic backscattering experiments in a well characterized sand sediment: Data/model comparisons using sediment fluid and biot models," IEEE J. Ocean. Eng. **27**, 376–387 (2002).
[6]M. A. Wilson, R. W. Farwell, and S. Stanic, "High-frequency propagation statistics in a shallow water environment," IEEE J. Ocean. Eng. **13**, 261–265 (1988).
[7]K. B. Briggs, D. Tang, and K. L. Williams, "Characterization of interface roughness of rippled sand off Fort Walton Beach, Florida," IEEE J. Ocean. Eng. **27**, 505–514 (2002).
[8]P. Elisseeff and H. Schmidt, "Acoustic propagation through a low Mach number, stratified flow," J. Acoust. Soc. Am. **101**, 1936–194 (1997).
[9]F. Ingenito, "Measurements of mode attenuation coefficients in shallow water," J. Acoust. Soc. Am. **53**, 858–863 (1973).
[10]F. Ingenito and S. N. Wolf, "Acoustic propagation in shallow water overlying a consolidated bottom," J. Acoust. Soc. Am. **60**, 611–617 (1976).
[11]R. J. Cederberg, W. L. Siegmann, and W. M. Carey, "Influence of geoacoustic modeling on predictability of low-frequency propagation in range-dependent, shallow-water environments," J. Acoust. Soc. Am. **97**, 2754–2766 (1995).
[12]A. K. Rogers, T. Yamamoto, and W. Carey, "Experimental investigation of sediment effect on acoustic wave propagation in the shallow ocean," J. Acoust. Soc. Am. **93**, 1747–1761 (1993).
[13]R. J. Cederberg, W. L. Siegmann, M. J. Jacobson, and W. M. Carey, "Predictability of acoustic intensity and horizontal wave numbers in shallow water at low frequency using parabolic approximations," J. Acoust. Soc. Am. **94**, 1034–1043 (1993).
[14]P. Abbot, I. Dyer, B. Gomes, J. Fulford, J. Lynch, G. Gawarkiewicz, and D. Volak, "Effects of Korean littoral environment on acoustic propagation," IEEE J. Ocean. Eng. **26**, No. 2, 266–284 (2001).
[15]M. Badiey, Y. Mu, J. Lunch, J. Apel, and S. Wolf, "Temporal and azimuthal dependence of sound propagation in shallow water with internal waves," IEEE J. Ocean. Eng. **27**, No. 1, 117–129 (2002).
[16]A. O. Williams, Jr., "Some effects of velocity structure on low-frequency propagation in shallow water," J. Acoust. Soc. Am. **32**, 363–371 (1960).
[17]W. M. Carey, J. Doutt, R. B. Evans, and L. M. Dillman, "Shallow-water sound transmission measurements on the New Jersey continental shelf," IEEE J. Ocean. Eng., **20**, No. 4, 321–336 (1995).
[18]R. H. Ferris, "Comparison of measured and calculated normal-mode amplitude functions for acoustic waves in shallow water," J. Acoust. Soc. Am. **52**, 981–988 (1972).
[19]H. P. Bucker, "Normal-mode sound propagation in shallow water," J. Acoust. Soc. Am. **36**, 251–258 (1964).
[20]C. T. Tindle, K. M. Guthrie, G. E. J. Bold, M. D. Johns, D. Jones, K. O. Dixon, and T. D. Birdsall, "Measurements of the frequency dependence of normal modes," J. Acoust. Soc. Am. **64**, 1178–1185 (1978).
[21]S. J. Hughes, D. D. Ellis, D. M. F. Chapman, and P. R. Staal, "Low-frequency acoustic propagation loss in shallow water over hard-rock seabeds covered by a thin layer of elastic-solid sediment," J. Acoust. Soc. Am. **88**, 283–297 (1990).
[22]A. Maguer, E. Bovio, W. L. J. Fox, and H. Schmidt, "*In situ* estimation of sediment sound speed and critical angle," J. Acoust. Soc. Am. **108**, 987–996 (2000).
[23]A. Maguer, W. L. J. Fox, H. Schmidt, E. Pouliquen, and E. Bovio, "Mechanisms for subcritical penetration into a sandy bottom: Experimental and modeling results," J. Acoust. Soc. Am. **107**, 1215–1225 (2000).
[24]H. J. Simpson and B. H. Houston, "Synthetic array measurements of acoustical waves propagating into a water-saturated sandy bottom for a smoothed and a roughened interface," J. Acoust. Soc. Am. **107**, 2329–2337 (2000).
[25]H. J. Simpson, B. H. Houston, S. W. Liskey, P. A. Frank, A. R. Berdoz, L. A. Kraus, and C. K. Frederickson, "At-sea measurements of sound penetration into sediments using a buried vertical synthetic array," J. Acoust. Soc. Am. **114**, 1281–1290 (2003).
[26]H. J. Simpson and B. H. Houston, "Laboratory measurements of sound scattering from a buried sphere above and below the critical angle," J. Acoust. Soc. Am. **113**, 39–42 (2003).
[27]K. L. Williams, D. R. Jackson, E. I. Thorsos, D. Tang, and S. G. Schock, "Comparison of sound speed and attenuation measured in a sandy sediment to predictions based on the Biot theory of porous media," IEEE J. Ocean. Eng. **27**(3), 413–428 (2002).
[28]N. P. Chotiros, "Biot model of sound propagation in water-saturated sand," J. Acoust. Soc. Am. **97**, 199–214 (1995).
[29]A. Tesei, A. Maguer, W. L. J. Fox, R. Lim, and H. Schmidt, "Measurements and modeling of acoustic scattering from partially and completely buried spherical shells," J. Acoust. Soc. Am. **112**, 1817–1830 (2002).
[30]L. M. Brekhovskikh, *Waves in Layered Media*, 2nd ed. (Academic, New York, 1980), pp. 1–23.
[31]M. D. Collins, "A split-step Padé solution for parabolic equation method," J. Acoust. Soc. Am. **93**, 1736–1742 (1993).
[32]M. D. Collins, "Generalization of the split-step Padé solution," J. Acoust. Soc. Am. **96**, 382–385 (1993).
[33]M. D. Collins, *User's Guide for RAM versions 1.0 and 1.0p* (Naval Research Laboratory, Washington DC, 2002).
[34]M. A. Leontovich and V. A. Fock, "Solution of the problem of propagation of electromagnetic waves along the earth's surface by the method of parabolic equation," J. Exp. Theor. Phys. **16**, 557–573 (1946).
[35]V. A. Fock, *Electromagnetic Diffraction and Propagation Problems* (Pergamon, New York, 1965), pp. 213–234.
[36]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, New York, 1977).
[37]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.

Sea surface effect on shallow-water reverberation

Ji-Xun Zhou[a] and Xue-Zhen Zhang
School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332-0405 and Institute of Acoustics, the Chinese Academy of Sciences, Beijing 100080, China

Zhaohui Peng
The National Laboratory of Acoustics, Institute of Acoustics, the Chinese Academy of Sciences, Beijing 100080, China

James S. Martin
School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332-0405

(Received 1 June 2006; revised 6 October 2006; accepted 22 October 2006)

Wideband reverberation measurements were made at a fixed location in the East China Sea on 3 and 5 June 2001 using the same measurement system. Sound-speed profiles were similar during both measurements. Wind speed (W) and rms surface-wave height (σ) changed from 2.74 m/s and 0.10 m on 3 June to 7.45 m/s and 0.33 m on 5 June. Thus, these measurements offer an opportunity to evaluate sea-surface effects on reverberation vertical coherence (RVC), RVC-inverted bottom acoustic parameters, and reverberation level (RL) in shallow water. The two sets of RVC and RL data, in a frequency range of 100–2500 Hz, show differences that are the apparent effects of the surface roughness. With increasing sea state, the RVC increases and the RL decreases. The effective bottom losses, inverted from the RVC data, correspond to the variation of sea state. This additional loss gives a physical explanation of the characteristics of both the measured RVC and RL. These preliminary findings show the importance of surface effects in shallow-water reverberation and propagation models. These effects would be pronounced for high frequencies and sea states. For the conditions reported here, the effects were most apparent for $f > 500$ Hz when $W = 7.45$ m/s. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2400623]

PACS number(s): 43.30.Gv, 43.30.Hw, 43.30.Pc, 43.30.Re [SLB] Pages: 98–107

I. INTRODUCTION

It is well known that a rough sea surface may cause strong scattering and redistribute energy from the forward-propagating field into the scattered (reverberant) field.[1–3] The rough sea surface may also increase coherent reflection loss in the specular direction because of scattering in nonspecular directions,[4–6] which would be expected to reduce long-range reverberation. In shallow-water waveguides these two mechanisms simultaneously affect the reverberation process and compete for dominance. A natural question is, as the sea surface-wave height or wind speed increases, how will the reverberation vertical coherence (RVC) and reverberation level (RL) change? If shallow-water sound propagation measurements or reverberation data are used to invert the geoacoustic parameters of the seabottom, it is important to know the impact of the sea surface on the accuracy of the inversion. Under some conditions, it may be necessary to account for surface effects in the inversion process.

Wideband reverberation measurements were made repeatedly at a fixed location in the center of the ASIAEX (Asian Sea International Acoustic Experiment) test site in the East China Sea on 3 and 5 June 2001 using a measurement-system deployed from the Chinese research vessel (R/V) *Shi Yan 3*. Sound-speed profiles during the two measurement periods were similar. However, the wind speed and rms surface-wave height (σ) changed from about 2.74 m/s and 0.10 m on 3 June to 7.45 m/s and 0.33 m on 5 June. Any differences in the two reverberation measurements should be attributable to these surface effects.

Recent papers have reported measurement logs and major field efforts on the ASIAEX East China Sea Program,[7] including the RVC data analysis, RVC-based seabed geoacoustic inversion, and RL data analysis.[8,9] This paper focuses on analyzing the sea-surface effects on the RVC, RL, and RVC-inverted bottom acoustic parameters. The paper describes the setup of the experiment and related environmental parameters, presents measurements of RVC as a function of reverberation time (range) and frequency, and describes an RVC-based geoacoustic inversion technique and the resulting effective bottom acoustic parameters (speed and attenuation). The measured RVC was found to increase with increasing sea state. The RVC-inverted effective bottom loss indicates that high sea state corresponds with high combined bottom-surface loss. This increased loss offers an explanation (or a physical analog) for the sea-surface effects on RVC. The measured RL as a function of time and frequency was found to decrease with increasing sea state. This is consistent with the analogous physical situation suggested by the RVC analysis.

II. MEASUREMENT SETUP AND ENVIRONMENTAL PARAMETERS

The ASIAEX experiment was conducted in the East China Sea between 28 May and 9 June 2001.[7] As part of

[a]Electronic mail: jixun.zhou@me.gatech.edu

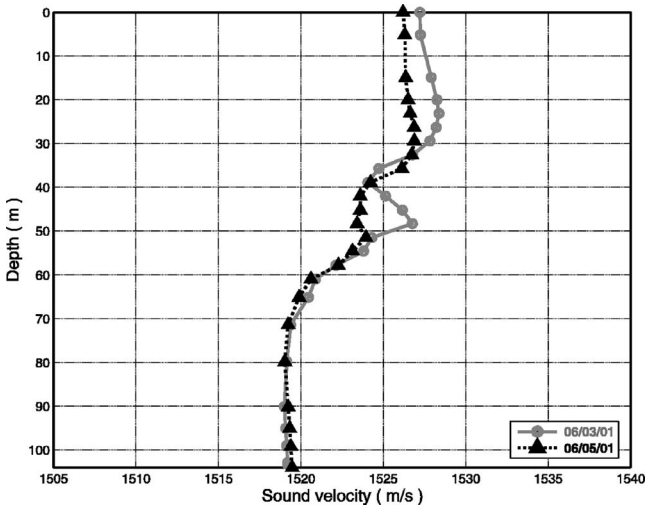


FIG. 1. Sound-speed profiles during two reverberation-measurement periods.

ASIAEX, wideband reverberation measurements were made repeatedly on 3 and 5 June using the same measurement system.[8,9] Reverberation data were collected from a 32-hydrophone vertical array. The sampling frequency for each hydrophone was 6000 Hz. The hydrophone array, which was made by the Institute of Acoustics of the Chinese Academy of Sciences (CAS), was suspended from the R/V *Shi Yan 3*, which was anchored at a location of 29°40.67′N and 126°49.39′E in the center of ASIAEX test area. The seabed was nominally flat and level, with an average water depth of 104 m over the extent of the reverberation measurements. The seabed at the ASIAEX site consists of very fine sand/sandy silt. The mean grain diameter over the entire ASIAEX area is 0.043 mm (4.54 Phi). Near the reverberation receiving array, it is 0.090 mm (3.47 Phi). Reverberation data were collected using both 38- and 1000-g explosive charges detonated at a nominal depth of 50 m. The horizontal separations between the source and the receiving array were 30.4 and 80.0 m, respectively on 3 and 5 June. Sound speed profiles during two reverberation measurement periods (09:13–09:21 on 3 June and 13:38–13:50 on 5 June, AWST-China) were similar. These profiles are shown in Fig. 1. Surface-wave data from a surface buoy were collected and analyzed by Dr. Peter Dahl of the Applied Physics Laboratory of University of Washington (APL-UW).[7] The wave heights were averaged over 30-min intervals during ASIAEX. These data indicate that σ changed from 0.10 m during the measurements on 3 June to 0.33 m during the measurements on 5 June. Figure 2 shows a 7-day history of the rms surface wave height (σ) and the wind speed (W), starting from 00:00 on 31 May (AWST-China). The wind speed was measured at a height of 14 m above the mean sea surface height. The rms wave height was directly calculated (integrated) from the measured frequency spectrum of the surface-wave field. The two time windows for reverberation data/model comparisons are marked with bars in Fig. 2. The two surface conditions that existed during the measurements can be seen to fit into slowly varying weather trends during ASIAEX.

III. SEA SURFACE EFFECTS ON RVC

In analyzing RVC, we are interested in a zero-delay value of the reverberation signal correlation between two hydrophones. Thus, the RVC between two time series from the ith and jth hydrophones with separation Δz_{ij} were obtained from

$$\rho_{Rvc}(t,\Delta z_{ij}) = \frac{\int_t^{t+\Delta t} p_i(\tau)p_j(\tau)\,d\tau}{[\int_t^{t+\Delta t} p_i^2(\tau)\,d\tau]^{1/2}[\int_t^{t+\Delta t} p_j^2(\tau)\,d\tau]^{1/2}} \approx \frac{\Sigma_{l=1}^{L} p_i(t+t_l)p_j(t+t_l)}{[\Sigma_{l=1}^{L}|p_i(t+t_l)|^2\Sigma_{l=1}^{L}|p_j(t+t_l)|^2]^{1/2}},$$

where ρ_{Rvc} is the cross-correlation coefficient, t is time, and $p(t)$ is the measured pressure. This expression was evaluated using MATLAB with Δt=341 ms, corresponding to 2048 samples of the measured signals.

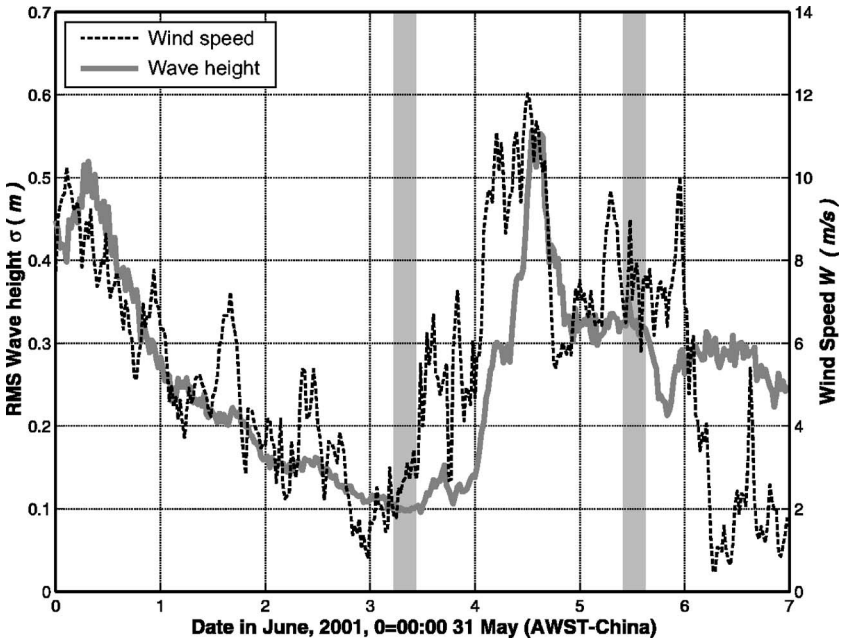


FIG. 2. Wind speed and surface wave height variabilities during the ASIAEX experiment.

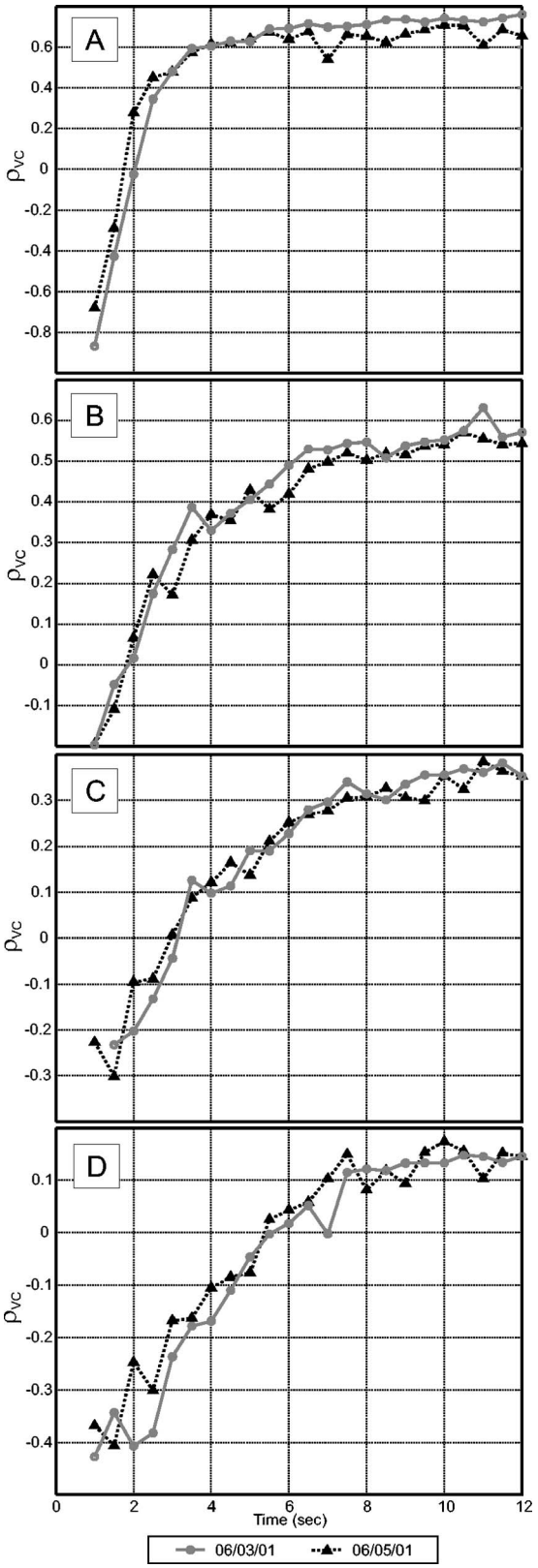


FIG. 3. RVC obtained on 3 June and on 5 June, Δz=4 m, (a) 200 Hz, (b) 300 Hz, (c) 400 Hz, and (d) 500 Hz.

Figures 3 and 4 show the reverberation vertical cross-correlation coefficient (ρ_{Rvc}) analyzed as a function of reverberation time and signal center frequency in $\frac{1}{3}$-oct bands. The circles in these figures represent the RVC data from the reverberation measurements on 3 June when W=2.74 m/s and σ=0.10 m. The triangles represent the RVC data from 5 June when W=7.45 m/s and σ=0.33 m. The separation between a pair of hydrophones (Δz) is 4 m for Fig. 3 and 1.7 m for Fig. 4. All the RVC data in these two figures are averaged values, obtained from three or four 1-kg wideband explosive sources and from several pairs of hydrophones. These pairs were located at depths between 56.5 and 90.5 m for the data depicted in Fig. 3 and between 23.7 and 30.5 m for the data depicted in Fig. 4. Different sensor pairs were used in Figs. 3 and 4, because the RVC data in these two figures are for different frequency ranges. For the RVC data at higher frequencies in Fig. 4, a smaller hydrophone separation was required to obtain enough dynamic range to show the RVC variation. In Fig. 4, the effect on the RVC of an echo arriving with a delay between 4 and 5 s can be seen between 700 and 1300 Hz. This was caused by the presence of an American vessel, the R/V *Melville*, which was about 2 miles away from the R/V *Shi Yan 3* during the measurements.

Figures 3 and 4 show that, at frequencies below about 500 Hz, the RVC coefficients have no significant dependence on the change in W from 2.74 to 7.45 m/s, or in σ from 0.10 to 0.33 m. At higher frequencies, the RVC increased with the increased sea state. This is a somewhat surprising phenomenon in that a rougher sea surface does not cause decoherence of shallow-water reverberation.

IV. RVC-BASED BOTTOM GEOACOUSTIC INVERSION

Sea-surface effects on shallow-water RVC can be explained by analogy to equivalent seabottom properties with calm surface conditions. For this it is useful to introduce a geoacoustic inversion technique for seabottom acoustic parameters that is based on RVC. More details of this inversion method may be found in a recent paper.[8]

An angular-spectrum method based on ray-mode analogies is used to derive an expression for the RVC. In this way, the normalized cross-correlation coefficient of shallow-water reverberation can be expressed by[8]

$$\rho_{Rvc}(\Delta z,r,z;h)=\left\{\sum_n \frac{e^{-2\beta_n r}\cos\left[k(z)\Delta z\,\sin\,\theta_n(z)\right]M[\theta_n(h)]}{S_n^2\tan\,\theta_n(z)k(h)\,\sin\,\theta_n(h)}\right\}\times\left\{\sum_n \frac{e^{-2\beta_n r}M[\theta_n(h)]}{S_n^2\tan\,\theta_n(z)k(h)\,\sin\,\theta_n(h)}\right\}^{-1} \quad (1)$$

or

$$\rho_{Rvc}(\Delta z,r,z;h)=\left\{\sum_n \overline{|\Phi_n(z)|^2}\,\overline{|\Phi_n(h)|^2}\cos\left[k(z)\Delta z\,\sin\,\theta_n(z)\right]\times M[\theta_n(h)]\exp\left(-2\beta_n r\right)/k_n\right\}\times\left\{\sum_n \overline{|\Phi_n(z)|^2}\,\overline{|\Phi_n(h)|^2}M[\theta_n(h)]\times\exp\left(-2\beta_n r\right)/k_n\right\}^{-1}. \quad (2)$$

Here Φ_n is the eigenfunction, k_n is the wave number, and β_n is the modal attenuation factor of the nth mode. Δz is the vertical separation of a pair of hydrophones located at depth z where the water depth is h. k is the acoustic wave number

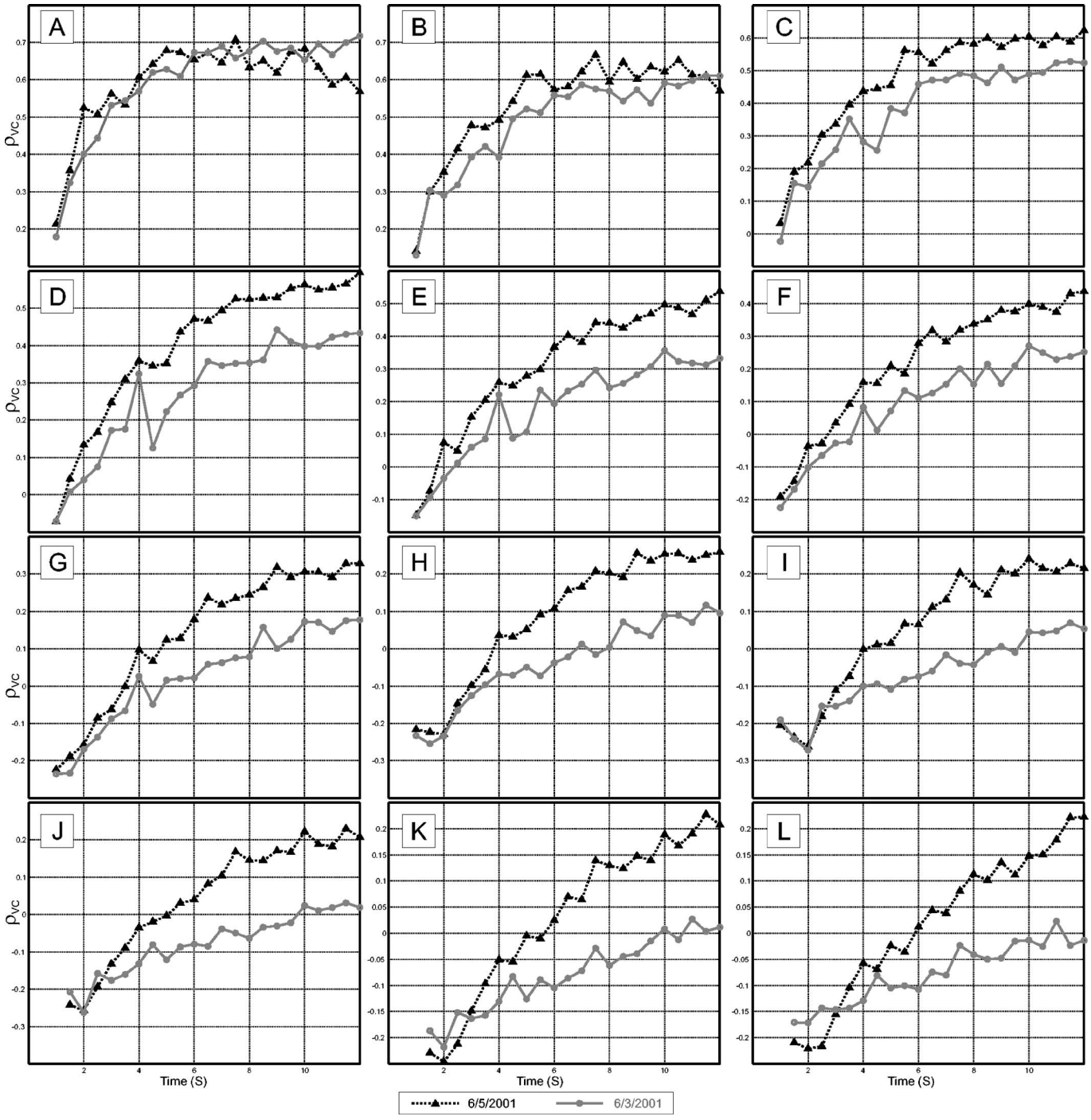


FIG. 4. RVC obtained on 3 June and on 5 June; the RVC increased with increased sea state. $\Delta z = 1.7$ m, (a) 500 Hz, (b) 600 Hz, (c) 700 Hz, (d) 800 Hz, (e) 900 Hz, (f) 1000 Hz, (g) 1100 Hz, (h) 1200 Hz, (i) 1300 Hz, (j) 1400 Hz, (k) 1500 Hz, and (l) 1600 Hz.

at the specified depth: $k(z)=\omega/c(z)$. $\theta_n(z)$ is the grazing angle of the nth mode-ray at the receiver depth z, $\sin\theta_n(z)=[1-k_n^2/k^2(z)]^{1/2}$. S_n is the cycle distance of the nth mode-ray: $S_n \approx 2\pi/(k_n - k_{(n+1)})$. $M[\theta_n(h)]$ is a bottom scattering function. $\overline{|\Phi_n(z)|^2}$ is the smoothed (slowly varying) energy depth distribution of the nth mode. These two expressions for the RVC can be calculated using a normal-mode numerical code, such as MOATL[10] or KRAKEN.[11] Sound speed and attenuation in the sea bottom at the ASIAEX site are determined by finding values that provide the best match between the measured and predicted RVC.

For the inversion, we assume that shallow-water reverberation is dominated by sea-bottom scattering and that the bottom scattering obeys a modified Lambert scattering law for bottom scattering in a shallow-water waveguide,[12]

$$M_b(\theta_n, \phi_m) = M(\theta_n)M(\phi_m) = \frac{\mu \sin\theta_n \sin\phi_m}{V(\theta_n)V(\phi_m)}. \tag{3}$$

Here $V(\theta_n)$ is bottom reflection coefficient, and μ is the Lambert scattering coefficient. For simplicity, we assume that the sediment scattering layer and the basement have uni-

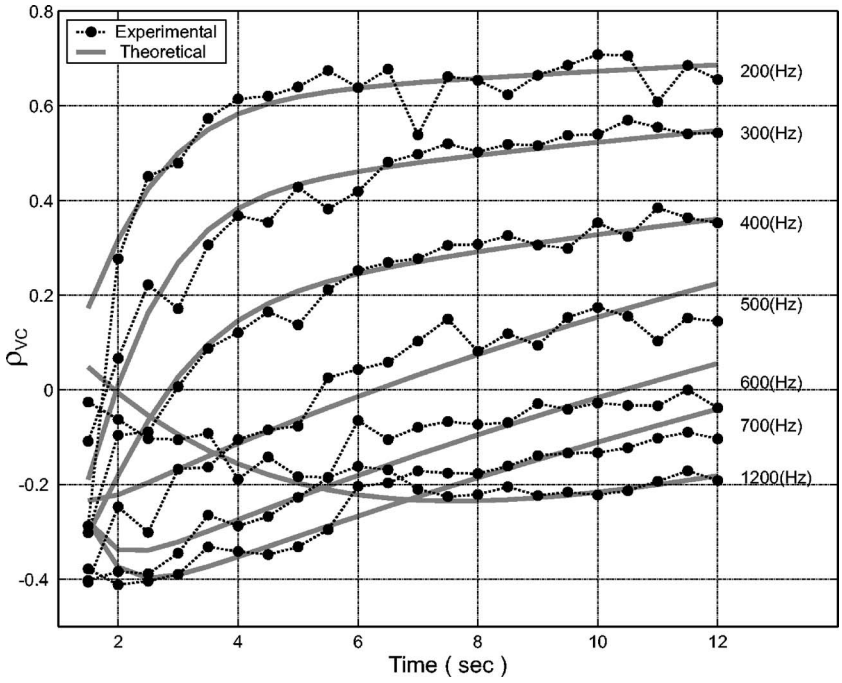


FIG. 5. The best match between the RVC data from 5 June and the model. Hydrophone separation=4 m.

form compressive sound speed and attenuation. The shear wave propagation in the sediment is assumed to be unimportant at experimental frequencies. The bottom sound speed and attenuation are the two unknown parameters that are inverted from the RVC data. Based on data from core samples, the bottom density was assumed to be uniformly 1.85 (g/cm^3). The average RVC data between 2 and 12 s are compared with theoretical predictions from Eq. (2) over all frequencies. Restricting the RVC data to times between 2 and 12 s lessens the effect of the bottom scattering at both very large and very small grazing angles, where the scattering might not be well approximated by Lambert's law.[13]

The theoretical RVC is numerically evaluated by using MOATL for bottom sound speeds between 1500 and 1900 m/s in intervals of 1 m/s with attenuation coefficients between 0 and $0.6f$ dB/m in intervals of $0.001f$ dB/m, where f is frequency in kHz. The pair of bottom sound-speed and attenuation values for which the numerical RVC curves best match the experimental RVC data is readily determined. The uncertainties associated with this inversion procedure have been discussed in a previous paper, which included the RVC data/model comparison for the data collected on 3 June.[8]

Representative results for the RVC data/model comparison using data collected on 5 June are shown in Fig. 5. The dots in these figures represent the measured reverberation vertical cross-correlation coefficients analyzed as a function of reverberation time and signal frequency f (for $\frac{1}{3}$-oct bands). The separation between pairs of hydrophones (Δz) for Fig. 5 is 4 m. The data are averaged values for four shots and eight pairs of hydrophones, located at depths from 56.5 to 90.5 m. Numerical results from Eq. (2), based on the best-matched sound speed and attenuation values, are shown by the solid lines in Fig. 5.

The pair of bottom sound-speed and attenuation values for which the numerical RVC curves best match the experimental RVC data in a least-error-squared sense are determined using a numerical search. RVC-inverted sound speed between 100 and 1600 Hz is plotted in Fig. 6 and equivalent attenuation in sediments between 100 and 1600 Hz is plotted in Fig. 7. Circles represent values inverted from the 3 June RVC data. Triangles represent values inverted from the 5 June RVC data. The bottom sound speed and effective attenuation for 100 and 200 Hz are inverted from the RVC data with a hydrophone separation of 8 m. For 300–1200 Hz, values are inverted from RVC data with hydrophone separations of 4 m. For 1300–1600 Hz, values are inverted from RVC data with a hydrophone separation of 1.7 m. The mean sound speed over 100–1600 Hz, inverted from the RVC measurements on 3 June, is 1614 m/s. On 5

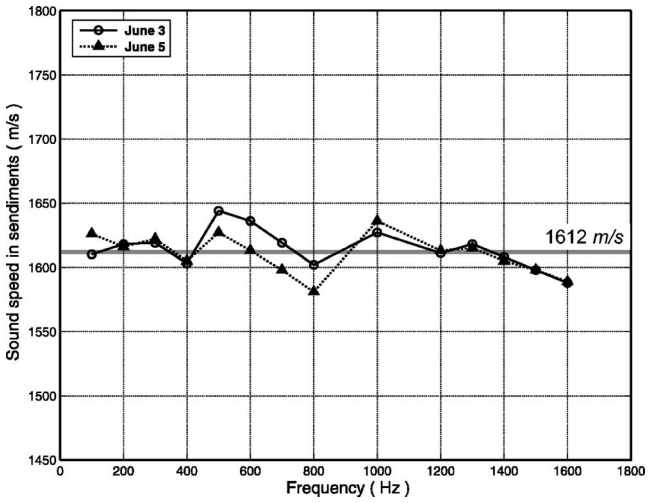


FIG. 6. Sound speed in the bottom at the ASIAEX site, inverted from two RVC measurements on 3 June and 5 June.

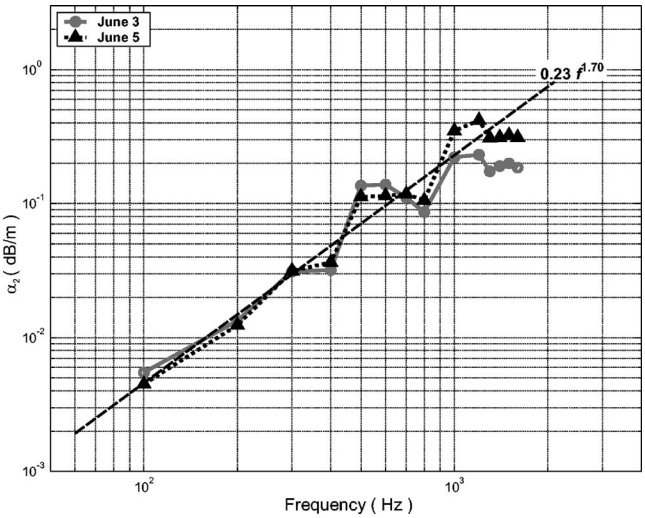


FIG. 7. Effective sound attenuation in the bottom at the ASIAEX site, inverted from two RVC measurements on 3 June and 5 June.

June it is 1610 m/s. The average of these two values (1612 m/s) is plotted in Fig. 6 as a solid gray line. Thus, the critical angle of seabed sound reflection, $\theta_c = \cos^{-1}(1519.2/1612.0)$, is 19.5°. The effective bottom attenuation, inverted from the RVC measurements, exhibits nonlinear frequency dependence in a frequency range of 100 to 1200 Hz. For the data measured on 3 June, the effective bottom attenuation can be approximated as $\alpha_2 = 0.20f^{1.57}$ dB/m, where f is frequency in kHz. For the RVC data measured on 5 June, it can be expressed as $\alpha_2 = 0.26f^{1.82}$ dB/m. Averaging the effective bottom attenuation over two days' measurements gives a frequency dependence of $\alpha_2 = 0.23f^{1.70}$ dB/m. This is plotted in Fig. 7 as a dashed line.

V. PHYSICS OF THE OBSERVED RVC CHARACTERISTICS

It is well known that the nth normal mode decay factor (β_n) in shallow water can be expressed by

$$\beta_n \approx -\frac{\ln|V_b(\theta_n)|}{S_n(\theta_n)}. \tag{4}$$

Here $V_b(\theta_n)$ is the effective bottom reflection coefficient when an equivalent ray of the nth mode is incident on the bottom with a grazing angle θ_n and S_n is the cycle distance of nth mode-ray.

The bottom reflection coefficient V_b at grazing angles that are smaller than the critical angle approximately satisfies the relationship[14,15]

$$-\ln|V_b(f,\theta)| \approx Q_b(f,\sigma)\theta \tag{5}$$

and the bottom reflection loss factor, Q_b, can be expressed by

$$Q_b(f) = \frac{0.0366(c_1^2/c_2)(\rho_2/\rho_1)\alpha_2}{[1-(c_1/c_2)^2]^{3/2}f}, \tag{6}$$

where c_2 and c_1 are sound speed in bottom and water in units of km/s, f is frequency in kHz, ρ_2 and ρ_1 are densities of the bottom and the water column, and α_2 is the sound attenuation in the bottom expressed in dB/(m·kHz).

Inputting the RVC-inverted bottom sound speed and equivalent attenuation shown in Figs. 6 and 7 into Eq. (6) gives the effective bottom loss factor as a function of frequency shown in Fig. 8. The effective bottom loss is not a characteristic of the seabottom but refers to a bottom condition that would reproduce the actual bottom loss and the additional loss that results from the coherent reflection loss of the rough surface. The losses from the RVC measurements on 3 June (Q_{E63}) are plotted as open circles. The data from the measurements on 5 June (Q_{E65}) are plotted as triangles. Figure 8 shows that the RVC-inverted effective bottom loss is higher for the data collected during the period of higher sea state. The RVC-inverted effective bottom loss represents the increased boundary reflection losses very well. This increased loss offers a physical explanation for the RVC characteristics reported in Sec. III. That is, with increasing sea state, the effective bottom loss becomes larger, and there is

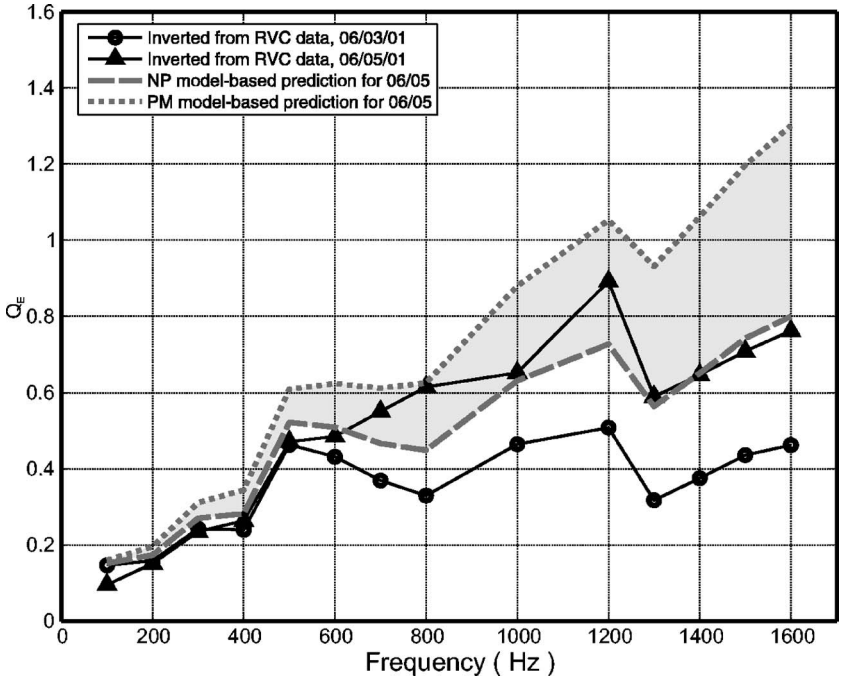


FIG. 8. Comparison of the RVC-inverted combined bottom-surface loss with the predictions, based on the NP and PM surface models.

faster mode stripping in the measured reverberation. Here, the effective grazing angle of reverberant energy arriving at the receiving array is smaller and there is an increase of the RVC.

As with any geoacoustic inversion method, the RVC-inversion method unavoidably involves uncertainties. A detailed discussion of these uncertainties is given in a previous paper.[8] For the current analysis, the same method and procedure were used for both RVC data sets. The uncertainties caused by the RVC-inversion method have no apparent effect on the systematic difference between the computed equivalent bottom losses.

The reflection coefficient of underwater sound from a flat sea surface is very close to −1. When the surface is rough, the magnitude of the reflection coefficient is less than unity because some acoustic energy goes into the diffusely scattered field. Based on Bass' perturbation theory, Brekhovskikh[4] and Brekhovskikh and Lysanov[5] derived the coherent mean reflection coefficient from an isotropic sea surface with large-scale roughness at a small grazing angle (θ) as follows:

$$|V_s(\theta)| = 1 - \frac{\sqrt{2}k^{3/2}}{\sqrt{\pi g}} \frac{\Gamma\left(\frac{3}{4}\right)}{\Gamma\left(\frac{5}{4}\right)} \sin\theta \int_0^\infty \Omega S(\Omega)\, d\Omega. \tag{7}$$

Here Γ is the Euler gamma function, g is the gravitational acceleration, and $k=2\pi f/c$ is the acoustic wave number. Ω is surface-wave frequency and $S(\Omega)$ is the frequency spectrum of the surface-wave field. For fully developed wind-driven seas, there are two widely used representations for surface wave spectra: the Neumann-Pierson (NP) spectrum and the Pierson-Moskowitz (PM) spectrum,[4,5] which are given by

$$S_{\mathrm{NP}}(\Omega) = \frac{2.4}{\Omega^6} \exp\left[-2\left(\frac{g}{\Omega W_{\mathrm{NP}}}\right)^2\right], \tag{8}$$

$$S_{\mathrm{PM}}(\Omega) = \frac{0.0081 g^2}{\Omega^5} \exp\left[-0.74\left(\frac{g}{\Omega W_{\mathrm{PM}}}\right)^4\right], \tag{9}$$

where W is the wind speed in m/s. The wind-speed corresponding to the NP spectrum (W_{NP}) is measured at the height of 5.5 m. The wind speed for the PM spectrum (W_{PM}) is measured at 19.5 m. The rms surface-wave height (σ) is defined as

$$\sigma = \left[\int_0^\infty S(\Omega)\, d\Omega\right]^{1/2}. \tag{10}$$

For small grazing angles, $|V_s(\theta)| \approx 1$, Eq. (7) may be rewritten as

$$-\ln|V_s(f,\theta)| = Q_s(f,W)\theta = Q_s(f,\sigma)\theta, \tag{11}$$

$$Q_s = \frac{\sqrt{2}k^{3/2}}{\sqrt{\pi g}} \frac{\Gamma\left(\frac{3}{4}\right)}{\Gamma\left(\frac{5}{4}\right)} \int_0^\infty \Omega S(\Omega) d\Omega. \tag{12}$$

At the ASIAEX site, the gravitational acceleration is $g \approx 9.793\ \mathrm{m/s^2}$. The near-surface sound speed was measured to be c=1526.7 m/s. From Eqs. (8)–(10), the rms wave height as a function of the wind speed can be found to be

$$\sigma = \left[\int_0^\infty S(\Omega)\, d\Omega\right]^{1/2} = \begin{cases} 1.769 \times 10^{-3} W_{\mathrm{NP}}^{5/2}, \\ 5.341 \times 10^{-3} W_{\mathrm{PM}}^{2}. \end{cases} \tag{13}$$

From Eqs. (8)–(12), we obtained the surface-reflection-loss factor Q_s as a function of the wind speed (W) or the rms wave height (σ) for two surface model spectra:

$$Q_s(f,W) = \begin{cases} 9.389 \times 10^{-5} f^{3/2} W_{\mathrm{NP}}^4, \\ 9.140 \times 10^{-4} f^{3/2} W_{\mathrm{PM}}^3, \\ \text{or} \\ 2.377 (f\sigma_{\mathrm{NP}})^{3/2} \sigma_{\mathrm{NP}}^{1/10} \quad \text{for NP spectrum}, \\ 2.341 (f\sigma_{\mathrm{PM}})^{3/2} \quad \text{for PM spectrum}, \end{cases} \tag{14}$$

where f is frequency in kHz, W is wind speed in m/s, and σ is given in meters.

Sea-surface scattering measurements show the importance of including wind history in scattering models and suggest that a 2- to 4-h back average for parameters directly tied to wind speed will give reasonable predictions.[2] Here, a 3-h record of W was used for back averaging. The wind speed shown in Fig. 2 was measured at 14 m elevation. This needs to be corrected to 5.5 m for W_{NP} and to 19.5 m for W_{PM}. Using the $\frac{1}{7}$ power law, the correction yields $W_{\mathrm{NP}}=2.40$ m/s and $W_{\mathrm{PM}}=2.87$ m/s for 3 June and $W_{\mathrm{NP}}=6.52$ m/s and $W_{\mathrm{PM}}=7.81$ m/s for 5 June. These values can be used to compute $\sigma_{Np}=0.016$ m and $\sigma_{PM}=0.044$ m for 3 June and $\sigma_{NP}=0.192$ m and $\sigma_{PM}=0.326$ m for 5 June. Inputting these values into Eq. (14) gives the following difference between the surface-reflection-loss values on 3 and 5 June:

$$\Delta Q_E = \begin{cases} 9.389 \times 10^{-5} f^{3/2} (6.52^4 - 2.40^4) & \textit{based on } W_{NP}, \\ 9.140 \times 10^{-4} f^{3/2} (7.81^3 - 2.87^3) & \textit{based on } W_{PM}, \end{cases} \tag{15}$$

or equivalently

$$\Delta Q_E = \begin{cases} 2.377 (f)^{3/2} (0.192^{1.6} - 0.016^{1.6}) & \textit{based on } \sigma_{NP}, \\ 2.341 (f)^{3/2} (0.326^{1.5} - 0.044^{1.5}) & \textit{based on } \sigma_{PM}. \end{cases} \tag{16}$$

ΔQ_E is defined such that $Q_{P65}=Q_{E63}+\Delta Q_E$, where Q_{P65} is the predicted loss value for 5 June based on the earlier measurements (Q_{E63}) and the known wind speed.

If the normal-mode decay shown in Eq. (4) is caused by combined bottom and surface-reflection losses (i.e., $V_E = V_b \cdot V_s$), expressed by

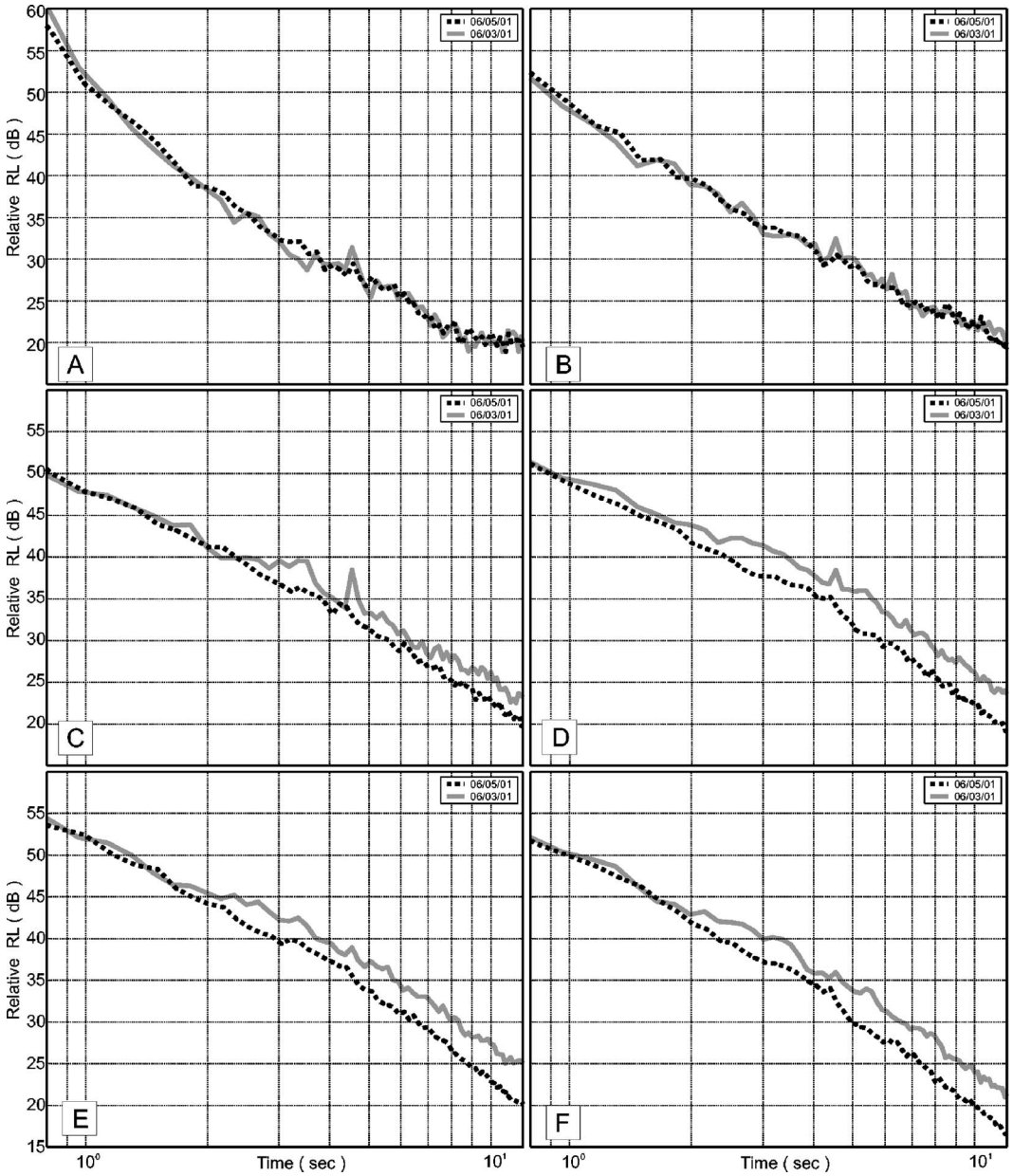


FIG. 9. RL obtained on 3 June and on 5 June; the RL decreased with increased sea state. (a) 300 Hz, (b) 500 Hz, (c) 1000 Hz, (d) 1500 Hz, (e) 2000 Hz, and (f) 2500 Hz.

$$-\ln|V_E(f,\theta)| = [Q_b + Q_s(f,\sigma)]\theta = Q_E\theta, \tag{17}$$

two prediction curves can be generated for the effective loss on 5 June predicted from the 3 June measurements. These predictions, from the NP and PM spectra, are plotted in Fig. 8 by a dashed line and a dotted line, respectively.

Figure 8 shows that the combined bottom and surface losses, inverted from the RVC measurements on 5 June, are in a range predicted by the NP and PM surface spectra.

The data/model comparison shown in Fig. 8 is based on W_{NP} and W_{PM}, or rms wave heights (σ_{NP} and σ_{PM}), derived from the wind speed by the NP and PM surface models. The values of σ_{NP} and the σ_{PM} in Eq. (16) are different from the data measured with the APL-UW buoy. There are several possible explanations for this difference. The sea states during two periods of reverberation measurements were not fully developed as the models assumed. Also, in shallow water, the JONSWAP-Mitsuyasu model has been found to be superior to the PM and NP models in predicting the rms sea surface-wave height.[16]

VI. THE EFFECT OF SEA SURFACE ON REVEBERATION LEVEL

Increasing sea state was found to increase shallow-water RVC because of faster mode stripping due to additional surface loss. The relationship of this to the effect of surface waves on the RL is not immediately apparent. On 3 June, three 1-kg explosive charges were detonated at 30.4 m from the receiver. On 5 June, four charges were detonated at 80.0 m from the receiver. Choosing the instant of detonation as time $t=0$ allows RLs to be compared in these two data sets. Figure 9 shows the relative RL as function of time in six $\frac{1}{3}$-oct bands centered at 300, 500, 1000, 1500, 2000, and

2500 Hz. The RL is the average value from three or four shots. Each data point corresponds to a time interval of 0.1707 s (1024 sample points). An echo from the R/V *Melville* can again be seen on the 3 June reverberation curve. Figure 9 shows that, for low frequencies ($f < 500$ Hz), the two days' RLs are almost identical. For higher frequencies, however, increasing sea state results in decreased RLs.[17] The same physical explanation that was applied to the observed RVC data can be used here: With the increased sea state, both higher surface scattering and reflection loss compete for dominance. For high frequency and high sea state, the surface reflection loss dominates the processes of both the RVC and RL.

VII. SUMMARY AND DISCUSSION

Two wideband reverberation measurements were made at a fixed location (the center of the ASIAEX site) in the East China Sea. Between the measurements, the wind speed and RMS surface wave height changed from 2.74 m/s and 0.10 m to 7.45 m/s and 0.33 m. The RVC and the RL from these two measurements have been analyzed as a function of reverberation time and frequency. The two days' RVC and RL data show differences that illustrate the effects of changing surface roughness. With increasing sea state, the RVC increases, but the RL decreases. The RVC-inverted combined bottom and surface losses from two reverberation measurements reflect the variation of the sea state. The data and findings, although preliminary, show the importance of including surface effects in reverberation and propagation models. Small changes in loss-per-bounce values can have cumulative effects on long-range reverberation and propagation, as well as on the accuracy of seabed geoacoustic inversion.

With increasing sea state, higher surface scattering and coherent-reflection loss compete for dominance. At high frequency and high sea state, the coherence reflection loss dominates both RVC and RL over the surface scattering at mid and long ranges. This was found to be the case for $W = 7.45$ m/s ($\sigma = 0.33$ m), $f > 500$ Hz, and reverberation times between 2 and 12 s (corresponding to 1.52 and 9.12 km) at the ASIAEX site. With increasing sea state, the effective bottom loss becomes larger, and there is faster mode stripping in the measured reverberation. Because of this, the measured RVC increases and the measured RL decreases.

The RVC and RL data in this paper were collected for only two cases where the variations of σ and W were limited to 0.23 m and 4.71 m/s, respectively. The RVC model did not explicitly include bottom or surface roughness. The RVC-based inversion technique involves some uncertainties.[8] Possible effects of sustained wind direction or near-surface bubble layers were not considered. The computed surface reflection losses were based on the NP and PM spectra. These may not be the best fit to real surface data in shallow water areas. The surface coherent reflection loss expressed by Eq. (7) and Eq. (14) is good in deep seas or in a surface channel. In a shallow-water waveguide, the energy scattered by the rough surface away from the specular direction will not be completely lost, because there is not total absorption in the seabottom. Thus, the physical explanation of the measured RVC, RL, and the RVC-inverted combined bottom-surface loss is preliminary. In shallow water, the effect of a rough sea surface on long-range sound propagation and reverberation at low and mid frequencies is a complicated physical process. It not only depends on the surface roughness and bubble distribution, but also significantly on the acoustic properties of the bottom sediment, the sound-speed profile in the water column, propagation range, receiving depth, signal frequency, and other factors. Other authors have investigated some of the complexities of this problem using numerical models.[18–22] More work needs to be done on this topic, including further measurements and comparisons between measured data and models.

ACKNOWLEDGMENTS

This work was supported by the Office of Naval Research and National Natural Science Foundation of China. The authors thank Dr. Ellen Livingston, Dr. Jeff Simmen, and Dr. Peter Rogers for their support and constructive comments. The authors would also like to express their gratitude to Dr. Peter Dahl and Jie Yang for assistance in conducting the experiment and analyzing the wave-height data. Coworkers at the Institute of Acoustics, CAS, are appreciated for at-sea assistance and for sharing experimental data. The professional and competent work of the Ship Manager (X.Y. Cai) and the captain and crew of the R/V *Shi Yan 3* are warmly acknowledged.

[1] S. T. McDaniel, "Sea surface reverberation: A review," J. Acoust. Soc. Am. **94**, 1905–1922 (1993).
[2] P. M. Ogden and F. T. Erskine, "Surface and volume scattering measurements using broadband explosive charges in the Critical Sea Test 7 experiment," J. Acoust. Soc. Am. **96**, 2908–2920 (1994).
[3] P. H. Dahl, "Forward scattering from the sea surface and the van Cittert-Zernike theorem," J. Acoust. Soc. Am. **115**, 589–599 (2004).
[4] L. M. Brekhovskikh, *Ocean Acoustics* (Nauka, Moscow, 1974), Part 4 (in Russian).
[5] L. M. Brekhovskikh and Y. Lysanov, *Fundaments of Ocean Acoustics* (Springer-Verlag, Berlin, 1982), Chap. 9.
[6] D. E. Weston and P. A. Ching, "Wind effects in shallow-water transmission," J. Acoust. Soc. Am. **86**, 1530–1545 (1989).
[7] P. H. Dahl, R. H. Zhang, J. H. Miller, L. R. Bartek, Z. H. Peng, S. R. Ramp, J. X Zhou, C. S. Chiu, J. F. Lynch, J. A. Simmen, and R. C. Spindel, "Overview of results from the Asian Seas International Acoustics Experiment in the East China Sea," IEEE J. Ocean. Eng. **29**, 920–928 (2004).
[8] J. X. Zhou, X. Z. Zhang, P. H. Rogers, J. A. Simmen, P. H. Dahl, G. L. Jin, and Z. H. Peng, "Reverberation vertical coherence and seabottom geoacoustic inversion in shallow water," IEEE J. Ocean. Eng. **29**, 988–999 (2004).
[9] J. X. Zhou and X. Z. Zhang, "Shallow-water reverberation level: measurement techniques and initial reference values," IEEE J. Ocean. Eng. **30**, 832–842 (2005).
[10] F. Ingenito, S. Wolf, and J. Miller, "Modal Acoustic Transmission Loss (MOATL)," NRL Report 8429 (1980).
[11] M. B. Porter, "The KRAKEN Normal Mode Program," SACLANT Undersea Research Centre SM-245, La Spezia, Italy, 1991.
[12] E. C. Shang, "Some new challenges in shallow water acoustics," in *Progress in Underwater Acoustics*, edited by H. M. Merklinger (Plenum, New York, 1986), pp. 461–471.
[13] Z. H. Peng, J. X. Zhou, and R. H. Zhang, "In-plane bistatic backward scattering from seabottom with random inhomogeneities sediment and rough interface," Sci. China, Ser. G **47**, 702–716 (2004).
[14] D. E. Weston, "Intensity-range relation in oceanographic acoustics," J. Sound Vib. **18**, 271–287 (1971).

[15]M. J. Buckingham, "Array gain of a broadside vertical line array in shallow water," J. Acoust. Soc. Am. **65**, 148–161 (1979).
[16]J. Yang, J. X. Zhou, and P. H. Rogers, "Effect of wind-generated surface waves on vertical coherence of sound propagation in the East China Sea," J. Acoust. Soc. Am. **119**, 3346 (2006).
[17]F. Li, J. Liu, and R. Zhang, "A model/data comparison for shallow-water reverberation," IEEE J. Ocean. Eng. **29**, 1060–1066 (2004).
[18]W. A. Kuperman and F. Ingenito, "Attenuation of the coherent component of sound propagating in shallow water with rough boundaries," J. Acoust. Soc. Am. **61**, 1178–1187 (1977).
[19]K. L. Williams, E. I. Thorsos, and W. T. Elam, "Examination of coherent reflection coefficient (CSRC) approximation in shallow water propagation," J. Acoust. Soc. Am. **116**, 1975–1984 (2004).
[20]R. Gauss, D. Fromm, K. LePage, J. Fialkowski, and R. Nero, "The influence of the sea surface and fish on long-range reverberation," in High Frequency Ocean Acoustics, edited by M. B. Porter, M. Siderius, and W. A. Kuperman, AIP Conf. Proc. **728**, 157–164 (2004).
[21]E. I. Thorsos, F. S. Henyey, W. T. Elam, S. A. Reynolds, and K. L. Williams, "Modeling shallow water propagation with scattering from rough boundaries," in High Frequency Ocean Acoustics, edited by M. B. Porter, M. Siderius, and W. A. Kuperman, AIP Conf. Proc. **728**, 132–140 (2004).
[22]E. I. Thorsos, W. T. Elam, F. S. Henyey, S. A. Reynolds, and K. L. Williams, "Scattering from rough surfaces in shallow water propagation," in *Proceedings of the 7th European Conference on Underwater Acoustics*, TU Delft (2004), pp. 559–564.

Separability of seabed reflection and scattering properties in reverberation inversion

Chris H. Harrison[a] and Peter L. Nielsen[b]
NATO Undersea Research Centre, Viale San Bartolomeo 400, 19138 La Spezia, Italy

(Received 19 July 2006; revised 4 October 2006; accepted 10 October 2006)

Separation of scattering properties (Lambert's μ) from reflection properties (the reflection loss' angle derivative α) presents difficulties in the geoacoustic inversion of long range reverberation in isovelocity water, and here it is shown that there is still a problem in a refracting environment. An alternative technique is proposed where reverberation is modified by altering the source or receiver beam pattern, for instance, using a triplet array or ring source, to provide a dipole and monopole pattern. Combinations of these two measures of reverberation then conveniently determine α and μ independently of other unknown quantities from long (or short) range data, in fact even from a single range. In addition the short range ratio of the two quantities determines the critical angle independently. The effects of refraction and other source or receiver beam patterns, including a horizontal beam and a tilted beam, are investigated by using analytical techniques. To enhance the credibility of these findings and demonstrate the benefits of the approach an example is posed as a standard inversion problem using a cost function based on both types of reverberation. Finally the technique is applied to some experimental data by forming simultaneous monopole and dipole beams in the vertical plane. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2384966]

PACS number(s): 43.30.Pc, 43.30.Gv [RCG] Pages: 108–119

I. INTRODUCTION

The idea of simultaneously separating out reflection and scattering properties from reverberation has developed over 15 years or so. It has potentially important consequences for sonar performance since it can make predictions not only of both types of seabed property but predictions over a large area at considerable range from the source ship. Amongst the earliest advocates are Zhou *et al.* (1982), Kamminga, Ellis, and Gerstoft (1993), Zhou *et al.* (1993), and Gerstoft and Ellis (1996) where scattering was assumed to follow Lambert's law. Preston (2000) and Ellis and Preston (1999) have analyzed Mediterranean reverberation extensively using manual inversion techniques, and Zhou and Zhang (2003) have analyzed data from the China Sea. Simulation studies include Muller *et al.* (2002), and Makris (1993), and extension by Rogers, Muncill, and Neumann (1998) to bistatics. Reverberation data from an exercise was selected as one of the test cases for the Geoacoustics Inversion Techniques workshop in May 2001 (Fulford, King, and Chin-Bing, 2004; and Gerstoft *et al.*, 2003). Holland (2005) investigated the effect on geoacoustic inversion of assuming a scattering law other than Lambert's, and Harrison (2005a) has attempted to determine the scattering law from simultaneous reverberation and propagation measurements. In addition there are data-model comparisons, for instance, by Desharnais and Ellis (1997), and variants combining reverberation with other information sources such as the direct blast by Heaney (2003).

At first sight one might expect the effects of scattering strength and boundary reflection to be quite distinct and separable since the first is a simple multiplier while the second, one might think, produces a decay in range. On the other hand most numerical inversion techniques rely on a given physical model with implicitly prechosen search parameters, and simply demonstrate the degree of mismatch between measurements and model. Dosso (2002) and Dosso and Nielsen (2002) have proposed an approach to quantify uncertainties stemming from the mismatch. Nevertheless, some physical intuition helps to make the initial choice of input parameters in any given inversion and to avoid unwanted correlations between them (Collins and Fishman, 1995). If one initially assumes a scattering law angle dependence of the form $\mu F(\theta_1, \theta_2, \phi)$, with known F but unknown constant μ (the three arguments defining incoming and outgoing grazing angles and bistatic angle) and one assumes that, of all the boundary interactions along a round-trip reverberation path, only one is a scatter, then it is inevitable that μ will indeed be merely a multiplier of the reverberation. So to be sure of separability one needs to be certain that the reflection parameter(s) are definitely *not* simply multipliers. Some intuition on this behavior can be gathered by applying analytical techniques to propagation and reverberation.

An earlier paper by Harrison (2003a) considered isovelocity water combined with Lambert's law and showed, by using analytical techniques, that beyond a certain range, within the so-called mode-stripping region, no such separation should be possible since the reverberation intensity was proportional to $\mu\alpha^{-2}r^{-3}$, where μ is the Lambert constant, α is the reflection derivative with angle (for small angles), and r is reverberation range (i.e., corrected travel time). In other words, in this limit, one could only determine $\mu\alpha^{-2}$, not μ and α separately. This point was recently taken up by Holland (2006) who argued that although there was little hope

[a]Electronic mail: harrison@nurc.nato.int
[b]Electronic mail: nielsen@nurc.nato.int

for extracting unique geoacoustic reflection properties from reverberation under these conditions it would still be possible to investigate or map their variability. In passing, we note that the inseparability results from any scattering law of the form $S(\theta_1, \theta_2, \phi) = \mu \times (\theta_1 \theta_2)^n f(\phi)$, and it is not just a peculiarity of Lambert's law.

The formulas for the isovelocity case were extended to uniform sound speed gradient (Harrison, 2003b, Harrison, 2005b) and showed a more complicated behavior with respect to α and μ although the tendency to inseparability was still manifest. In Sec. II we demonstrate that even with a uniform sound speed gradient an increase in μ can be approximately matched by changing α, and exactly matched by changing α and the sound speed gradient c' at the same time. From a geoacoustic inversion point of view this is a statement that given experimental reverberation with poorly known sound speed gradient c' there is an ambiguity or correlation between μ and α (and c'). If however c' is known (and not zero) then the ambiguity is weaker. In other words, the exact shape of the decay in range is unaffected by μ (since it is just a multiplier) but it is slightly affected by α. On the other hand if, in changing α, we were to keep $(\alpha \times c')$ constant then the decay shape would also be unaffected. These findings are confirmed with a modified version of the wave model C-SNAP (Ferla, Porter, and Jensen, 1993).

The main point of this paper is to suggest an alternative and in Sec. III we propose a way of separating μ and α, still from measurements of reverberation alone, even at long range and possibly a single ping, depending on configuration. The reason for the inseparability of μ and α in conventional long range reverberation is that in the mode-stripping propagation regime α becomes a simple multiplier (see Harrison, 2003a) and is indistinguishable from μ. Because the propagation is directly dependent on the angle distribution of the multipath rays (or modes) it is affected not only by the scattering law (or kernel) angle dependence but also by the directivity of the source and receiver. By changing one or both of these directivities one can modify the propagation range dependence *in situ*, and the result is a modified reverberation. Given a measurement of "conventional" reverberation with a point source and receiver, and a measurement of "modified" reverberation with some given directivity one can deduce α alone from the relative range laws. The value of $\mu\alpha^{-2}$ derived from the reverberation then permits a separation of μ. The theory can already be extended to range-dependent and refracting environments and different scattering laws. Section IV discusses the robustness of the findings in this context.

Numerical geoacoustic inversion by matching measurements with a coherent propagation and reverberation model is a powerful technique that can take advantage of aspects of reverberation behavior that are not reflected by the incoherent diffuse reverberation behavior. We might still expect better separation performance with the added information from the modified reverberation. In Sec. V we reinforce these arguments by again using the wave model C-SNAP to apply standard inversion techniques to simulated data to investigate parameter correlations and cost function minima both with and without the modified reverberation input, in the latter case achieving a much better performance.

Finally some experimental demonstrations with a steered line array are given in Sec. VI, and the reflection and scattering parameters extracted. The resulting values of α agree with other estimates and the μ is within the expected bounds.

II. REVERBERATION DEPENDENCE ON SCATTERING, REFLECTION, AND REFRACTION: CORRELATIONS

A. Isovelocity

Harrison (2003a) provided formulas for propagation, reverberation, and target echo in isovelocity, range-dependent environments with Lambert's law scattering. Treating the eigenrays or modes as a continuum in angle their angle distribution is Gaussian and they decay exponentially with the range. In a range-independent environment these formulas naturally demonstrate three clear propagation regimes according to range, already noted by Weston (1971). In the first, the Gaussian is very wide so the angle range is limited by the critical angle, and there are many contributing rays, all about the same strength. In the second, there is "mode-stripping" and the resulting Gaussian angle distribution is narrower than the critical angle. In the third, the Gaussian is so narrow that only one mode remains. These regimes can also be seen in incoherent mode solutions where the angle distribution is similarly truncated by the critical angle. In fact there is also a zeroth regime at very short range, of little interest in this context, and predicted neither by the formulas nor by discrete mode theory, where the direct path propagation follows an inverse square law. In the context of geoacoustic inversion of long- or short-range reverberation, only two of these regimes and one transition are of interest. An appropriate formula for isovelocity reverberation (Harrison, 2003a) is

$$I = \frac{\mu \Phi p}{\alpha^2 r^3} \left\{ 1 - \exp\left(\frac{-\alpha r \theta_c^2}{2H} \right) \right\}^2, \tag{1}$$

where μ is the Lambert constant, α is related to the derivative of reflection loss with angle α_{dB} (dB/rad) through $\alpha = \alpha_{\mathrm{dB}}/(10 \log_{10}(e))$, θ_c is critical angle, H and r are water depth and range, Φ is the horizontal beam width, and p is the spatial pulse length $p = ct_p/2$. At long range this becomes

$$I = \frac{\mu \Phi p}{\alpha^2 r^3} \tag{2}$$

and at short range, expanding the exponential to first order, we have

$$I = \frac{\mu \Phi p \theta_c^4}{4 r H^2}. \tag{3}$$

The transition from short- to long-range occurs when the exponential term becomes negligible. An exact transition range can be defined by equating the neighboring limiting formulas (see Appendix)

$$r_o = \frac{2H}{\alpha\theta_c^2}. \tag{4}$$

The inseparability of μ and α at long range is seen explicitly in Eq. (2) where reverberation depends on μ/α^2. In passing we note that in principle the short-range reverberation could be used to separate the two since it has no α dependence. However, poor definition of the regimes combined with spatial interference effects, excluded by these formulas, usually rule this out as a practical option.

B. Refraction

Since formulas are also available for the case of a uniform sound speed gradient (Harrison, 2003b; Harrison, 2005b) it is interesting to check the effect of refraction on separability. In a range-independent environment there are now two distinct sets of contributing propagation angles for each range: in the first set rays interact with only the boundary on the low sound speed side; in the second set they interact with both. For the same reason there are two sets of reverberation with different behavior, one from the low sound speed side of the duct, the other from the high sound speed side. We use L and H subscripts to denote quantities associated with the low and high speed boundaries. The sound speed gradient c' is embedded in the ray's radius of curvature $\rho = c/c'$. For monostatic sonar the formulas are:

Low sound speed side:

$$I_L = (I_{L1} + I_{L2})^2 r\Phi p, \tag{5}$$

where

$$I_{L1} = \frac{\mu_L^{1/2}}{2r\rho}\log\left(\frac{H}{h}\right)\exp\left(\frac{-\alpha_L r}{2\rho}\right), \tag{6}$$

where h is the greater of the distance from source, receiver, or scatterer from the low sound speed boundary,

$$I_{L2} = \frac{2\mu_L^{1/2}}{r^2(\alpha_L+\alpha_H)}(A+B)\exp\left(\frac{-(\alpha_L-\alpha_H)r}{4\rho}\right) \tag{7}$$

with

$$A = \frac{1}{2}\left\{\exp\left(\frac{-(\alpha_L+\alpha_H)r}{4\rho}\right) - \exp\left(\frac{-(\alpha_L+\alpha_H)r\theta_c^2}{2H}\right)\right\}, \tag{8}$$

$$B = \frac{(\alpha_L+\alpha_H)r}{8\rho}\left\{E_1\left(\frac{(\alpha_L+\alpha_H)r}{4\rho}\right) - E_1\left(\frac{(\alpha_L+\alpha_H)r\theta_c^2}{2H}\right)\right\}, \tag{9}$$

where $E_1(x)$ is the exponential integral (Abramowitz and Stegun, 1972).

High sound speed side:

$$I_H = I_{H2}^2 r\Phi p, \tag{10}$$

where

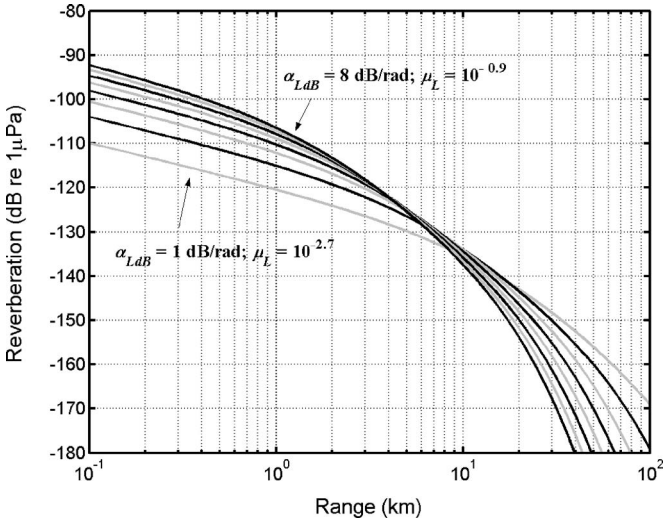


FIG. 1. Family of reverberation curves with reflection loss derivative α_{LdB} taking uniform steps between 1 and 8 dB/rad and simultaneously $\mu_L = 10^{-2.7} \times \alpha_{LdB}^2$. Other variables are fixed: $H=100$ m, $h=50$ m, $\theta_c=20°$, $c_{\min}=1500$ m/s, $c_{\max}=1520$ m/s.

$$I_{H2} = \frac{2\mu_H^{1/2}}{r^2(\alpha_L+\alpha_H)}(A-B)\exp\left(\frac{-(\alpha_L-\alpha_H)r}{4\rho}\right). \tag{11}$$

The magnitudes of the first and second exponents in Eqs. (8) and (9) are important. Those containing the seabed's critical angle θ_c are usually much larger than those containing ρ because the equivalent to θ_c in the latter terms is $\sqrt{H/(2\rho)}$, which is half the critical angle in the water column. Thus in the long-range limit the terms containing θ_c are large, making the second exponential and the exponential integral vanish. However the other exponents may, or may not be large. Note that these formulas revert to the isovelocity case when $\rho\to\infty$ since $B\to 0$ and I_{L1} vanishes.

Clearly the relationship between μ, α, and r is not so simple with refraction even if we single out reverberation from the L side. Although μ_L is still a multiplier α_L is not. This can be demonstrated numerically by plotting Eq. (5) for various values of α_L as in Fig. 1. Here α_{LdB} is varied in uniform steps between 1 and 8 dB/rad [$\alpha = \alpha_{\mathrm{dB}}/(10\log_{10}(e))$] but with $\mu_L = 10^{-2.7}\times\alpha_{LdB}^2$ simultaneously, while other parameters are constant $\alpha_H=0$ dB/rad, $\mu_H=0$, $\theta_c=20°$, $H=100$ m, $h=50$ m, $c_L=1500$ m/s, and $c_H=1520$ m/s. For clarity source strength and Φp are set to unity throughout. In effect, Φ is one radian and p is one meter, corresponding to a 1.3 ms pulse duration. A closer inspection of the formulas Eqs. (6)–(9), however, shows that if surface losses are small, $\alpha_H=0$, then in the long-range limit the subformulas can all be written in the form,

$$I = \frac{2}{r^2}\left(\frac{\mu}{\alpha^2}\right)^{1/2} \times F\left(\frac{\alpha r}{\rho}\right). \tag{12}$$

(In the case of I_{L1}, one needs to multiply the numerator and denominator by $\alpha_L r$.) So if ray radius of curvature ρ is maintained proportional to α_L the form of the reverberation range dependence is invariant as α_L is changed. Therefore it is always possible to match a change in scattering constant, μ_L, exactly with a change in reflection and sound speed proper-

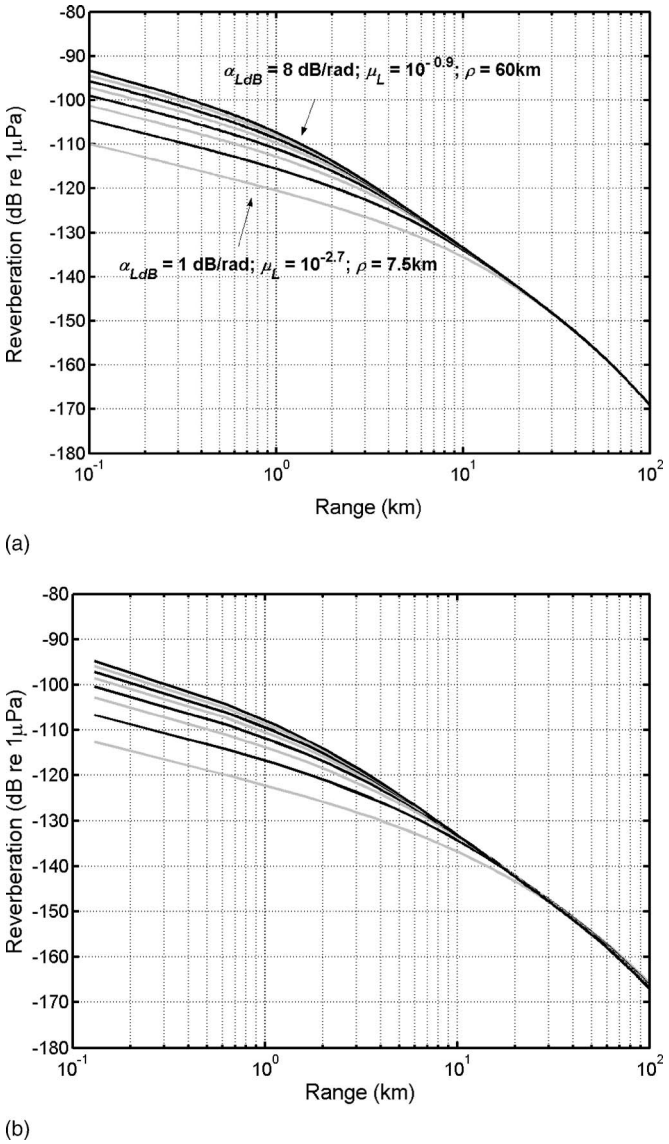


FIG. 2. Family of reverberation curves as in Fig. 1 but with μ_L and c_H simultaneously varied such that $\mu_L = 10^{-2.7} \times \alpha_{LdB}^2$ and the ray radius of curvature $\rho = 7500 \times \alpha_{LdB}$. Note the convergence at long range. (a) closed-form solution; (b) modified C-SNAP.

ties, α_L and ρ. From the geoacoustic inversion point of view one expects a strong correlation between μ_L and α_L and c'. This effect is shown in Fig. 2(a), where variation of reverberation with α_L keeping α_L/ρ constant shows clear convergence, i.e., invariance, at long range. In fact, numerically α_{LdB} again takes values 1–8 dB/rad while μ_L is set to $10^{-2.7} \times \alpha_{LdB}^2$ and c_H is varied so that $\rho = 7500 \times \alpha_{LdB}$.

C. Numerical confirmation of closed-form solutions

The findings of the previous section are here reinforced by using an independent wave model to calculate reverberation from an incoherent mode sum under the same conditions. The model uses C-SNAP (Ferla, Porter, and Jensen, 1993) operating at 1 kHz to calculate the mode shapes and horizontal wave numbers, with a separate module to do the mode summation, conversion of mode to ray angle, and calculation of group velocities, etc., according to Ellis (1995). To ensure compatible environmental inputs the seabed was composed of a half-space for which the bottom reflection loss can be calculated from the geoacoustic parameters through a relationship noted by Weston (1971), namely,

$$R_{\mathrm{dB}} = \alpha_{\mathrm{dB}} \times \theta = \frac{a_2}{\pi} \frac{d_2 c_1^2}{d_1 c_2^2 \sin^3 \theta_c} \times \theta,$$

where a_2 is the sediment volume absorption in dB/wavelength, and d, c are density and sound speed with subscripts 1, 2 referring to the upper and lower media. The actual geoacoustic properties corresponding to $\alpha_{\mathrm{dB}} = 3.98$ dB/rad, ($\alpha = 0.916$ rad^{-1}) are given in Table I. Other values of α_{dB} were mimicked by adjusting the volume absorption in proportion. As shown in Fig. 2(b) the results are virtually identical to the analytical examples in Fig. 2(a). The minor discrepancies reduce to a fraction of a dB throughout if one reverts to the variant of the analytical formula that includes the focusing or caustic effect dictated by the WKB depth dependence (Harrison, 2003b; Weston, 1980). In any case, the main point is to obtain a second opinion on the predicted long range convergence with a correlated set of α, μ, and sound speed gradient, and Fig. 2(b) leaves no doubt.

TABLE I. Baseline half-space parameters.

	Sound speed c (m/s)	Density d (kg/m^3)/1000	Vol. absorption a (dB/wavelength)
Water	1500	1.0	0.0
Sediment	1600	1.5	0.4

III. MODIFIED REVERBERATION

Harrison (2003a) assumed the propagation to follow mode-stripping and the scattering to follow Lambert's law. Other assumptions are possible, but one might expect that during any one experiment in one locality the scattering law and the propagation would be given and fixed, therefore one has no control over the reverberation. In fact there are various controls, for instance, in a refracting environment the source and receiver depth alter the results. Similarly placing either source or receiver near the sea surface creates a dipole directionality which alters the propagation by introducing an extra square-of-grazing-angle term in one or both of the outward and return angle integrals. In general, the waveguide imposes a near-Gaussian angle distribution which is multiplied by the directivities at the two ends of each propagation leg, i.e., source and scatterer for outward, and scatterer and receiver for return. There are therefore many ways one could alter the propagation integral *in situ* given the locally fixed scattering law. Some possibilities include:

- *modify source* (1): with a multiring source use one ring for the "point" source; three rings for a more directional source;
- *modify source* (2): run two rings in antiphase creating a dipole source;
- *modify source* (3): place the source near the sea surface to create a dipole (possibly problematic because of physical size and efficiency near surface);

- *modify receiver*: modify steering of a horizontal triplet array to form vertical dipoles;
- *measure angle dependence of reverberation*: use the vertical array to derive α directly from the width of the Gaussian angle distribution.

In the following we investigate three possibilities, all involving some vertical aperture, reverting to the isovelocity environment for simplicity. First the dipole, enhancing steep angles, second a horizontal beam, weakening steep angles, and third a tilted beam. In all cases we consider application to the return path only, although it is easy to extend this to the two-way path. To track the changes we note that Eq. (1) could have been written in terms of the outward and return integrals L_O, L_R as

$$I = \frac{\mu \Phi p}{rH^2} L_O L_R, \tag{13}$$

where

$$L_O = L_R = \int_0^{\theta_c} \theta \exp\left(-\frac{\alpha \theta^2 r}{2H}\right) d\theta. \tag{14}$$

We now leave L_O unchanged and calculate the new L_R as required. Introducing the shorthand

$$X = \frac{\alpha \theta_c^2 r}{2H} \tag{15}$$

we write L_R as

$$L_R = \frac{\theta_c^2}{2X}(1 - e^{-X}). \tag{16}$$

The short- and long-range limits of L_R are, respectively, $\theta_c^2/2$ and $\theta_c^2/2X$. We emphasize that we do not advocate these formulas as serious contenders for inversion directly. Rather they point the way to dependencies or correlations which can be utilized in numerical inversion by choice of the search parameters and combinations of them.

A. Vertical dipole

Assuming the dipole to be composed of two unit monopoles a distance z apart, there is a gain factor of $k^2 z^2 \sin^2 \theta$ which we approximate as $k^2 z^2 \theta^2$,

$$L_{RD} = k^2 z^2 \int_0^{\theta_c} \theta^3 \exp\left(-\frac{\alpha \theta^2 r}{2H}\right) d\theta$$

$$= k^2 z^2 \theta_c^4 \left(\frac{1 - (1+X)\exp(-X)}{2X^2}\right). \tag{17}$$

Thus the ratio of dipole to monopole reverberation F is

$$F = \frac{I_{\text{dipole}}}{I_{\text{monopole}}} = \frac{L_{RD}}{L_R} = k^2 z^2 \frac{\theta_c^2 (1 - (1+X)\exp(-X))}{X(1 - \exp(-X))}. \tag{18}$$

For long range $(X > 1)$

$$F = k^2 z^2 \frac{2H}{\alpha r}. \tag{19}$$

For short range $(X < 1)$

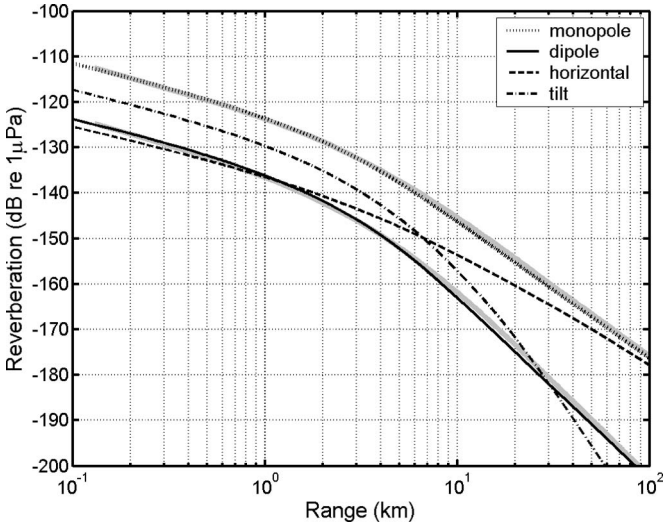


FIG. 3. Various reverberations: Monopole reverberation from the closed-form solution (dotted black) and C-SNAP (thick grey); dipole reverberation from the closed-form (solid black) and C-SNAP (thick grey); horizontal beam closed-form (dashed black); tilted beam closed-form (dashed-dotted black).

$$F = k^2 z^2 \theta_c^2/2. \tag{20}$$

The full behavior is shown as the solid black line in Fig. 3 (suppressing the $k^2 z^2$ term) for the baseline parameters ($H = 100$ m, $\mu = 10^{-2.7}$, $\alpha_{\text{dB}} = 4$ dB/rad, $\theta_c = 20°$, see Table I) and compared to the conventional monopole reverberation (dotted line). The equivalent ratio F is shown in Fig. 4. The most important point is that because of Eq. (19) the range dependence is completely altered in a very simple way at long ranges. In fact, the ratio of this modified reverberation to conventional reverberation *F is independent of* μ at all ranges, as is clear in Eq. (18). If we plot $r \times F$ as shown in Fig. 5 we expect a long range plateau with value of $2H/\alpha$. This means that if one measures monopole and dipole reverberation, then α and μ are separable after all. In fact in this example the value of the plateau is 23.37 dB

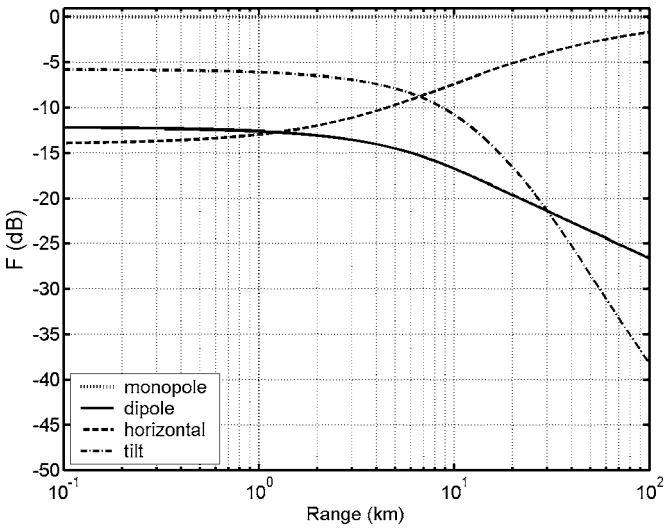


FIG. 4. The closed-form ratio (function F) of the quantities in Fig. 3 to monopole reverberation. Monopole (dotted); dipole (solid); horizontal beam (dashed); tilted beam (dashed-dotted).

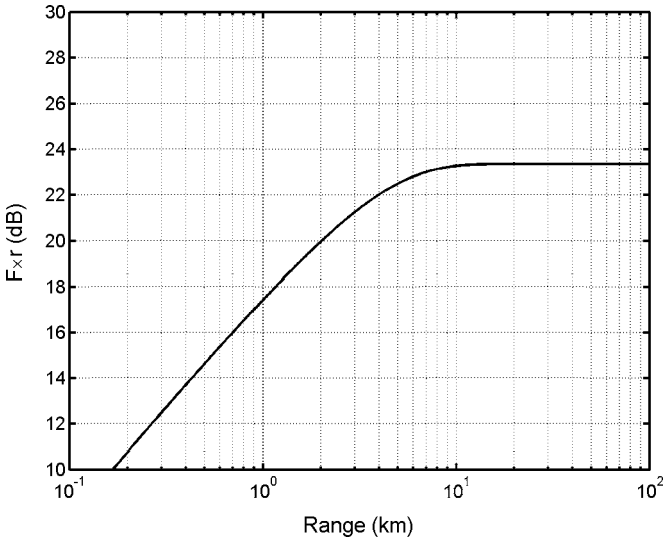


FIG. 5. The ratio multiplied by range $F \times r$ showing the plateau at long range.

which leads to $\alpha_{\mathrm{dB}}=3.9977$ dB/rad and agrees with the input value. Having separated α one can deduce μ from long range monopole reverberation. Although of less practical interest, the short range plateau in F (solid line in Fig. 4) occurs at −12.19 dB, which according to Eq. (20) gives an independent estimate of the critical angle. The estimate is $\theta_c=19.9°$ which again agrees with the input value. Despite the inaccuracy of the formulas in the presence of spatial interference, these findings suggest that a more sophisticated model of the ratio F would still be sensitive to α alone at long range and θ_c alone at short range. Notice that it is the nature of the function in Eq. (18) that the short range to long range transition "knee" is about two times further out in range than the "knee" in Eq. (1). This is a peculiarity of the function as explained in the Appendix. Numerical confirmation of these findings for the monopole and dipole is shown by the thick grey lines in Fig. 3 for the same parameters (determined as discussed in Sec. II C). Agreement is excellent.

B. Horizontal beam

A horizontal beam is conveniently represented as the Gaussian $\exp(-\theta^2/\theta_B^2)$. Although the effect on the integral is trivial the effect on reverberation is more interesting since it affects range dependence [through Eq. (15)],

$$L_{RH}=\int_0^{\theta_c}\theta\exp\left(-X\frac{\theta^2}{\theta_c^2}-\frac{\theta^2}{\theta_B^2}\right)d\theta = \frac{\theta_c^2}{2\left(X+\frac{\theta_c^2}{\theta_B^2}\right)}\left\{1-\exp\left(-\left(X+\frac{\theta_c^2}{\theta_B^2}\right)\right)\right\} \tag{21}$$

and the ratio of reverberation obtained with a horizontal beam and with an isotropic source is

$$F=\frac{L_{RH}}{L_R}=\frac{X}{\left(X+\frac{\theta_c^2}{\theta_B^2}\right)}\frac{\left\{1-\exp\left(-\left(X+\frac{\theta_c^2}{\theta_B^2}\right)\right)\right\}}{\{1-\exp(-X)\}}. \tag{22}$$

The transition determined by the magnitude of the exponent is now when

$$X+\frac{\theta_c^2}{\theta_B^2}\sim 1. \tag{23}$$

If the beam is narrower than the critical angle then there is essentially no transition in the exponential—it is negligible for all ranges—but there is one in the denominator. For

$$X<\frac{\theta_c^2}{\theta_B^2},\quad L_{RH}=\theta_B^2/2, \tag{24}$$

whereas for

$$X>\frac{\theta_c^2}{\theta_B^2},\quad L_{RH}=H/\alpha r, \tag{25}$$

which leads to the same r^{-3} behavior for isotropic source and receiver ($F=1$). The effect is shown in Figs. 3 and 4 for $\theta_B=4°$. In summary a narrow horizontal beam has an effect, but it is too weak to be useful for inversion.

C. Tilted beam

If the beam of width θ_B is tilted upwards at θ to the horizontal we have

$$\begin{aligned}L_{RT}&=\frac{1}{2}\int_{-\theta_c}^{\theta_c}|\theta|\exp\left(-X\frac{\theta^2}{\theta_c^2}\right)\exp\left(-\left(\frac{\theta-\theta_o}{\theta_B}\right)^2\right)d\theta\\&=\left\{\int_0^{\theta_c}\theta\exp(-A\theta^2-B\theta-C)d\theta+\int_0^{\theta_c}\theta\exp(-A\theta^2+B\theta-C)d\theta\right\}/2,\end{aligned} \tag{26}$$

where

$$A\equiv\frac{\alpha r}{2H}+\frac{1}{\theta_B^2}=\frac{X}{\theta_c^2}+\frac{1}{\theta_B^2};\quad B\equiv\frac{2\theta_o}{\theta_B^2};\quad C\equiv\frac{\theta_o^2}{\theta_B^2}.$$

It is possible to solve this exactly for arbitrary θ_o, θ_B, θ_c but a reasonable approximation is for large θ_c (see 3.462/5, Gradshteyn and Ryzhik, 1980),

$$L_{RT}=\frac{\exp(-C)}{2A}\left\{1+\frac{B\sqrt{\pi}}{2\sqrt{A}}\exp\left(\frac{B^2}{4A}\right)\mathrm{erf}\left(\frac{B}{2\sqrt{A}}\right)\right\}. \tag{27}$$

The corresponding reverberation and F function are shown in Figs. 3 and 4 for $\theta_B=4°$ and $\theta_o=15°$. A poorer but more insightful approximation is a narrowish beam inside the critical angle, i.e:, $\theta_B\ll\theta_c$; $\theta_o<\theta_c$,

$$L_{RT}=\frac{\sqrt{\pi}}{2}\theta_B\theta_o\exp\left(-\frac{\alpha\theta_o^2 r}{2H}\right)=\frac{\sqrt{\pi}}{2}\theta_B\theta_o\exp\left(-X\frac{\theta_o^2}{\theta_c^2}\right), \tag{28}$$

$$F = \frac{\alpha r}{H}\frac{\sqrt{\pi}}{2}\theta_B\theta_o \exp\left(-\frac{\alpha\theta_o^2 r}{2H}\right)$$
$$= \frac{\sqrt{\pi}}{2}\frac{\theta_B\theta_o}{\theta_c^2} X \exp\left(-X\frac{\alpha\theta_o^2}{\theta_c^2}\right). \quad (29)$$

In theory, this behavior is virtually the same as that of the integrand of Eqs. (26) or (14), and so it contains more information (given a variable beam angle) than, say the dipole beam. Nevertheless the simple relationship between dipole and conventional reverberation makes it more attractive.

IV. ROBUSTNESS AND EXTENSIONS

The main point of this paper, and in particular Sec. III, is to demonstrate that reflection properties can, after all, be separated out from reverberation. Since the derivations have made certain assumptions we explore their weaknesses here.

A. Range-dependence

If we assume variable bathymetry, isovelocity reverberation formulas are given by Harrison (2003a) in terms of an effective depth, conceived by Weston (1976). These have been justified and favorably compared with the model BiStaR (LePage and Harrison, 2003; Harrison, 2005c). As is apparent from the original definition, a variable α can also be accommodated as part of the effective depth integral, so the original term in the exponent $\alpha\theta_s^2\int dr/H(r)^3$ becomes $\theta_s^2\int \alpha(r)dr/H(r)^3$ (see also Holland, 2006) and we define a modified effective depth

$$\chi_{\text{eff}}(r) = \frac{H_s^2 H(r)^2}{r}\int_0^r \alpha(r)dr/H(r)^3 \quad (30)$$

in between a notional fixed source at range zero and receiver at r. Following Harrison's derivation through, and making the substitution [cf. Eq. (15)],

$$X(r) = \frac{\theta_c^2 H_c^2 \chi_{\text{eff}} r}{2H_s^2 H(r)^2} \quad (31)$$

the monopole reverberation with variable α and μ is

$$I_{\text{monopole}} = \frac{\mu\Phi p}{r^3}\left\{\frac{H_s}{\chi_{\text{eff}}}(1-\exp(-X))\right\}^2. \quad (32)$$

Following Sec. III A the dipole reverberation is

$$I_{\text{dipole}} = \frac{\mu\Phi p}{r^3}\left\{\frac{H_s}{\chi_{\text{eff}}}(1-\exp(-X))\right\} \times\left\{\frac{H_s}{\chi_{\text{eff}}}(1-(1+X)\exp(-X))\right\}\frac{2H(r)^2}{r\chi_{\text{eff}}} \quad (33)$$

and the ratio of the two (suppressing the k^2z^2) is

$$F(r) = \frac{2H(r)^2}{r\chi_{\text{eff}}}\frac{(1-(1+X)\exp(-X))}{(1-\exp(-X))}. \quad (34)$$

Since range and bathymetry are assumed known, the ratio F at long range tells us χ_{eff} (r), and from its definition, Eq. (30), range-smoothing followed by differentiation of the quantity $\chi_{\text{eff}}r/(H_s^2H(r)^2)$ gives $\alpha(r)/H(r)^3$. Therefore the range-dependent $\alpha(r)$ can be computed directly, in principle. In this same long range limit, knowing $\alpha(r)$ and $H(r)$ one can then compute $\mu(r)$ from the absolute value of I_{monopole} at each range. Note that θ_c, whether range-dependent or -independent is not obtainable in the long range limit.

B. Scattering law

From Eqs. (13) and (14) it is clear that an angle-separable scattering law of the more general form $S(\theta_{\text{in}},\ \theta_{\text{out}})=\mu(\theta_{\text{in}}\times\theta_{\text{out}})^n$ can be accommodated in the derivations of Sec. III. The ratio of dipole to monopole reverberation in the long range limit becomes

$$\frac{I_{\text{dipole}}}{I_{\text{monopole}}} = \frac{\int_0^\infty \theta^{n+2}\exp(-a\theta^2)d\theta}{\int_0^\infty \theta^n\exp(-a\theta^2)d\theta}, \quad (35)$$

where a is a constant, representing some function of α, H, r, etc., and changing the variable to $Z=a\theta^2$ we have

$$\frac{I_{\text{dipole}}}{I_{\text{monopole}}} = \frac{1}{a}\frac{\int_0^\infty Z^{(n+1)/2}\exp(-Z)dZ}{\int_0^\infty Z^{(n-1)/2}\exp(-Z)dZ}$$
$$= \frac{1}{a}\frac{\Gamma(1+(1+n)/2)}{\Gamma((1+n)/2)} = \frac{1}{a}\times\frac{(1+n)}{2} \quad (36)$$

for any positive value of n (not necessarily integer) where Γ is the gamma function (Abramowitz and Stegun, 1972). We obtain the same a from the ratio, whether as in Eq. (19) or the range-dependent version, Eq. (34), but with a small change in the scaling factor of $(1+n)/2$, i.e., 1 for Lambert's law, 0.75 for Lommel-Seeliger (or square root equivalent, see Holland, 2005), 0.5 for angle-independent. Thus the measure of α is relatively insensitive to discrepancies between the actual and assumed scattering law. However one expects the assumption of Lambert in the presence of a lower angle power in reality to lead to an overestimate of α.

In the extreme case of a nonangle-separable scattering law $S(\theta_{\text{in}},\theta_{\text{out}})=\mu f(\theta_{\text{in}},\theta_{\text{out}})$ one can still say several things. First, $f(\theta_{\text{in}},\theta_{\text{out}})=f(\theta_{\text{out}},\theta_{\text{in}})$. Second, the functional form of f will clearly affect the result. But, most importantly, the dipole to monopole ratio does not depend on the strength of the scattering, μ, since it cancels. It therefore *does* depend on reflection properties modulated by the function f.

C. Refraction

The effects of refraction on reverberation were explored in Sec. II B, Eqs. (5)–(11). Of interest here is the effect of refraction on the long range ratio F. First, consider orders of magnitude. In the long range limit the second terms in Eqs. (8) and (9) tend to zero. For strong refraction we might have $\rho=c/c'\sim 1500\times 100/20=7.5$ km and $\alpha=0.5$ rad^{-1} so that the decay distance for the first exponential in Eq. (8) (i.e., $4\rho/\alpha$) would be 60 km. In a rough calculation we can ne-

glect the B term since it is proportional to the *difference* between reverberation from the low sound speed and the high sound speed side of the duct. Equations (6) and (7) reduce to

$$I_{\text{monopole}} = \frac{\mu\Phi p}{r^3\alpha^2}\left(\frac{\alpha r}{2\rho}\log\left(\frac{H}{h}\right)+1\right)^2 \exp(-\alpha r/\rho). \quad (37)$$

The ray angle at the array is always in between that at the sea surface and seabed, but since we have already neglected the difference, it is straightforward to form I_{dipole} by inserting a θ^2 term in the integrals that resulted in Eqs. (6) and (7). The result is

$$I_{\text{dipole}} = \frac{\mu\Phi p}{r^3\alpha^2}\left(\frac{\alpha r}{2\rho}\log\left(\frac{H}{h}\right)+1\right)\left(\frac{\alpha r}{2\rho}\frac{2(H-h)}{\rho}+\frac{H}{2\rho}+\frac{2H}{r\alpha}\right) \times\exp(-\alpha r/\rho). \quad (38)$$

Writing $b \equiv \alpha r/(4\rho)$ and $\eta \equiv h/H$ the ratio is

$$F = \frac{2H}{\alpha r}\times\frac{1+b+8b^2(1-\eta)}{1-2b\log(\eta)}. \quad (39)$$

Clearly the dependence of F on H, α, and r is unchanged except for a numerical factor, which for $b=1$ and $\eta=0.5$ is 2.51. Thus the numerical inversion technique is robust in the presence of refraction.

D. Formation of the dipole beam pattern

In practice it may be difficult to produce a perfect dipole response. It goes without saying that the null needs to be as deep as possible since it is attempting to reject near horizontal paths. If there is a hint of monopole behavior of the form $B(\theta)=\delta+k^2z^2\sin^2\theta$, then the expected weak returns and rapid fall-off with range will be contaminated and the function F [Eq. (19)] will be poorly estimated.

V. STANDARD INVERSION TECHNIQUES

A. Standard geoacoustic inversion of conventional reverberation only

Here the objective is to demonstrate, by examining a cost function, that application of standard inversion techniques to conventional reverberation alone would yield poor estimates of α and μ. The reverberation was again simulated by the modified C-SNAP already described in Sec. II C for an isovelocity water column and flat bottom environment as in Table I, again varying α by adjusting the volume absorption a_2 in proportion. The baseline values of α and μ are 0.92 rad^{-1} (corresponding to $\alpha_{\text{dB}}=4$ dB/rad) and -27 dB. For demonstration purposes we deliberately avoid introducing frequency variation in these parameters; in practice there may be an additional benefit in including such variation in the cost function. Figure 6 shows the misfit ambiguity surface (least-mean-square) as a function of μ and α^2 and averaging over ranges from 3 km to 50 km. The baseline values are indicated by a plus sign. There is a clear trough along the diagonal where $\mu\propto\alpha^2$ is predicted by Eq. (2). However the bottom of the trough is not absolutely flat so it is possible to obtain a fit (indicated by the circle), though only in the absence of noise. "Noise" can mean several things in this context. First, from a pure inversion theory standpoint, one might envisage some randomness added in the α, μ plane. Secondly ambient noise, when added to reverberation measurements only affects the longest ranges, and one therefore should, as far as possible, arrange the experiment to minimize ambient noise (see Sec. VI). Thirdly, a discrepancy between the chosen physical model and the experimental measurements may result in spatial (or temporal) fluctuations which spoil the goodness of match. Obviously the slightest disagreement between the physical model and measurements will result in arbitrary shifts predominantly along the trough. This is a manifestation of the inseparability of α and μ that this paper addresses. How the inversion algorithm copes with this trough and noise depends on which algorithm is chosen, and a separate investigation is out of the scope of this paper.

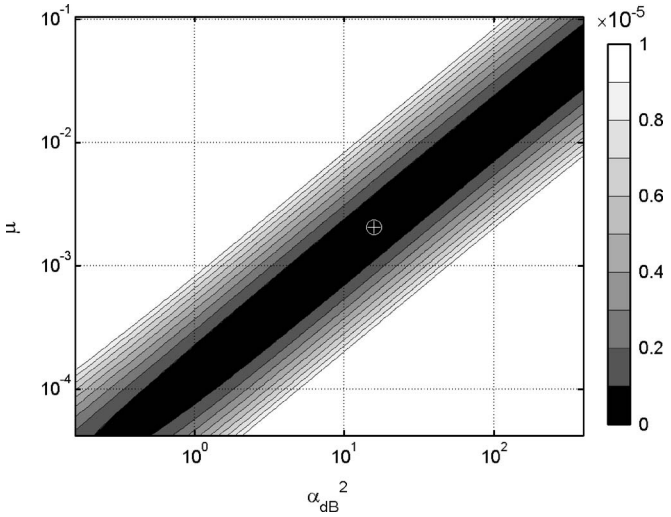


FIG. 6. Behavior of least-mean-square cost function based on monopole reverberation alone. The baseline values of α_{dB} and μ are indicated by a +, and the deduced solution by a ◦.

B. Simultaneous geoacoustic inversion of reverberation and dipole reverberation

With the addition of, effectively, a second measurement there are several options for processing. One possibility is to construct a joint cost function following the approach used by Nielsen (2006) where the two measurements were reverberation and propagation. In the present context where we are only searching for two parameters we have two measurements (even at a single range) and two unknowns, μ and α. If one were to assume the long range versions of Eqs. (1) and (18), namely Eqs. (2) and (19), to be true one could write down a formula for α only and a formula for μ only. Instead, for reasons that will become apparent, we *define* a quantity A

$$\mathrm{A} \equiv k^2z^2\frac{2HI_{\text{monopole}}}{rI_{\text{dipole}}} = k^2z^2\frac{2H}{rF} = \alpha\times\frac{(1-\exp(-X))}{(1-(1+X)\exp(-X))} \quad (40)$$

such that, in the long range limit it would converge on α. Similarly we define a quantity M

$$\mathrm{M} \equiv \frac{I_{\text{monopole}}^3 r}{I_{\text{dipole}}^2} \frac{(2Hk^2z^2)^2}{\Phi p}$$
$$= \mu \times \left(\frac{(1-\exp(-X))^2}{(1-(1+X)\exp(-X))} \right)^2 \quad (41)$$

that in the long range limit would converge on μ. From the point of view of numerical inversion the two parameters A and M are already orthogonal in the long range limit and do not require any coordinate rotations (see Collins and Fishman, 1995). Their orthogonality suggests the following joint cost function,

$$E(\alpha_S, \mu_S) = \sum_i (\mathrm{A}_i - \mathrm{A}_i'(\alpha_S, \mu_S))^2 + w \times (\mathrm{M}_i - \mathrm{M}_i'(\alpha_S, \mu_S))^2, \quad (42)$$

where A, M denote measurements and A', M' denote modeled quantities, w is a relative weighting found by trial and error, and the summation is assumed to be over a selection of ranges. Figure 7 is a plot of E as a function of the search parameters α_S, μ_S for ranges between 3 km and 50 km. The first and second terms of Eq. (42) are shown individually in Figs. 7(a) and 7(b), so that one can see the closeness to orthogonality of the functions, then combined in Fig. 7(c). The improvement over Fig. 6 is dramatic, and there is a well defined solution for both parameters which agrees with the simulated experimental values of $\alpha=0.92$ and $\mu=10^{-2.7}$ as shown by the plus symbol in the figure. In fact the fit is still good for shorter ranges, and the reason for this is that the function of X on the right-hand side of Eq. (41) tends to a constant (actually 2^2) for small X as well as tending to a constant (1) for large, so numerically μ can be fixed. Because μ is fixed, α can also be fixed despite the fact that the function of X at the right-hand side of Eq. (40) tends to $2/X$ leading to the whole right-hand side being a constant, independent of α. By a useful quirk, that constant $4H/(\theta_c^2 r)$ depends only on the critical angle.

In retrospect it is clear from Eqs. (40) and (41) that a measurement of monopole and dipole reverberation at even a single range (given H, z, etc.) is enough to define A and M and therefore α and μ if that range is long. For instance, in this simulated example (or Fig. 3) we have $I_{\text{monopole}}=155.3$ dB and $I_{\text{dipole}}=174.9$ dB at 20 km. This leads to $\alpha=0.91$, $\mu=10^{-2.7}$ again agreeing with the simulated experimental values. Obviously the sensitivity to the two absolute reverberation levels requires some kind of averaging, but not necessarily over range. Although this appears to obviate numerical inversion altogether, in a more realistic inversion one might be searching for more than just two parameters (for instance, critical angle or geoacoustic layer parameters) so one can justify using the more robust standard inversion technique.

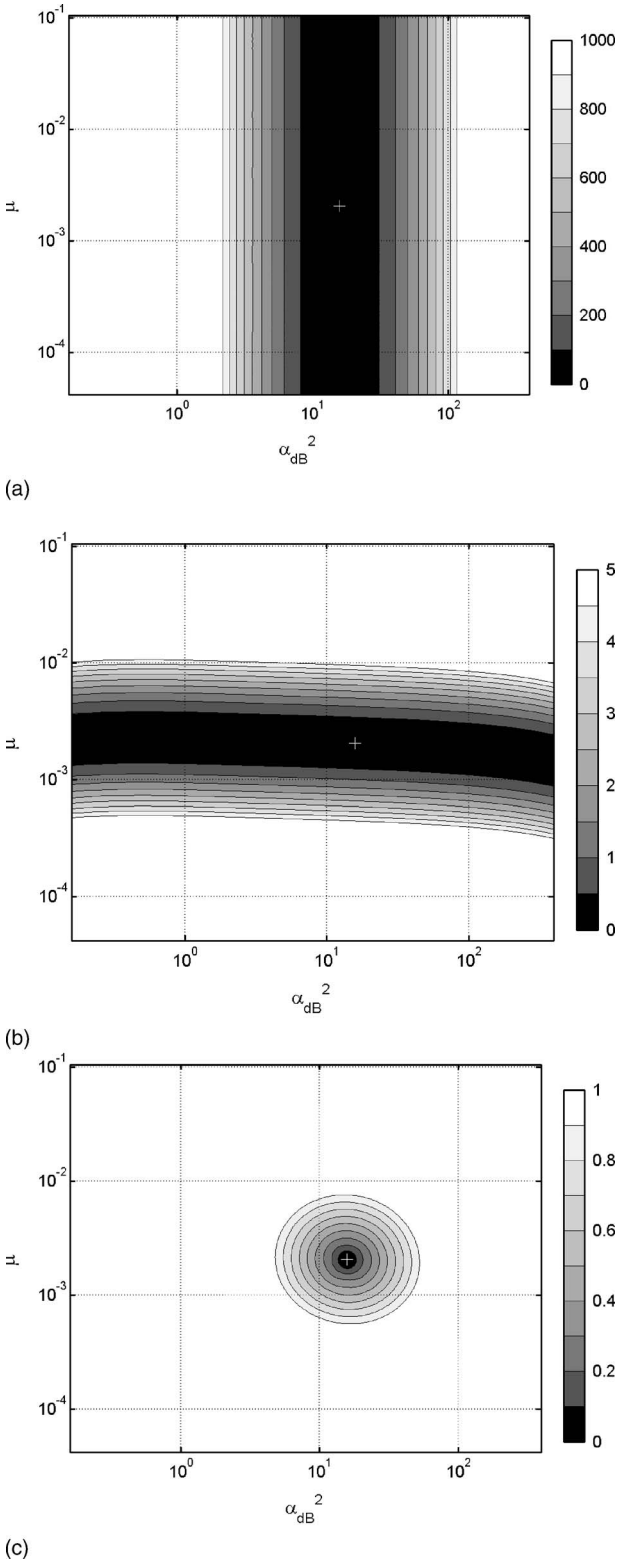


FIG. 7. Behavior of the first (a) and second (b) quasiorthogonal term in the cost function [Eq. (42)], and the joint cost function (c) derived from monopole and dipole reverberation. Baseline values are indicated by a +.

VI. EXPERIMENTAL DETERMINATION OF α FROM THE REVERBERATION RATIO

An ideal dipole receiver would be a towed horizontal triplet array since it is readily available as a means of resolving the left/right ambiguity by cardioid beam-forming. In principle the triplets can be weighted as $[\sin(\phi); \sin(\phi+2\pi/3); \sin(\phi+4\pi/3)]$, knowing the roll angle for each triplet, to form an effective vertical dipole. Unfortunately suitable data have not been found by these authors. Instead we use here data gathered with a vertical array (VLA) to mimic a sine-squared beam since all beams are known. Of

course, in this case, one could treat each vertical beam as a "tilted beam" (Sec. III C), but the simplicity of the dipole formulas and the appeal of a dipole or triplet array compels us to demonstrate that approach. The data were collected at three sites on the Malta Plateau during BOUNDARY2003 (July 8th, 2003) and BOUNDARY2004 (May 17 and 20, 2004), and have already been described in Harrison (2005a). Each set consists essentially of monostatic reverberation on a gradual slope with the source and array in about 150 m water depth but with returns from water shallowing to about 80 m over some tens of kilometers. There was a strong summer profile in 2003 with a velocity contrast of 30 m/s, whereas in 2004 the gradients were much weaker with a contrast of about 10 m/s. The source was a sweep from 700 to 1500 Hz at 209 dB re 1 μPa at 1 m, and the 32-element VLA had angle resolution of a few degrees at its design frequency, 1500 Hz. Given the smoothed beam responses it is an easy matter to sum over angles (monopole) and multiply by $\sin^2(\theta)$ then sum (dipole). The match-filtered response for the three dates extending out to about 25 km are shown in Figs. 8(a)–8(c), and these can be compared with Fig. 3. The spike near 10 km on the 8th and the 20th is the Campo Vega oil rig and tender. From approximately 1 to 25 km in all cases one can see the expected divergence of the monopole and dipole curves. The weaker dipole curves tend to flatten off into ambient noise at a shorter range. Also the Campo Vega spike is weaker compared with its surrounding in the dipole case, which is to be expected since it is a predominantly horizontal return. Both the continuum of angles theory in Secs. II and III and the discrete mode theory of C-SNAP exclude angles greater than the critical angle. They therefore do not model behavior correctly at short range. However, it is interesting to note that the direct path propagation (through its angle-independence) would lead to the same formula for F [Eq. (20)] in this region. In fact the dB difference of 12 or 13 dB (which is strikingly uniform despite the spatial fluctuations of the individual curves) is close to that expected for a critical angle of about 20°.

The equivalent of Fig. 5 for all cases is shown in Fig. 9. Now one can clearly see the effects of ambient noise beyond about 10 km, and the fall-off at short range, leaving a plateau in the middle at 23, 21, 21 dB, respectively. If a flat bottom and Lambert's law were assumed with corresponding depths of 149, 143, 165 m we would arrive at α values of 1.49, 2.27, 2.62 rad^{-1}. Allowing for the fact that the dominant reverberation is from the shallow side of the array a more realistic average depth would be 100 m, so the values of α would reduce to 1.00, 1.58, 1.58 rad^{-1}. One could also include more detailed bathymetry and azimuthal effects through an effective depth, as in Harrison (2003a) or Harrison (2005c), where three-dimensional slope effects were considered. However the effective depth with a uniform bottom slope is just the average of the endpoint depths so we would expect the above estimate to remain valid. Note that we have no reason to suspect spatial changes in α in this environment so we do not invoke Eq. (30).

Since the sound speed profiles were not isovelocity it is interesting to calculate the correction to the long range ratio F using Eq. (39). In fact $\eta=h/H\approx 0.5$ and assuming α $=1$ rad^{-1} and a (uniform) sound speed gradient of 0.1 s^{-1} the extra factor is 1.04 at 10 km. Increasing the sound speed gradient to 0.3 s^{-1} the extra factor becomes 1.5 at 10 km. These factors are close enough to unity to justify the isovelocity estimates above.

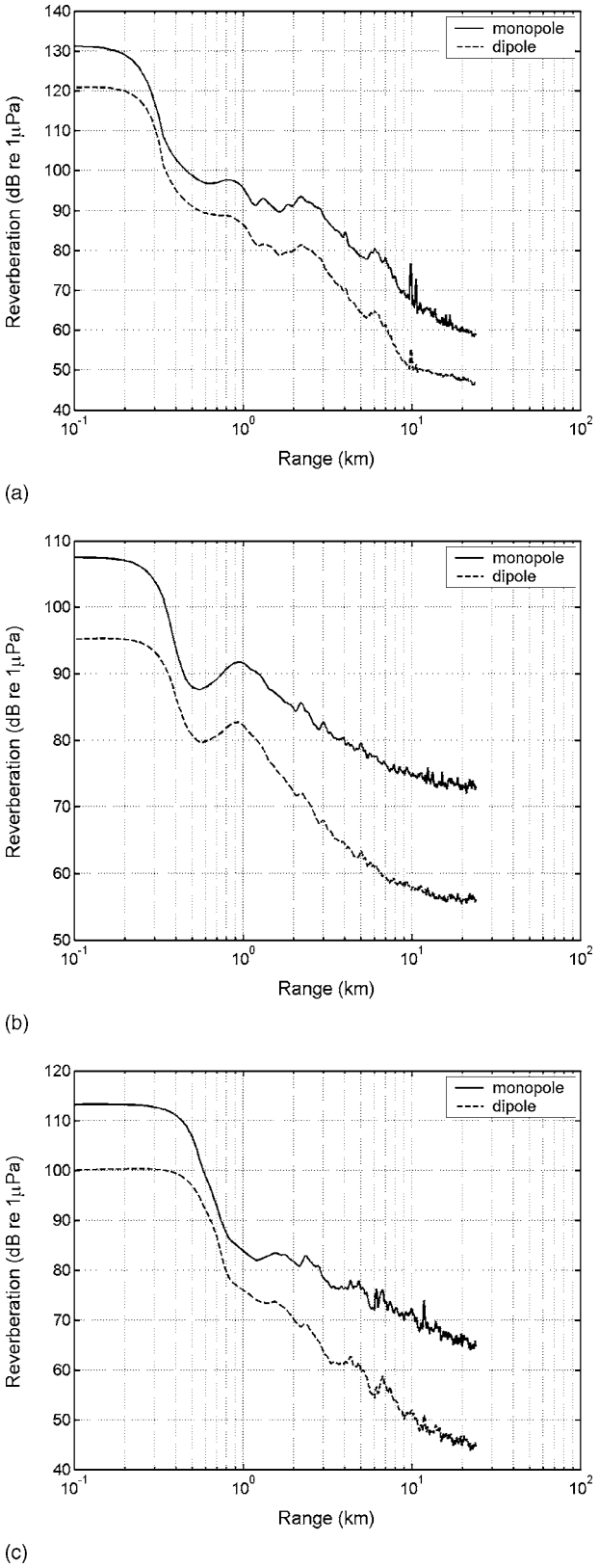


FIG. 8. Monopole (solid) and dipole (dashed) reverberation on the Malta Plateau for (a) July 8, 2003; (b) May 17, 2004; (c) May 20, 2004.

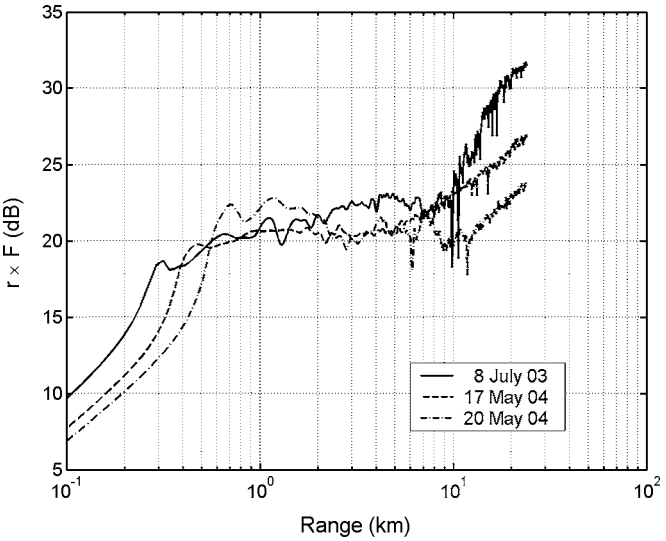


FIG. 9. The ratio of dipole to monopole reverberation multiplied by range for July 8, 2003 (solid), May 17, 2004 (dashed), May 20, 2004 (dashed-dotted).

There is evidence that the scattering law may have a weaker dependence on angle than Lambert (Holland, 2006), and it was shown in Sec. IV B that a law, such as Lommel-Seeliger, with $n=0.5$, would result in an extra factor of 0.75 in the dipole-monopole ratio. This would reduce the α values further to 0.75, 1.19, 1.19 rad^{-1}. These values are close to the mean (1.1 rad^{-1}) and well within the bounds set by an independent measurement technique, already reconciled with other measurements, for the same area based on one-way propagation paths (Prior and Harrison, 2004). In fact the upper Lambert values and the lower Lommel-Seeliger values lie on the edge of their error distributions.

Inserting $\alpha=1.1$ rad^{-1} into Eq. (1) and taking the monopole reverberation value of 68 dB at a range of 10 km, with $\Phi=2\pi$, $p=1.87$ m (1.25 ms pulse), and vertical beam width 3.5°, we find $\mu=-33$ dB. This value is close to other measurements for the area (see for instance, Holland, 2006), for which there is a wide spread, and of course this evaluation depends on the sonar calibration, unlike the evaluation of α. In addition, for comparison purposes, long range measurements of μ that are immune from propagation uncertainty are hard to come by for the reasons central to this paper.

VII. CONCLUSIONS

An analytical approach resulting in closed-form solutions has been used and extended to propose a more robust way of separating and extracting geoacoustic and scattering parameters from reverberation. From a philosophical point of view the marrying of analysis with a numerical technique has been demonstrably fruitful.

An earlier paper (Harrison, 2003a) suggested on theoretical grounds that in an isovelocity environment long range reverberation is a function of μ/α^2 (where α is the angle derivative of reflection loss and μ is the scattering coefficient). So α and μ cannot be separated by inverting reverberation alone. In this paper we have shown that in a refracting environment there is a strong correlation, i.e., relationship, between μ, α and sound speed gradient (ray radius of curvature ρ). Because of this, the shape of the reverberation curve at long range for a given μ, can always be matched by a change in α *and* ρ. Thus if ρ is predetermined then, in principle, α and μ *can* be separated since they affect the curve in different ways. However if, ρ is unknown there is enough freedom to match a change in μ with a joint change in α and ρ.

This paper has proposed an alternative geoacoustic inversion technique in which the reverberation is deliberately modified by biasing the propagation angles through a source or receiver beam, thus providing two separate measures of reverberation. The relative behavior of these two measures then provides the required separability between scattering and geoacoustic parameters. Of the various possibilities, three were investigated: a dipole beam, a narrow horizontal beam, and a tilted beam. The dipole reverberation combined with monopole (conventional) reverberation leads to some surprisingly powerful, simple, and robust results. For instance, from the ratio at long range we can find α directly and independently of other parameters (other than water depth). Having found α one can then deduce μ. The short range ratio is sensitive to critical angle and insensitive to μ and α. These deductions were made by extending Harrison's closed-form solutions, and they were confirmed using a modified version of the wave model C-SNAP. Although the choice of measurements (i.e., the ratio) was suggested by appeal to closed-form solutions this paper does not advocate using closed-form solutions for inversion.

Simulations using C-SNAP were used to test the separability of μ and α in the context of geoacoustic inversion of reverberation. A comparison was made between an approach based on monopole reverberation alone and joint monopole and dipole reverberation. The first case leads to a strong correlation between μ and α^2, as expected, and therefore the prospect of difficult separation with noisy data. The second used the theory to suggest orthogonal search parameters which then led to a very well defined minimum.

Although the simulation assumed an isovelocity range-independent environment it was shown mathematically that the approach can handle range-dependent environments and still separate α, μ in the long range limit. The approach is robust to refraction and the derived α is relatively insensitive to the assumed scattering law.

Separability of α and μ was demonstrated with experimental reverberation data from a vertical array. Since elevation angles were known it was possible to form an exact analog of the monopole and the dipole receiver beams by multiplying by, respectively, unity and sine squared of beam angle, before summing over all angles at each delay. The results clearly showed the expected divergence of the two varieties of reverberation at long range from which one could derive separately the three quantities, α, μ, and θ_c. This operation ought to, in principle, be possible with a towed horizontal triplet array steered as a vertical dipole or monopole.

ACKNOWLEDGMENTS

The authors would like to thank the Captain and crew of the *NRV Alliance* and Enzo Michelozzi, Luigi Troiano and

Piero Boni for taking care of the hardware and data acquisition during the BOUNDARY experiments.

APPENDIX: TRANSITION FROM "SHORT RANGE" TO "LONG RANGE"

Although it might seem a bit fussy to investigate as vague a phenomenon as the transition from short range to long range behavior (the "knee"), its position has important consequences for the viability of this technique in practice since it influences the lower limit of "long range" where the theory is at its best. As a working definition of a knee we take the well-defined point on a $\log(r)$ graph where the short range and long range asymptotes intersect. Taking Eq. (1), the long- and short-range limiting forms are already specified in Eqs. (2) and (3). The intersection is found by simply equating the two, thus the knee is at

$$r_o = \frac{2H}{\alpha\theta_c^2} \quad \text{or } X = 1$$

as we might have guessed. To do the same operation on Eq. (18), we equate the long and the short range asymptotes given by Eqs. (19) and (20), and we find

$$r_o = \frac{4H}{\alpha\theta_c^2} \quad \text{or } X = 2.$$

Thus the knee of the ratio of dipole to monopole reverberation occurs at twice the range of the monostatic reverberation knee. As pointed out in Harrison (2003a) the latter tends to occur at about 20 water depths for many bottom types.

Abramowitz, M. and Stegun, I. A. (**1972**). *Handbook of Mathematical Functions* (Dover, New York).

Collins, M. D. and Fishman, L. (**1995**). "Efficient navigation of parameter landscapes," J. Acoust. Soc. Am. **98**, 1637–1644.

Desharnais, F. and Ellis, D. D. (**1997**). "Data-model comparisons of reverberation at three shallow-water sites," IEEE J. Ocean. Eng. **22**, 309–316.

Dosso, S. E. (**2002**). "Quantifying uncertainty in geoacoustic inversion. I. A fast Gibbs sampler approach," J. Acoust. Soc. Am. **111**, 129–142.

Dosso, S. E. and Nielsen, P. L. (**2002**). "Quantifying uncertainty in geoacoustic inversion. II. Application to broadband, shallow-water data," J. Acoust. Soc. Am. **111**, 143–159.

Ellis, D. D. (**1995**). "A shallow-water normal-mode reverberation model," J. Acoust. Soc. Am. **97**, 2804–2814.

Ellis, D. D. and Preston, J. R. (**1999**). "Extracting sea-bottom information from reverberation data," J. Acoust. Soc. Am. **105**, 1042.

Ferla, C. M., Porter, M. B., and Jensen, F. B. (**1993**). "C-SNAP: Coupled SACLANTCEN normal mode propagation loss model," NATO Undersea Research Centre Memorandum Report No. SM-274, December 1993.

Fulford, J. K., King, D. B., Chin-Bing, S. A., and Chapman, N. R. (**2004**). "Benchmarking geoacoustic inversion methods using range-dependent field data," IEEE J. Ocean. Eng. **29**, 3–12.

Gerstoft, P. and Ellis, D. D. (**1996**). "Application of multifrequency inversion methods to obtain seabed properties from broadband reverberation data," J. Acoust. Soc. Am. **100**, 2665.

Gerstoft, P., Hodgkiss, W. H., Kuperman, W. A., and Song, H. (**2003**). "Phenomenological and global optimization inversion," IEEE J. Ocean. Eng. **28**, 1–10.

Gradshteyn, I. S. and Ryzhik, I. M. (**1980**). *Table of Integrals, Series, and Products* (Academic, New York).

Harrison, C. H. (**2003a**). "Closed-form solutions for reverberation and signal-excess with mode-stripping and Lambert's law," J. Acoust. Soc. Am. **114**, 2744–2756.

Harrison, C. H. (**2003b**). "Signal and reverberation formulae including refraction," NATO Undersea Research Centre Report No. SR-370, April 2003.

Harrison, C. H. (**2005a**). "Experimental determination of seabed scattering law and environmental parameters from reverberation," NATO Undersea Research Centre Report No. SR-403, February 2005.

Harrison, C. H. (**2005b**). "Closed form bistatic reverberation and target echoes with variable bathymetry and sound speed," IEEE J. Ocean. Eng. **30**, 660–675.

Harrison, C. H. (**2005c**). "Fast bistatic signal-to-reverberation-ratio calculation," J. Comput. Acoust. **13**, 317–340.

Heaney, K. D. (**2003**). "Active rapid geo-acoustic characterization," J. Acoust. Soc. Am. **113**, 2191.

Holland, C. W. (**2005**). "On errors in estimating seabed scattering strength from long-range reverberation (L)," J. Acoust. Soc. Am. **118**, 2787–2790.

Holland, C. W. (**2006**). "Mapping seabed variability: Rapid surveying of coastal regions," J. Acoust. Soc. Am. **119**, 1373–1387.

Kamminga, S. D., Ellis, D. D., and Gerstoft, P. (**1993**). "Extraction of both bottom backscattering strength and reflection loss by inversion of reverberation measurements," J. Acoust. Soc. Am. **94**, 1844.

LePage, K. and Harrison, C. H. (**2003**). "Bistatic reverberation benchmarking exercise: BiStaR versus analytic formulas," J. Acoust. Soc. Am. **113**, 2333.

Makris, N. C. (**1993**). "Imaging ocean-basin reverberation via inversion," J. Acoust. Soc. Am. **94**, 983–993.

Muller, S. H. E., Ainslie, M. A., Boek, W., and Simons, D. G. (**2002**). "Inversion for bottom parameters using low frequency reverberation data," edited by A. Stepnowski, R. Salamon and A. Partyka, in *Proceedings of the 6th European Conference on Underwater Acoustics (ECUA)*, Poland, 24–27 June 2002, pp. 147–152.

Nielsen, P. L. (**2006**). "Simultaneous geoacoustic inversion of propagation and reverberation data," edited by S. M. Jesus and O. C. Rodriguez, in *Proceedings of the 8th European Conference on Underwater Acoustics (ECUA)*, Portugal, 12–15 June 2006, pp. 541–546.

Preston, J. R. (**2000**). "Inversion of reverberation data for rapid environmental assessment," J. Acoust. Soc. Am. **107**, 2772.

Prior, M. K. and Harrison, C. H. (**2004**). "Estimation of seabed reflection loss properties from direct blast pulse shape," J. Acoust. Soc. Am. **116**, 1341–1344.

Rogers, A., Muncill, G., and Neumann, P. (**1998**). "Inversion of bistatic reverberation and scattering measurements for seabottom properties," J. Acoust. Soc. Am. **103**, 3096.

Weston, D. E. (**1971**). "Intensity-range relations in oceanographic acoustics," J. Sound Vib. **18**, 271–287.

Weston, D. E. (**1976**). "Propagation in water with uniform sound velocity but variable-depth lossy bottom," J. Sound Vib. **47**, 473–483.

Weston, D. E. (**1980**). "Acoustic flux formulas for range-dependent ocean ducts," J. Acoust. Soc. Am. **68**, 269–281.

Zhou, J., Guan, D., Shang, E., and Luo, E. (**1982**). "Long-range reverberation and bottom scattering strength in shallow water," Chinese J. Acoust. **1**, 54–63.

Zhou, J., Zhang, X., Luo, E., and Wang, D. (**1993**). "Vertical coherence of long-range reverberation in a Pekeris shallow-water waveguide," J. Acoust. Soc. Am. **93**, 2286.

Zhou, J. and Zhang, X. (**2003**). "Seabottom geoacoustic inversion from reverberation vertical coherence in shallow water," J. Acoust. Soc. Am. **113**, 2204.

Scintillation index of high frequency acoustic signals forward scattered by the ocean surface

Benjamin Cotté,[a] R. Lee Culver, and David L. Bradley
Graduate Program in Acoustics and Applied Research Laboratory, The Pennsylvania State University, P.O. Box 30, State College, Pennsylvania 16804

(Received 13 April 2006; revised 27 September 2006; accepted 3 October 2006)

Ocean measurements of the scintillation index (SI) of surface forward-scattered signals made in August 2002 are presented and compared with a model developed by Yang and McDaniel [Waves in Random Media **1**, 419–439 (1991)]. The acoustic measurements employed continuous wave (CW) pulses and linear frequency modulated (LFM) sweeps with center frequencies of 20 and 40 kHz. Simultaneously, measurements of wind speed, directional surface wave height spectrum, and ocean sound speed profile were made. The sea state was between 0 and 1 during the four days of the experiment, in part because the location is very much in the lee of San Clemente Island. The measured values of SI are found to agree with Yang and McDaniel model predictions, except for measurements with the largest signal bandwidth and/or the narrowest beamwidths, which violate model assumptions of continuous signals and omnidirectional projectors and hydrophones. In addition, the data show that SI decreases with increasing signal bandwidth (or decreasing temporal extent). An extension to the Yang and McDaniel model is developed that accounts for a reduction in signal temporal extent or ocean surface ensonification. The extended model is in qualitative agreement with the measurements, in that SI is predicted to decrease with increasing signal bandwidth. [DOI: 10.1121/1.2382457]

PACS number(s): 43.30.Re, 43.30.Hw [SLB] Pages: 120–131

I. INTRODUCTION

The intensity of a received signal can display temporal variations due to phenomenon such as relative motion between a source and a receiver, fluctuations in the index of refraction along the propagation path, or motion of the ocean medium. The intensity variance divided by the square of the mean intensity, referred to as the *scintillation index* (SI), is widely used in underwater acoustics to quantify signal level variation.

$$\mathrm{SI} = \frac{\langle I^2 \rangle}{\langle I \rangle^2}. \tag{1}$$

This paper presents measurements and theory for the SI of acoustic signals that have been forward scattered by the ocean surface. This work represents the first measurement-model comparison for the SI ocean surface forward-scattered signals.

From a theoretical point of view, there are several models for the SI of surface forward-scattered signals (Macaskill and Kachoyan, 1988; Frankenthal, 1990; Yang and McDaniel, 1991). The Yang and McDaniel (henceforth referred to as YMcD) model is the focus of this paper. It involves a scattering strength or roughness parameter that is proportional to the ratio of root mean square (rms) ocean surface wave height to acoustic wavelength. When the strength/roughness parameter is small, scattering is weak and the SI is close to 0; this is called *unsaturated* scattering by Flatté (1983) in the context of scattering by ocean volume inhomogeneities. When the strength/roughness parameter is large, scattering is strong and the SI approaches 1; this is called *saturated* scattering. The SI can follow two behaviors in the transition region, as shown in Fig. 1 (Colosi and Baggeroer, 2004), which was developed from ocean volume scattering considerations but is (shown to be, in Sec. II) applicable to scattering by the ocean surface. In the diffractive path to saturation, the SI increases monotonically from 0 to 1 in transitioning from unsaturated to saturated scattering. In the geometric path to saturation, the SI reaches a maximum and decreases to 1. In this case, SI values greater than 1 can occur due to constructive and destructive interference between a few, discrete acoustic paths (Flatté, 1983; Colosi and Baggeroer, 2004); this phenomenon is referred to as *partially saturated* scattering. The behavior in the partially saturated region is controlled by scatterer size or (for the purposes of this work) a diffraction parameter related to the size of the first Fresnel zone relative to the correlation length of surface height fluctuations. Both the strength/roughness and size/diffraction parameters will be discussed in detail below.

From an experimental point of view, there are few published measurements of SI for ocean surface forward-scattered signals. Several papers published in the 1960s and 1970s dealt with second order moments of the scattered pressure (Scrimger, 1961; Melton and Horton, 1970; Nichols and Senko, 1974). More recently, Stroud, Marston, and Williams (1997) published theoretical predictions and experimental data for fourth order (and higher) moments of the scattered pressure. The measurements were conducted in a tank using a rigid one-dimensional surface with a Gaussian roughness spectrum. Their experiment showed that SI falls off as pulse

[a]Present address: Ecole Centrale de Lyon, Laboratoire de Mécanique des Fluides et d'Acoustique, UMR CNRS 5509, BP 163, 69131 Ecully Cedex, France. Electronic address: benjamin.cotte@ec-lyon.fr

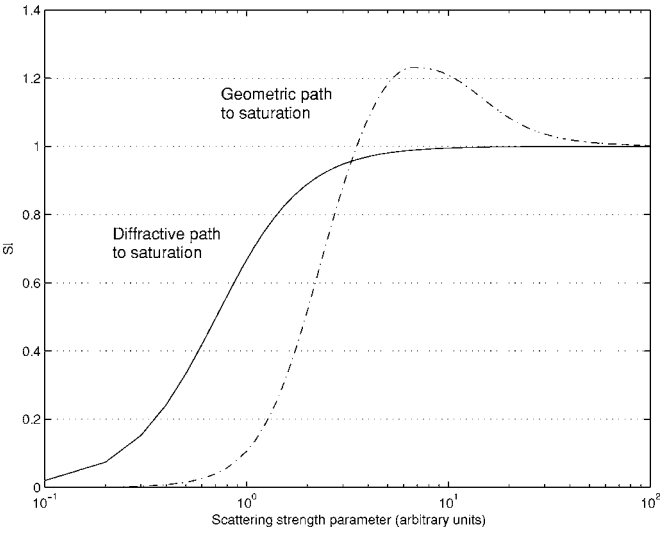


FIG. 1. Sketch of SI vs strength parameter illustrating the two possible paths to saturation. After Colosi and Baggeroer (2004, Fig. 1).

length is reduced in the saturated region, which is consistent with the measurements presented in this report.

This paper presents measurements of the SI of surface forward-scattered signals made in August 2002 near the eastern shore of San Clemente Island, California, and compares them with theory developed by Yang and McDaniel (1991) (YMcD). The acoustic measurements employed 1.0 and 0 .14 ms continuous wave (CW) pulses and 8 ms linear frequency modulated (LFM) sweeps with center frequencies of 20 and 40 kHz; LFM signal bandwidths range from 1 to 22 kHz. Source-receiver separation was approximately 700 m. In most cases, source and receiver depths were such that the direct and surface forward-reflected arrivals were separated in time. Environmental measurements of the wind speed, the directional surface wave height spectrum, and the ocean sound speed profile accompanied the acoustic measurements. The sea state was between 0 and 1 during the experiment, in part because the location is very much in the lee of San Clemente Island.

As will be discussed in Sec. II, the YMcD model is based upon the Kirchhoff approximation for scattering from the rough surface. Thus the issue of the validity of the Kirchhoff approximation for the experimental conditions associated with the measurements presented in this paper has to be considered. Thorsos (1988; 1990) has investigated the conditions under which the Kirchhoff approximation is valid and found it accurate near the specular direction, even for quite small incident angles (as low as 10 deg). Both single spatial scale Gaussian and multiscale Pierson-Moskowitz surface wave height distributions were considered. Away from specular, but still far from very shallow grazing angles where shadowing or multiple scatter can occur, the validity of the Kirchhoff approximation was found to depend upon the ratio of the surface correlation length l to the acoustic wavelength λ. This is the range of angles which dominate the scintillation of forward-scattered signals, and here the Kirchhoff approximation is found to be valid as long as $\frac{l}{\lambda} > 1$. For the data presented here, the grazing angles are 10 to 20 deg; values of l range from 0.6 to 4.9 m and λ is 0.0325 or

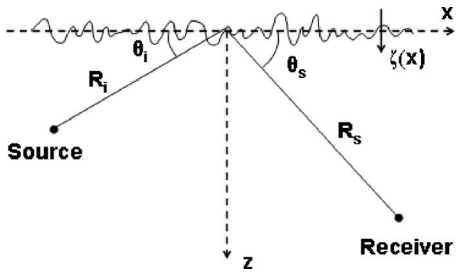


FIG. 2. Geometry for the YMcD model for scintillation index (SI).

0.075 m (see Sec. IV), so that $\frac{l}{\lambda}$ is clearly much greater than 1. The validity of the Kirchhoff approximation is therefore valid and it is appropriate to compare the YMcD model with the measurements.

The experimental conditions did not satisfy two important assumptions of the YMcD model, namely, that the signal was continuous and the projector and hydrophones were omnidirectional. The measurements were made with pulses with finite bandwidth (larger than 1 kHz), and half of the acoustic projectors and all of the receive hydrophones were directional. Not withstanding these shortcomings, many SI measurements agree well with the YMcD model. The SI measurements that did not agree with the model were made using signals with the largest bandwidth and/or projectors/hydrophones that ensonified a significantly reduced area of the ocean surface, thus violating model assumptions to the greatest degree. Increasing signal bandwidth results in increased temporal resolution at the output of the matched filter, and equivalently in reduced signal temporal extent. The YMcD development has been modified to handle pulses with finite bandwidth and directional beams by limiting the ensonified area of the ocean surface, hereafter referred to as the XYMcD model. The XYMcD model predicts a reduction in SI when the signal temporal extent is shortened, which is in agreement with the measurements.

The Yang and McDaniel (YMcD) model and its extension for limited ensonification of the ocean surface (XYMcD) are described in Secs. II and III, respectively. In Sec. IV acoustic and environmental measurements are presented, and the method of estimating the surface wave height spectrum and the correlation function of surface wave height fluctuations is detailed. SI measurements are compared to YMcD and XYMcD model predictions in Secs. V and VI, and a summary is given in Sec. VII.

II. THE YANG AND MCDANIEL (YMcD) MODEL FOR SCINTILLATION INDEX (SI)

The Yang and McDaniel (YMcD) model (Yang and McDaniel, 1991) utilizes the two-dimensional geometry shown in Fig. 2 and assumes an omnidirectional source and receiver, a CW signal, and validity of the Kirchhoff approximation. Derivation of the scattered pressure for the Kirchhoff approximation is taken from Clay and Medwin (1977). Using the geometry and notation of Fig. 2 and the assumptions discussed above, the scattered pressure p_s can be written

$$p_s = \int_{-\infty}^{\infty} \exp\left[-i\left(\frac{x^2}{x_f^2} + 2\gamma\zeta(x) \right) \right] dx, \tag{2}$$

where

$$x_f^2 \equiv \frac{2}{k}\left(\frac{\sin^2\theta_i}{R_i} + \frac{\sin^2\theta_s}{R_s} \right)^{-1}, \tag{3}$$

$$\gamma \equiv -\frac{k}{2}(\sin\theta_i + \sin\theta_s). \tag{4}$$

In writing Eq. (2), the terms describing time and range dependence have been omitted because they are not important to this derivation. Note that $\zeta(x)$ denotes surface elevation. The parameter x_f can be interpreted as $\frac{1}{\sqrt{\pi}}$ times the radius of the first Fresnel zone (Clay and Medwin, 1977, p. 50), projected onto the x axis. The $\exp\left(-i\frac{x^2}{x_f^2}\right)$ term thus provides sinusoidal dependence upon x as x moves through the Fresnel zones. Also, k is the wave number and γ is the mean vertical component of the wave number of the incident and scattered waves, so that the term $2\gamma\zeta(x)$ parametrizes the effect of the rough surface to alter the phase of the received signal (Clay and Medwin, 1977, p. 340). Equation (2) is the starting point for calculating the mean intensity $\langle I\rangle = \langle |p_s|^2\rangle$ and the mean squared intensity $\langle I^2\rangle = \langle |p_s|^4\rangle$. After some manipulations, Yang and McDaniel obtain the following expression for SI:

$$\begin{aligned} \mathrm{SI} &= \frac{\langle I^2\rangle}{\langle I\rangle^2} - 1 \\ &= \frac{4}{\pi}\,\mathrm{Re}\left[\int_0^{\infty} \left(\int_0^{P} \exp(-iPS)\exp[-\Phi^2 H(P,S)]\,dS \right) dP \right] \\ &\quad - 1, \end{aligned} \tag{5}$$

where Re means "real part of," and P and S are functions of position on the surface defined for convenience. Equation (5) is the same as Eq. (32) of YMcD. It is similar to Eq. (9) of Yang, Fennemore, and McDaniel (1992), in which the YMcD theory is applied to a two-dimensional (2D) ocean surface; note that for consistency the factor in front of the integral should be $\frac{4}{\pi}$ in the latter paper. The factor H depends on the correlation function of a nondimensional parameter $C(X) = C_\zeta\left(\frac{Xx_f}{\sqrt{2}}\right) = C_\zeta(Xl\sqrt{\Lambda})$

$$H(P,S) = 2 - 2C(P) - 2C(S) + C(P+S) + C(P-S). \tag{6}$$

Φ and Λ are the strength/roughness and size/diffraction parameters that control SI behavior. They are defined as follows:

$$\Phi^2 \equiv 4h^2\gamma^2, \tag{7}$$

$$\Lambda \equiv \frac{x_f^2}{2l^2}, \tag{8}$$

where $h = \langle \zeta^2\rangle^{1/2}$ is the rms ocean surface wave height and l is the scale size for the ocean surface correlation function.

To understand what Φ and Λ physically mean, consider the case of specular reflection ($\theta_i = \theta_s = \theta$ in Fig. 2). Then Eq. (7) becomes the Rayleigh roughness parameter $\Phi = 4\pi h \sin\theta/\lambda$, where λ is the acoustic wavelength, as defined, for instance, by Melton and Horton (1970) and Dahl (1999).[1] When Φ is large ($\Phi \gg 1$), the surface appears very rough and there are many acoustic paths to the receiver that are, on average, uncorrelated. SI approaches 1, and the scattering is referred to as *saturated*. When Φ is small ($\Phi \ll 1$), the surface appears smooth and approaches only one acoustic path. SI approaches 0, and the scattering is referred to as *unsaturated*. This behavior can be seen on the Λ-Φ diagram shown in Fig. 3 (Flatté, 1983). The axes in Fig. 3 are $\Phi_{\mathrm{Flatte}} = k_0\langle\mu^2\rangle^{1/2}\sqrt{RL_P}$ and $\Lambda_{\mathrm{Flatte}} = R/6L^2k_0$, where $\langle\mu^2\rangle^{1/2}$ is the rms sound speed fluctuation, R is source-receiver separation, L and L_P are the sound speed inhomogeneity correlation lengths in the vertical direction and along the path, respectively, and k_0 is the mean acoustic wave number (Flatté *et al.* 1979). There is strong physical correspondence between Φ_{Flatte} and $\Lambda_{\mathrm{Flatte}}$, which were derived for ocean volume scattering, and Φ and Λ defined in Eqs. (7) and (8) for surface-forward scattering. The rms wave height and correlation length of the ocean surface in the YMcD theory correspond to rms sound speed fluctuation and the correlation length of the sound speed fluctuations, respectively, in Flatté's theory. In both theories, the diffraction parameter Λ is proportional to the square of the Fresnel radius divided by the square of a correlation length. Both definitions of the strength parameter include the product of a wave number and rms fluctuation, but in Flatté's theory an additional range times along-path correlation length term is present. Scattering in the ocean volume occurs along the path, and $\sqrt{RL_P}$ is a Fresnel term with L_P replacing λ.

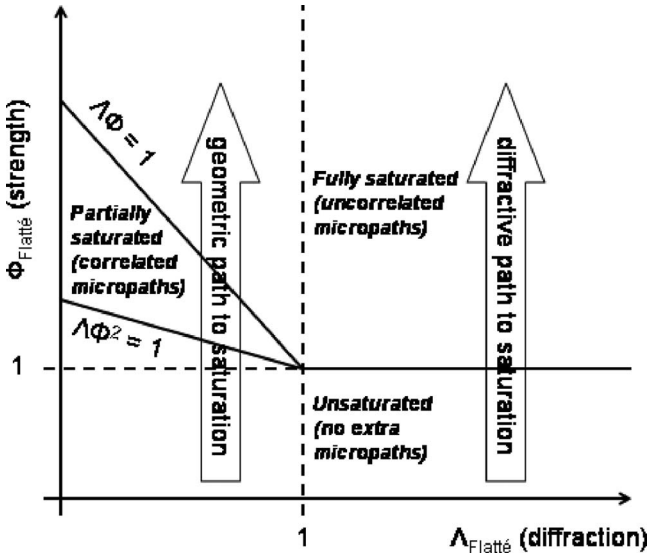


FIG. 3. Λ-Φ diagram representing the different scattering regimes with respect to the size/diffraction parameter $\Lambda_{\mathrm{Flatte}}$ and the strength/roughness parameter Φ_{Flatte}. After Flatté (1983, Fig. 12).

Λ plays an important role in the transition region between the unsaturated and saturated regions in both volume and surface scattering. As stated above, x_f can be interpreted as $\frac{1}{\sqrt{\pi}}$ times the radius of the first Fresnel zone projected onto the x axis. A spherical wave front can be divided into half-wavelength annular zones called Fresnel zones; these zones

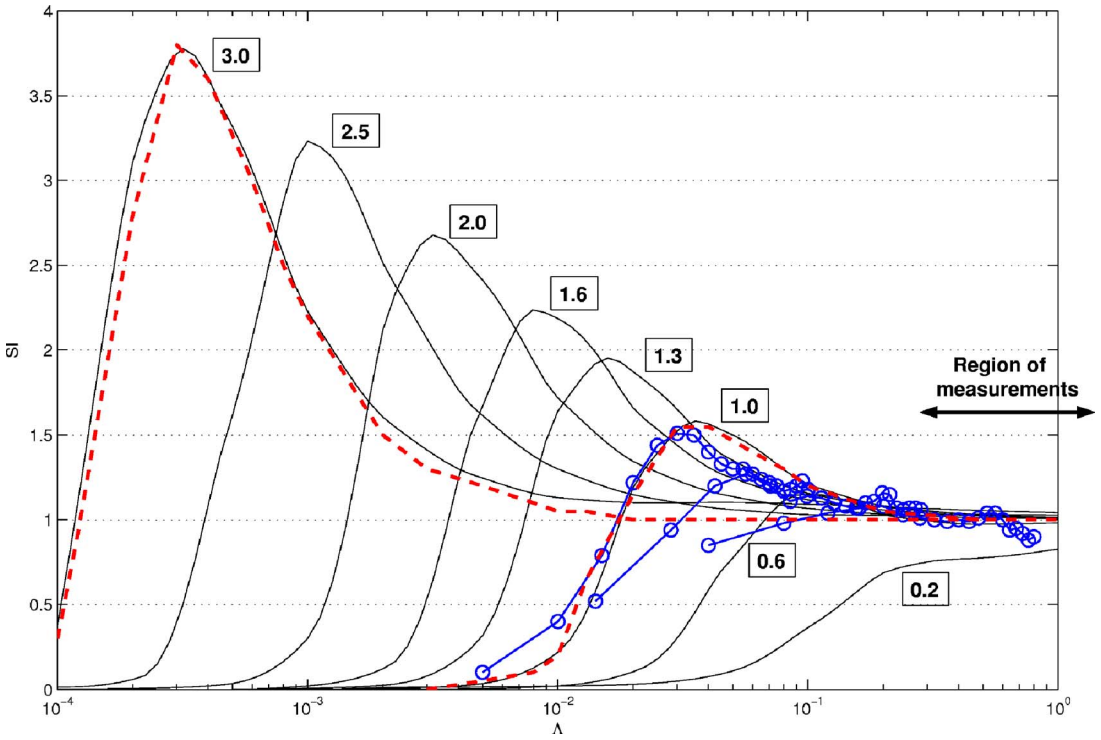


FIG. 4. (Color online) Comparison of three models for SI of ocean forward-scattered signals. A Gaussian surface correlation spectrum is used, and results are plotted and scaled with respect to the Yang and McDaniel Λ and Φ parameters. The Yang and McDaniel predictions are given for $\log_{10}(\Phi)=0.2$, 0.6, 1.0, 1.3, 1.6, 2.0, 2.5, and 3.0 (solid lines), as given in the text boxes. Macaskill and Kachoyan predictions are given for $\log_{10}(\Phi)=0.70$, 0.85, and 1.0 (open circles), and Frankenthal predictions are given for $\log_{10}(\Phi)=1.0$ and 3.0 (dashed lines).

define phase relationships, since points from adjacent zones are out-of-phase from each other. The parameter l is a measure of the area over which the surface elevation is correlated. Thus for surface scattering, Λ compares the size of the first Fresnel zone to the size of the correlated region of the ocean surface, and is therefore called size parameter. Λ provides a similar comparison for volume scattering, although it is called a diffraction parameter by Flatté (1983) since for small values of Λ (roughly $\Lambda<1$), the geometric approximation is valid, while for large values of Λ (roughly $\Lambda>1$), diffraction occurs. The corresponding geometric and diffractive paths to saturation are represented by two block arrows in Fig. 3, which are related to the SI behavior shown in Fig. 1. For surface scattering, in the geometric path to saturation ($x_f<l$), the surface elevation is correlated over several Fresnel zones. This leads to constructive and destructive interference between acoustic arrivals (also referred to as *multipath correlation* by Colosi and Baggeroer (2004)) for a range of Φ values that cause large variations and SI values greater than 1. In the diffractive path to saturation ($x_f>l$), acoustic arrivals from Fresnel zones other than the first zone are uncorrelated and SI transitions directly from unsaturated scattering to fully saturated scattering.

YMcD predictions for dependence of SI on Λ for various values of Φ using a Gaussian surface correlation spectrum are shown in Fig. 4. Note that this representation is different from the one used in Fig. 1. Curves in Fig. 1 are equivalent to vertical paths in the Λ-Φ diagram (see Fig. 3), while curves in Fig. 4 are equivalent to horizontal paths in the Λ-Φ diagram. As will be discussed in Sec. V, the range of Λ values over which measured SI can be compared with the YMcD model is 0.3 to 1.9, and this range is indicated in the figure. Two other models are compared in Fig. 4 using Yang and McDaniel Λ and Φ parameters. The predictions by Macaskill and Kachoyan (1988) are obtained by numerical simulation based upon inversion of the full integral equation describing surface scattering, for which the Kirchhoff approximation corresponds to the first iteration. As shown there and by Thorsos (1988), the Kirchhoff approximation compares well with the full integral equation for scattered angles near the specular, for small to moderate rms surface waveheight, and for a Gaussian surface waveheight spectrum. Reflection coefficients are found for different realizations of a random sea surface, and the SI is calculated by ensemble averaging over the reflection coefficients. No size parameter is explicitly defined. The model by Frankenthal (1990) uses a phase screen model for the surface interaction that can be applied to a refracting as well as a reflecting scattering surface. Numerical results are presented for large scattering strength parameters and for normal incidence. Frankenthal's definition of Λ is the same as Yang and McDaniel's for normal incidence $\left(\theta_i=\theta_s=\frac{\pi}{2}\right)$. It can be seen that his model fits Yang and McDaniel's very well. To calculate Λ using Macaskill and Kachoyan parameters, a rms wave height h (or equivalently a correlation length l) needs to be chosen. h is assumed to be 0.01 m to obtain the best fit with the other two models for $\log_{10}(\Phi)=1.0$. Doing so this model fits the Yang and McDaniel predictions relatively well. Note that all three models predict unsaturated scattering for the smallest values of Φ. They also predict that the peak value of SI above 1,

characteristic of the partially saturated region, increases and shifts to lower Λ as Φ increases. This behavior is consistent with Fig. 3.

III. YMcD MODEL EXTENSION FOR FINITE SIGNAL BANDWIDTH AND LIMITED ENSONIFICATION OF THE OCEAN SURFACE (XYMcD)

The conditions of the SI measurements reported in this paper violate the assumptions of the YMcD model in two respects. First, pulses with finite bandwidth were transmitted rather than continuous signals, and second, directional acoustic projectors and hydrophones were used. A continuous signal has broad temporal extent or equivalently, very small effective bandwidth. The signals used in the 2002 field experiment have small temporal extent or non-zero effective bandwidth, which means that at any given time only a small part of the ocean surface will contribute to the received signal. The size of the area of the ocean surface that contributes to the received signal is limited by the signal characteristics because matched-filtering provides temporal extent equal to the inverse signal bandwidth. For instance, the extent of the ocean surface contributing to the received signal for a 1 ms pure tone pulse and an 8 ms, 1 kHz LFM is approximately $\frac{c}{\beta}=1.5$ m, where $c=1500$ m/s is the speed of sound in the water and β is the effective bandwidth. Similarly, hydrophone directionality resulted in incomplete illumination of the surface for some projector-receiver pairs. In some cases the specular point was not strongly illuminated. Many measured SI values agreed with the YMcD model in spite of these differences, but the SI measurements with the shortest signal temporal extent (or largest bandwidth) and narrowest beams did not agree. Therefore, the YMcD model has been modified to account for finite signal bandwidth and hydrophone beam width by introducing an illumination factor $\mathfrak{I}(x)$ (Clay and Medwin, 1977). The illumination factor $\mathfrak{I}(x)$ is limited to values between 0 and 1 and indicates the combined projector-hydrophone pattern response for the point x on the surface ($x=0$ is the specular point, as may be seen in Fig. 2). $\mathfrak{I}(x)=1$ corresponds to a completely ensonified surface (omnidirectional projectors and hydrophones) and $\mathfrak{I}(x)=0$ corresponds to no ensonification of the surface.

Introducing the illumination factor into the equation for the scattered pressure, Eq. (2) becomes

$$p_s=\int_{-\infty}^{\infty}\mathfrak{I}(x)\exp\left[-i\left(\frac{x^2}{x_f^2}+2\gamma\zeta(x)\right)\right]dx. \tag{9}$$

Equation (9) has been evaluated by Clay and Medwin (1977) for the case of a Gaussian beam. The mean intensity is given by

$$\langle I\rangle = x_f^2\int_0^{\infty}\left[\int_{-\infty}^{\infty} IL'(\omega,\xi)\cos(\omega\xi)d\omega\right] \times\exp[-\Phi^2(1-C(\xi))]d\xi, \tag{10}$$

where

$$IL'(\omega,\xi)=\mathfrak{I}\left(\frac{x_f}{\sqrt{2}}\omega+\frac{x_f}{2\sqrt{2}}\xi\right)\mathfrak{I}\left(\frac{x_f}{\sqrt{2}}\omega-\frac{x_f}{2\sqrt{2}}\xi\right). \tag{11}$$

A quadruple integral must be evaluated in order to calculate the mean squared intensity

$$\langle I^2\rangle=\frac{x_f^4}{4}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} IL(t,P,S,q) \times\exp(-itq)\exp[\Phi^2H_1(P,S,q)]dtdq\right] \times\exp(-iPS)\exp[-\Phi^2H_2(P,S)]dSdP, \tag{12}$$

where t and q are temporary variables of integration with no physical significance

$$H_1(P,S,q)=C\left(P+\frac{q}{2}\right)+C\left(P-\frac{q}{2}\right)+C\left(S+\frac{q}{2}\right)+C\left(S-\frac{q}{2}\right),$$

$$H_2(P,S)=2+2C(P+S)+C(S-P), \tag{13}$$

and

$$\begin{aligned}IL(t,P,S,q)=&\,\mathfrak{I}\left(\frac{x_f}{\sqrt{2}}t+\frac{x_f}{2\sqrt{2}}P+\frac{x_f}{2\sqrt{2}}S+\frac{x_f}{4\sqrt{2}}q\right)\\&+\mathfrak{I}\left(\frac{x_f}{\sqrt{2}}t-\frac{x_f}{2\sqrt{2}}P+\frac{x_f}{2\sqrt{2}}S-\frac{x_f}{4\sqrt{2}}q\right)\\&+\mathfrak{I}\left(\frac{x_f}{\sqrt{2}}t-\frac{x_f}{2\sqrt{2}}P-\frac{x_f}{2\sqrt{2}}S+\frac{x_f}{4\sqrt{2}}q\right)\\&+\mathfrak{I}\left(\frac{x_f}{\sqrt{2}}t+\frac{x_f}{2\sqrt{2}}P-\frac{x_f}{2\sqrt{2}}S-\frac{x_f}{4\sqrt{2}}q\right).\end{aligned} \tag{14}$$

The effect of the illumination factor on SI has been evaluated using the XYMcD model and an illumination factor defined by

$$\mathfrak{I}_{x_0}\equiv\begin{cases}\dfrac{1+\cos\left(\dfrac{\pi x}{2x_0}\right)}{2} & \text{for } x\le 2x_0,\\ 0 & \text{elsewhere,}\end{cases} \tag{15}$$

where x_0 parametrizes the width of the ensonified area. In Fig. 5, SI is plotted versus x_0 for different values of $\log_{10}(\Phi)$. It can be seen that SI decreases as the size of the ensonified area decreases. As mentioned earlier, the range of Λ values used for comparison with measured SI is 0.3 to 1.9. Results are shown for Λ equal to 0.5 only in Fig. 5 since extended model predictions depend very little on the size parameter Λ for Λ between 0.3 and 1.9.

IV. OCEAN MEASUREMENTS

Acoustic and environmental measurements were made over a four-day period during August 2002 about 2.5 km east of San Clemente Island, California (32° 38.2′ N, 117° 57.4′ W). Water depth was approximately 500 m. The acoustic

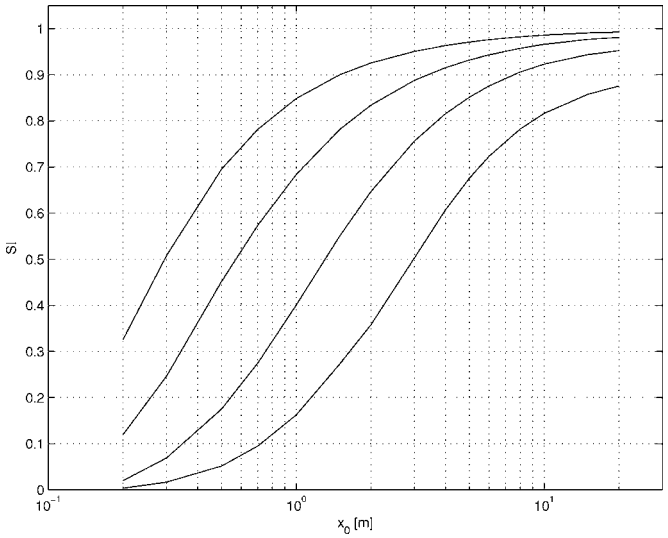


FIG. 5. XYMcD model predictions calculated using the illumination factor $\mathfrak{I}_{x_0}$ and the swell enhanced Plant spectrum described in Sec. IV C. $\Lambda=0.5$ for all curves and $\log_{10}(\Phi)=0.4$, 0.6, 0.8, and 1.0 from the lowest to the highest curve.

measurement system and the signals used to measure SI are described in Sec. IV A. Environmental measurements are identified in Sec. IV B. The surface waveheight spectrum, the correlation function of surface height fluctuations, the rms waveheight, and the correlation length are estimated in Sec. IV C.

A. Acoustic measurements

The acoustic measurement system consists of four projectors deployed from a moored buoy and a set of six receive hydrophones deployed from the Motor Vessel (M/V) Acoustic Explorer as shown in Fig. 6. Projectors P_1 and P_3 are ITC 6084 hydrophones, with vertical half-beam widths of 13.3° at 20 kHz and 6.4° at 40 kHz. Projectors P_2 and P_4 are ITC 1001 projectors, which are basically omnidirectional. All receive hydrophones are ITC 6080C, which have vertical half-beamwidths of 42.0° at 20 kHz and 20.6° at 40 kHz. Hydrophone H_3 malfunctioned during the experiment and recorded no data. Seven acoustic data sets were recorded during the experiment. Data for projector P_4 were only available for

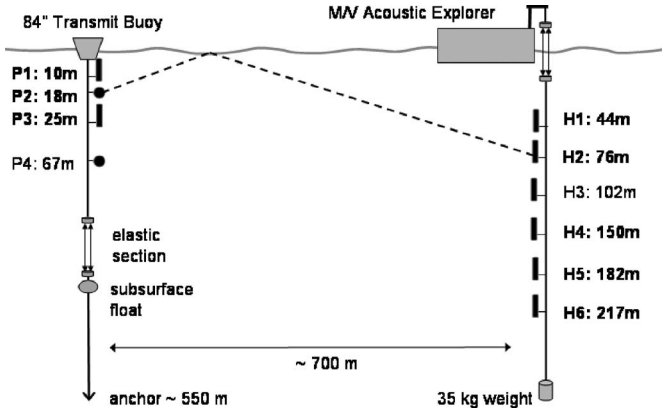


FIG. 6. Acoustic measurement system configuration. The specular ray for *P2-H2* (projector 2 - hydrophone 2) is shown.

TABLE I. Acoustic signal characteristics.

Name	Type	Center frequency (kHz)	Bandwidth (kHz)	Duration (ms)
Short CW 20	CW	20	7	0.14
Long CW 20	CW	20	1	1
Short CW 40	CW	40	7	0.14
Long CW 40	CW	40	1	1
LFM 20-1	LFM	20	1	8
LFM 20-7	LFM	20	7	8
LFM 20-13	LFM	20	13	8
LFM 20-22	LFM	20	22	8
LFM 40-1	LFM	40	1	8
LFM 40-7	LFM	40	7	8
LFM 40-13	LFM	40	13	8
LFM 40-22	LFM	40	22	8

LFM signal duration is 10 ms for projector P3 only.

two of the seven measurement periods and are not used in this paper. Also, direct and surface bounce paths could not be separated for the shallowest projectors and receivers and those data are not used in this paper.

For each data set, the 12 signals identified in Table I were transmitted by each projector. Center frequencies are 20 and 40 kHz. CW signals with duration 0.14 and 1 ms are referred to as "short" and "long" pulses, respectively. LFM signals have bandwidths ranging from 1 to 22 kHz. For each signal and each projector, 300 pings were transmitted using a 0.1 s repetition rate. Received signals from all hydrophones were digitized and recorded simultaneously using a sample rate of 312.5 kHz. Subsequently, received signals were matched filtered and the peak of the filter output located. The SI was calculated over the 300 peak values of the surface scattered path in the match-filtered output.

B. Environmental measurements

Sound speed in the water column, wind speed, and directional surface waveheight spectrum were measured during the four-day period of the experiment. The sound speed profile was estimated approximately once a day using a cast conductivity-temperature-depth (CTD) profiler dropped down to a depth of approximately 300 m. As shown in Fig. 7, a downward refracting profile is found; it is very similar from day to day. Ray tracing is used to determine whether the specular point is strongly illuminated for each projector-receiver pair. The time-averaged directional wave height spectrum was measured every 30 min using a Triaxys Directional wave rider buoy. However, sea surface waveheight during the experiment was too small to be measured accurately by the wave rider buoy[2] and those data are not used in this paper. Wind speed was measured using a Handar ultrasonic anemometer every 10 s, except that there are no wind speed data for one of the four days of the experiment. Measured wind speed ranges from 6 to 12 m/s. Because of the relatively steady conditions, the wind speed is taken to be 8 ± 2 m/s during the experiment. The rms waveheight is very low considering the wind speed measured during the experi-

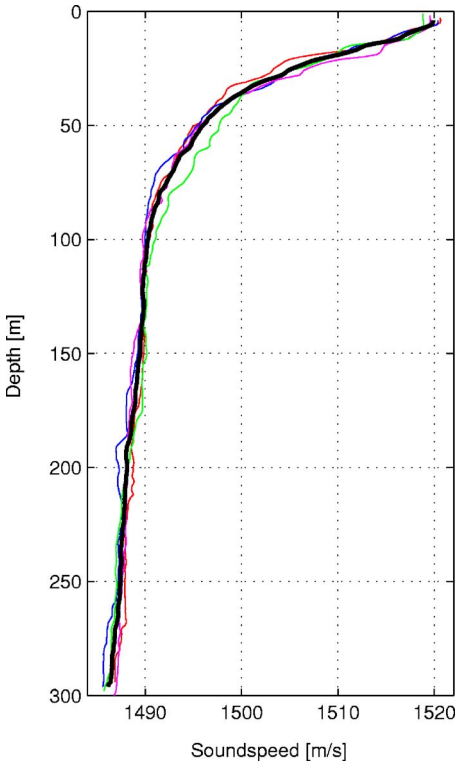


FIG. 7. (Color online) Sound speed profiles measured between 15 and 18 August 2002 at the experiment site (thin lines). The thick line is the averaged sound speed profile over the four days.

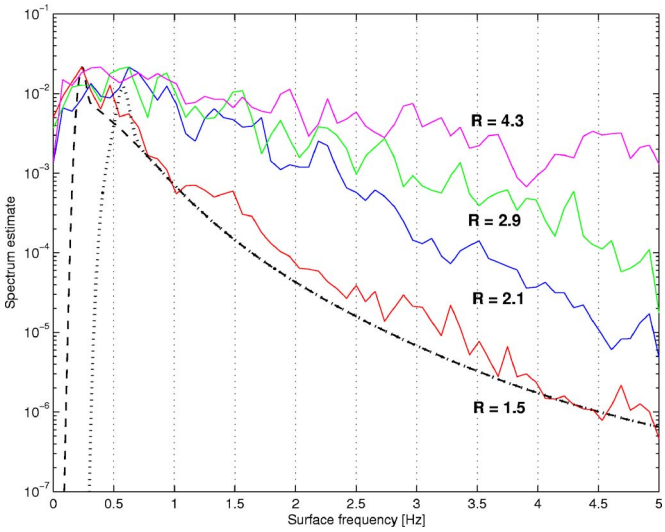


FIG. 8. (Color online) Spectra of peak long pulse acoustic intensity values (solid lines) for different values of the Rayleigh roughness parameter R (as labeled). Each spectrum is an average over three measurement periods. The dotted line corresponds to Plant's model calculated using a wind speed of 8 m/s and a fetch of 2500 m. The dashed line is Plant's model extended to include a swell component at 0.23 Hz. All of the acoustic spectra have been scaled so that their maxima equal the swell enhanced Plant spectrum maximum.

ment; this is because of the close proximity of the experiment to San Clemente Island and the attendant reduced fetch (McDaniel and McCammon, 1987).

C. The surface waveheight spectrum and correlation function

As stated above, surface conditions during the experiment were too benign to be measured by the Triaxys wave rider buoy. However, the correlation function of surface height fluctuations is an important input parameter to the models described in Secs. II and III. Accordingly, a model was employed to estimate the surface waveheight spectrum, the correlation function and the correlation length. Following Dahl (1999), the isotropic correlation function $C_\zeta(\rho)$ can be computed via the Bessel transform relation

$$C_\zeta(\rho) = \frac{\int_0^\infty KJ_0(K\rho)F(K)dK}{\int_0^\infty KF(K)dK}, \tag{16}$$

where K is the magnitude of the ocean surface wave number, $F(K)$ is the directionally averaged surface height wavenumber spectrum, and ρ is the horizontal distance between two points on the surface. Dahl showed that at high acoustic frequencies (approximately 10 kHz and more), directionality of the surface waveheight spectrum does not play an important role since scattering is dominated by capillary waves (Dahl, 1999; Dahl, 2004).

The surface waveheight spectrum has been modelled using Plant's D spectrum (Plant, 2002). The model input parameters are wind speed and fetch. Because the experiment location was about 2500 m east of San Clemente Island, and the wind was out of the west northwest, the fetch is limited to approximately 2500 m. Diffraction around the north-east tip of the island could cause the effective fetch to be higher. On the other hand, because of shadowing by hills on the island, the effective fetch could be lower than 2500 m. For this analysis, the fetch is taken to be 2500 ± 1500 m.

Use of Plant's model with a wind speed of 8 m/s and a fetch of 2500 m produces the waveheight spectrum shown in Fig. 8. The spectra of acoustic intensity fluctuations of selected measurements are also shown in Fig. 8 for comparison. They correspond to Rayleigh roughness parameter $R = 2kh \sin(\theta)$ [see Eq. (7) in the case of specular reflection ($\theta_i = \theta_s = \theta$)] of 1.5 to 4.3, all normalized to the same maximum value. (The variation in R is due to change in grazing angle and acoustic frequency.) A swell enhanced Plant spectrum (explained below) is also plotted in Fig. 8.

It can be observed in Fig. 8 that the acoustic spectrum for the smallest roughness parameter follows Plant's model well above 0.5 Hz. However, as the roughness parameter increases, higher frequency energy increases and a flattening of the acoustic spectrum is observed. As a result, the fit with Plant's model degrades. The spectral broadening is related to the change in the probability density function (pdf) of the peak pressure at the receiver, as described by Clay and Medwin (1977). For small roughness parameter, the scattered pressure follows a Gaussian distribution, and the field is dominated by the coherent contribution corresponding to specular reflection. For the large roughness parameter, the scattered pressure follows a Rayleigh distribution; there are a large number of randomly phased contributions (contribu-

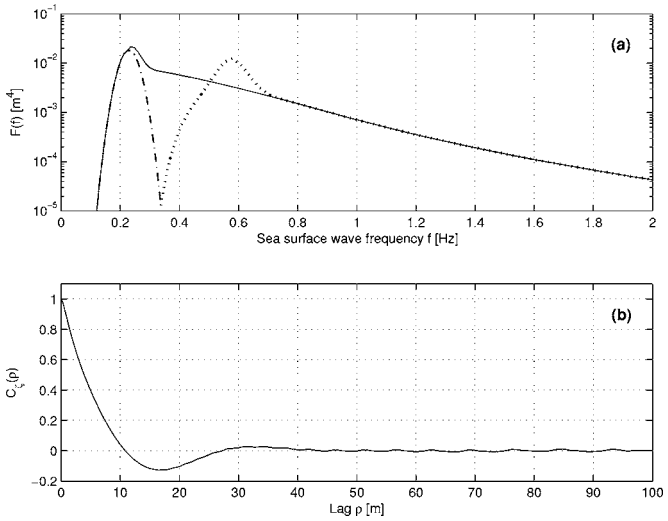


FIG. 9. (a) Directionally averaged wave-number spectra for Plant's model with a wind speed of 8 m/s and a fetch of 2500 m (dotted line), for the Gaussian contribution due to swell centered at 0.23 Hz (dash-dotted line), and for the swell enhanced Plant spectrum (solid line). (b) Correlation function of surface height fluctuations calculated using the swell enhanced Plant spectrum.

tions from many random facets of the ocean surface), and the incoherent component of the field dominates. This broadening of the acoustic spectra for large roughness parameter was observed experimentally by Scrimger (1961) and Brown (1969). The result is that the spectrum of the acoustic pressure (or intensity) fluctuations closely follows the ocean surface fluctuation spectrum for the small roughness parameter, while for larger roughness parameter, the acoustic fluctuation spectra contain high frequency energy that is not present in the surface wave height spectrum.

In general, the peaks of the acoustic spectra in Fig. 8 occur at frequencies below the 0.6 Hz peak of the Plant model spectrum. The acoustic spectrum with the smallest roughness parameter has a peak at 0.23 Hz (with a frequency resolution of 0.04 Hz). This peak could be due to swell from distant storms, which is not predicted by Plant's model, but which could have been diffracted around the northeast tip of San Clemente Island. The peak is modeled as a Gaussian spectrum centered at 0.23 Hz. Taking into account the peak in the acoustic spectra, a new wave-number spectrum is constructed by combining the Gaussian spectrum and Plant's D spectrum extrapolated down to 0.23 Hz. Figure 9(a) illustrates the construction of the swell enhanced Plant spectrum.

Surface waveheight frequency spectra are used to calculate the wave-number spectra, using the dispersion relation

$$(2\pi f)^2 = gK, \tag{17}$$

where f is the surface wave frequency and $g=9.8$ m/s^2 is the gravitational constant. A correlation function calculated using Eq. (16) and the extended waveheight spectrum is plotted in Fig. 9(b). The rms wave height is obtained from the wave spectrum $F(K)$ using

TABLE II. Rms wave height h and correlation length l for Plant's D spectrum, for the swell enhanced Plant spectrum shown in Fig. 9(a), and for two extended spectra illustrating the effect of uncertainties on wind speed, fetch and swell peak frequency.

Type of spectrum	Wind speed (m/s)	Fetch (m)	Peak frequency (Hz)	h (cm)	l (m)
Plant's D spectrum	8	2500	N/A	4.0	0.6
Swell enhanced Plant spectrum	8	2500	0.23	4.4	3.2
"Lowest-energy" spectrum	6	1000	0.27	2.1	2.2
"Highest-energy" spectrum	10	4000	0.19	7.1	4.9

$$h = \sqrt{\langle \zeta^2 \rangle} = \left(\int_0^\infty KF(K)dK \right)^{1/2}. \tag{18}$$

Yang, Fennemore, and McDaniel (1992) define the correlation length as the normalized first moment of the correlation function

$$l = \frac{\int_0^\infty \rho C_\zeta(\rho)d\rho}{\int_0^\infty C_\zeta(\rho)d\rho}. \tag{19}$$

Equation (19) is used in this paper for comparison with the YMcD model. The lag at which the correlation function crosses zero is used as the upper limit for the integrals in Eq. (19) because the most significant part of the correlation function is the region of small lags, and integrating to infinity can yield small or even negative correlation lengths, which does not make sense physically. The rms waveheight h and the correlation length l for Plant's D spectrum and the extended spectrum are given in Table II. The low-frequency extension is seen to have a small effect on h but a large effect on l, which is about five times larger.

To understand the effect of uncertainties in wind speed, fetch, and peak frequency, h and l are also calculated for two extreme cases. A "lowest-energy" spectrum is calculated using a wind speed of 6 m/s, a fetch of 1000 m, and a peak frequency of 0.27 Hz. Similarly a "highest-energy" spectrum is calculated using a wind speed of 10 m/s, a fetch of 4000 m, and a peak frequency of 0.19 Hz. As shown in Table II, using the highest and lowest energy spectra changes the rms waveheight by about a factor of 2 and the correlation length by about a factor of 1.5.

V. MODEL-DATA COMPARISON

In this section, the YMcD and XYMcD predictions are compared to measured SI for similar strength/rouhgness Φ and size/diffraction Λ parameters. For the measurements, Φ and Λ are calculated using the values of h and l shown in Table II, together with the source-receiver geometry. Model parameter values are given in Table III for the 13 SI measurements that are used in the paper. Data omitted because of

TABLE III. Parameters associated with the SI measurements presented here. See Fig. 6 for definition of paths. Also, θ is incident angle, f_c is center frequency, and Φ and Λ were defined in Eqs. (7) and (8). The model parameter values are calculated with h=4.4 cm and l=3.2 m (parameters of the swell enhanced Plant spectrum given in Table II).

Path	θ (deg)	f_c (kHz)	$\log_{10}(\Phi)$	Λ	Measurement number
P_1-H_4	12.6	20	0.16	1.1	1
		40	0.47	0.5	2
P_1-H_5	15.0	20	0.24	0.6	3
P_1-H_6	17.6	20	0.31	0.4	4
P_2-H_4	13.2	20	0.18	1.6	5
		40	0.49	0.8	6
P_2-H_5	15.6	20	0.26	1.0	7
		40	0.56	0.5	8
P_2-H_6	18.1	20	0.32	0.6	9
		40	0.62	0.3	10
P_3-H_4	13.7	20	0.20	1.9	11
P_3-H_5	16.1	20	0.27	1.2	12
P_3-H_6	18.7	20	0.33	0.8	13

projector and/or hydrophone directionality are P_1-H_5 at 40 kHz, P_1-H_6 at 40 kHz, P_3-H_4 at 40 kHz, P_3-H_5 at 40 kHz, and P_3-H_6 at 40 kHz. It can be seen in Table III that Λ ranges from 0.3 to 1.9 and $\log_{10}(\Phi)$ ranges from 0.16 to 0.62 for this set of measurements.

Measurements of SI from different measurement periods are "noisy." That is, large variations are observed from day to day, but no measurement period has a consistently higher or lower SI (regardless of source/receiver geometry or signal type). This is consistent with the environment remaining relatively constant throughout the experiment. In order to reduce noise in the SI measurements, results from the seven measurements are averaged.

A. Comparison of long pulse SI with YMcD model

Figure 10 shows SI measurements made using the long pulse and projector-hydrophone pairs for which the omnidirectional assumption is not seriously violated. The mean and mean ±1 standard deviation calculated over all seven measurement periods are shown. Data are grouped by size/diffraction parameter Λ equal to 1.9 (measurement numbers 5 and 11), 1.0 (measurement numbers 1, 6, 7, 12, and 13), 0.5 (measurement numbers 2, 3, 4, 8, and 9) and 0.3 (measurement number 10). Three YMcD model curves are plotted in Fig. 10. The solid curve corresponds to the extended spectrum with a wind speed of 8 m/s, a fetch of 2500 m, and a peak frequency of 0.23 Hz. These are the best estimates of the environmental conditions at the time of the experiment. The dashed curves give an idea of the variation in SI prediction due to environmental uncertainties. The upper curves are calculated using the "lowest-energy" spectrum (l=2.2 m), and the lower curves are calculated using the "highest-energy" spectrum (l=4.9 m). Parameters for these two spectra are shown in Table II. It can be seen in Fig. 10 that SI measured using the long pulse and omni directional projector-hydrophones generally follows the YMcD model predictions over the limited range of strength/roughness parameter Φ for which data are available.

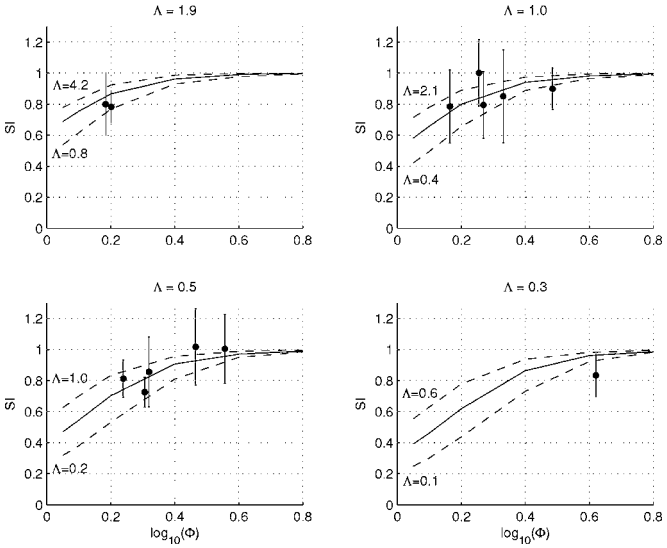


FIG. 10. Mean of SI measured over the seven measurement periods for the long pulse compared with model predictions for three environment conditions. Error bars denote the mean ±1 standard deviation. Parameters for the model predictions are given in Table II. The solid lines utilize the swell enhanced Plant spectrum with a wind speed of 8 m/s, a fetch of 2500 m, and a peak frequency of 0.23 Hz. The upper and lower dashed lines are calculated using the "lowest-energy" and "highest-energy" spectra, respectively.

B. Effect of signal bandwidth on SI

Dependence of SI on signal bandwidth (or temporal extent) is now discussed. Figure 11 shows mean measured SI for the long (1 ms) CW and 1 kHz LFM pulses, both of which possess a 1 kHz bandwidth, plotted using solid dots. Mean SI is shown for short (0.14 ms) CW and 7 kHz LFM pulses, both of which possess a 7 kHz bandwidth, plotted using open circles. CW and LFM data having the same effective bandwidth, or equivalently, temporal extent, are seen to have about the same SI values. Note that LFM signals are much longer (8 or 10 ms) than the CW signals, indicating that SI does not depend on the signal duration in a simple

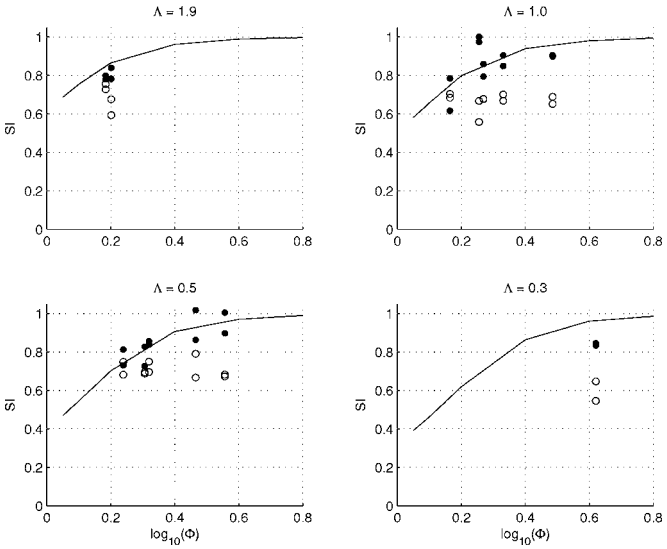


FIG. 11. Comparison of measured SI with the YMcD model for 1 kHz LFM and 1 ms CW signals (solid dots) and 7 kHz LFM and 0.14 ms CW signals (open circles).

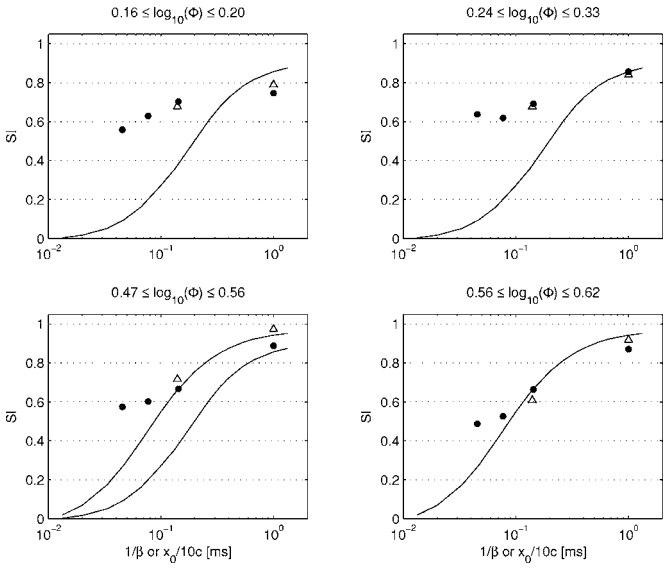


FIG. 12. Measured SI for LFM (solid dots) and CW data (open triangles) vs inverse signal bandwidth $1/\beta$ (equal to temporal extent after match filtering). SI measurements are grouped by values of Φ as indicated, and Λ is equal to 0.5 for all plots. For comparison, SI predicted using the XYMcD model is plotted vs $\frac{x_0}{10c}$ (solid lines), where x_0 is a measure of illumination factor width defined in Eq. (15). XYMcD predictions are calculated using $\log_{10}(\Phi)=0.4$ for the two first Φ groups, $\log_{10}(\Phi)=0.4$ and 0.6 for the third Φ group, and $\log_{10}(\Phi)=0.6$ for the last Φ group.

way. Figure 11 shows that measured SI is generally lower for signals with greater bandwidth (or smaller temporal extent), although for small values of $\log_{10}(\Phi)$ the trend is not as clear. Also the 1 kHz bandwidth (1 ms temporal extent) signals are in better agreement with the YMcD model predictions than the 7 kHz bandwidth (0.14 ms temporal extent) signals, which is consistent with the former being closer to the CW signal that is assumed by the YMcD model.

C. Comparison with the XYMcD model

Measured dependence of SI on signal bandwidth is now compared to the XYMcD model, which accounts for finite signal bandwidth and hydrophone beam width. Only the effect of signal bandwidth has been studied in the present paper. Figure 12 shows measured SI for both LFM and CW signals as a function of inverse signal bandwidth $\frac{1}{\beta}$, which (after match filtering) is equal to signal temporal extent. SI appears to increase approximately linearly with log inverse bandwidth. For comparison, Fig. 12 also shows SI computed using the XYMcD model and the illumination factor in Eq. (15) versus $\frac{x_0}{10c}$, where c is the speed of sound in water, and $\frac{1}{10}$ is chosen to fit SI measurements around the 1 ms temporal extent. The parameter x_0 is a measure of the width of the ensonified area, thus it is proportional to $\frac{c}{\beta}$. Strictly speaking the width of the ensonified area is given by $\frac{c}{\beta\cos\theta}$, but $\cos\theta$ is larger than 0.95 for all projector-receiver paths used in this paper so the angular dependence is neglected.

In Fig. 12, the SI measurements are partitioned into four groups according to Φ. XYMcD predictions are calculated for the value of $\log_{10}(\Phi)$ closest to the experimental values, except that computational difficulties limited model predictions to $\log_{10}(\Phi)\geq 0.4$. As mentioned in Sec. III, XYMcD model predictions depend very little on Λ for Λ between 0.3 and 1.9, and XYMcD predictions are given for $\Lambda=0.5$ only. The model predictions in the two upper panels of Fig. 12 are for $\log_{10}(\Phi)=0.4$, which is higher than that of the measurements. The model-data comparison is not particularly good, although the measured SI does increase with inverse bandwidth and the model prediction does increase with the ensonified area. The model predictions in the lower left panel of Fig. 12 are for $\log_{10}(\Phi)=0.4$ and $\log_{10}(\Phi)=0.6$. The $\log_{10}(\Phi)=0.6$ model prediction matches the data reasonably well for the lower bandwidth data. The model prediction in the lower right panel is for $\log_{10}(\Phi)=0.6$, and it also matches the lower bandwidth SI measurements reasonably well. In all panels the rates of decline do not agree very well, and there is some indication that some mechanism prevents measured SI from dropping below a threshold for the largest bandwidth. This point is discussed further in Sec. VI.

VI. DISCUSSION

The measurement-model comparison in Fig. 10 indicates that values of SI measured under conditions that did not seriously violate the continuous signal and omnidirection transducer assumptions of the YMcD model matched the trend of the model reasonably well. The range of Φ-Λ represented by the measurements is modest—certainly it would be valuable to measure SI under conditions that more fully explored the Φ-Λ parameter space. Figure 11 characterizes the clear dependence of measured SI on signal temporal resolution, or equivalently, inverse bandwidth. Higher bandwidth signals resulted in lower SI and poorer agreement with the YMcD model, which assumes continuous signals with infinitely narrow bandwidth.

Figure 12 shows that measured SI decreases from about 0.85 to about 0.5 as bandwidth increased from 1 to 22 kHz. A physical explanation is that smaller temporal extent resulting from increased bandwidth corresponds to a smaller ensonified surface area, so that fewer surface facets contribute to the received signal at any point in time; as a result fluctuations due to interference between micropaths are reduced. Figure 12 also contains SI predictions made using an extended (XYMcD) model, in which ensonified area is restricted via an illumination factor. The XYMcD model predicts a decrease in SI with reduced ensonified area consistent with the measured trend. However, the difference between the measurements and the XYMcD model predictions becomes large for the largest bandwidth measurements. The main reason for this discrepancy is thought to be the spatial averaging used in the model to calculate SI. As can be seen in Eq. (5) for the YMcD model, or in Eqs. (10) and (12) for the XYMcD model, SI is calculated by integrating contributions over the ocean surface, which corresponds to a spatial averaging. On the other hand the measured SI is calculated by temporal averaging (ping-to-ping averaging). When a large area of the ocean surface is ensonified, the two are equivalent. However, for small ocean surface ensonification, the number of contributions becomes small and the spatial averaging yields smaller SI than the temporal averaging does. To push this analysis to the extreme, when ocean sur-

face ensonification is very close to 0 (contributions from specular only), the mean intensity [see Eq. (10)] and the mean squared intensity [see Eqs. (12) and (13)] only depend on the normalized surface correlation function at zero lag which is 1; as a result the dependence on the surface statistics is removed and SI is equal to 0. That is why the XYMcD model predicts that SI tends to 0 when temporal extent tends to 0, as shown in Fig. 12. In the measurements, there is still ping-to-ping variations due to ocean "facets" moving with time, so SI does not decrease below a certain threshold that depends on the strength/roughness parameter Φ.

Two other bandwidth effects could be important. First the transmit voltage response (TVR) of the projectors used in the experiment is frequency dependent. This causes the effective bandwidth of the signals in the water to be lower for the 13 and 22 kHz bandwidth signals, and consequently moves the corresponding points in Fig. 12 to the right. Second, model predictions are performed at the center frequency (i.e., 20 or 40 kHz); for the largest bandwidth (and smallest temporal extent), many other frequencies are present in the received signal. This effect is not accounted for in the XYMcD model.

A caveat to the model-data comparison presented in this paper is that the YMcD model predicts SI due to interaction with the rough surface only, while the measured SI may be caused by other factors as well, such as inhomogeneities in the volume or near-surface bubbles. Thus generally speaking, for the measured SI

$$(\mathrm{SI})_{\mathrm{total}} = (\mathrm{SI})_{\mathrm{surf}} + (\mathrm{SI})_{\mathrm{other}} \geq (\mathrm{SI})_{\mathrm{surf}}, \tag{20}$$

where $(\mathrm{SI})_{\mathrm{total}}$ is the measurement result, $(\mathrm{SI})_{\mathrm{surf}}$ is that due to the interaction with the ocean surface, and $(\mathrm{SI})_{\mathrm{other}}$ is scintillation due to other effects. $(\mathrm{SI})_{\mathrm{surf}}$ is the quantity predicted by YMcD and XYMcD. Romond (2004) looked at the SI of the direct path for the same data set considered in the present study. The objective of that study was to investigate the effect of an intermittent current jet (localized in depth) on SI. Looking at four of the seven measurement periods only, Romond observed that SI was less than 0.1 for all projector-hydrophone pairs during three of the periods when the magnitude of the current jet was very low. However, during a measurement period when the current was higher, and for shallow projector-hydrophone pairs for which the direct path propagated through the current jet in a largely horizontal direction, the SI was found to be as high as 0.4 (the maximum value of 0.4 corresponded to P_3-H_1). Thus SI values greater than 0.1 were attributed to interaction of the acoustic waves with the current jet.

In the present study of SI for surface forward-reflected signals, shallow projectors and receivers could not be used. Surface-reflected rays connecting the deeper projector-hydrophone pairs propagate at much higher angles than the direct path rays and thus spend comparatively less time in the depth stratum where the current jet was observed. Therefore, since $(\mathrm{SI})_{\mathrm{total}}$ is always greater than 0.4 for surface forward-scattered signals, and surface bounce rays are steeper than direct path rays, $(\mathrm{SI})_{\mathrm{other}}$ should not significantly contribute to $(\mathrm{SI})_{\mathrm{total}}$ and $(\mathrm{SI})_{\mathrm{total}} \approx (\mathrm{SI})_{\mathrm{surf}}$.

VII. SUMMARY

Ocean measurements of scintillation index (SI) for acoustic signals forward scattered by the ocean surface have been presented and compared with a model by Yang and McDaniel (1991). To our knowledge, this is the first such comparison for ocean surface forward-scattered signals. The model uses the Kirchhoff approximation and assumes two-dimensional scattering geometry, CW signal, and omnidirectional projectors and hydrophones. The model utilizes two parameters: Φ which is proportional to surface roughness (strength or roughness parameter) and Λ which is a function of geometry and the surface waveheight correlation length (size or diffraction parameter). The experiment employed a range of geometries and signal frequencies, which in turn provided measurements over a range of values of Φ and Λ. The measurement took place close to, and on the leeward side of, San Clemente Island, with the result that the sea surface was quite benign and basically constant throughout the experiment. Measurements of SI reported here were averaged over the seven measurement periods. In general, measured SI was found to compare favorably with the Yang and McDaniel's model as long as the model assumptions (continuous signals and omnidirectional acoustic transducers) were not seriously violated. However, SI measured using signals with large bandwidth or directional transducers was lower than that predicted by the model. Approximately the same SI was measured using LFM signals and CW pulses having the same effective bandwidth, or equivalently, the same temporal extent, although the LFM signals were 8 to 70 times longer, indicating no simple dependence upon signal duration.

In order to investigate how increasing signal bandwidth (shorter temporal extent) reduces SI, the Yang and McDaniel model was extended to include an ocean surface illumination factor that controls the extent of the ocean surface that contributes to the scattered signal. SI predicted using the extended (XYMcD) model is found to decrease with decreasing ensonified surface area, but the relationship between ensonified area and signal temporal extent has not been quantified. Also, since spatial averaging is used in the model to calculate SI, XYMcD model predictions underestimate SI when the ensonified area is very small, which explains the discrepancy between model predictions and measured SI for the largest bandwidth signals. Although transducer directionality was not considered in the paper, it can be studied using the XYMcD model with an illumination factor corresponding to the projector/hydrophone directionality.

The YMcD and XYMcD models can be used to design future experiments. For instance, it is quite challenging to obtain SI measurements in the partially saturated region, where SI is above 1. However it can be done using short ranges and a range of acoustic frequencies, as long as the surface wave height elevation is not too small.

ACKNOWLEDGMENTS

Support by the Office of Naval Research is gratefully acknowledged. Special thanks go to Captain Jim Binford, USN (retired), NAVSEA PMS 415; Captain Bill Gaines,

USN (retired), and the technical staff of Marine Physical Lab, Scripps Institute of Oceanography; Penn State acoustics graduate students Tom Weber, Rachel Romond, and Steven Lutz; the Captains and crews of M/V Independence and M/V Acoustic Explorer. The authors also thank the reviewers for their constructive comments and suggestions.

[1]The Rayleigh roughness parameter is sometimes defined as $2\pi\frac{h\sin\theta}{\lambda}$ [see, for instance, Brown (1969) and Nichols and Senko (1974)].
[2]Personal communication with Al Todd of Axys Technologies.

Brown, M. V. (**1969**). "Intensity fluctuations in reflections from the ocean surface," J. Acoust. Soc. Am. **46**, 196–204.
Clay, C. S., and Medwin, H. (**1977**). *Acoustical Oceanography: Principles and Applications* (Wiley, New York).
Colosi, J. A., and Baggeroer, A. B. (**2004**). "On the kinematics of broadband multipath scintillation index and the approach to saturation," J. Acoust. Soc. Am. **116**, 3515–3522.
Dahl, P. H. (**1999**). "On bistatic sea surface scattering: Field measurements and modeling," J. Acoust. Soc. Am. **105**, 2155–2169.
Dahl, P. H. (**2004**). "Forward scattering from the sea surface and the van Cittert-Zernike theorem," J. Acoust. Soc. Am. **115**, 589–599.
Flatté, S. M. (**1983**). "Wave propagation through random media: Contributions from ocean acoustics," Proc. IEEE **71**(11), 1267–1294.
Flatté, S. M., Dashen, R., Munk, W. H., Watson, K. M., and Zachariasen, F. (**1979**). *Sound Transmission through a Fluctuating Ocean* (Cambridge University Press, Cambridge).
Frankenthal, S. (**1990**). "The bichromatic intensity correlation of radiation reflected off a randomly irregular surface," J. Acoust. Soc. Am. **87**, 2467–2476.
Macaskill, C., and Kachoyan, B. J. (**1988**). "Numerical evaluation of the statistics of acoustic scattering from a rough surface," J. Acoust. Soc. Am. **84**, 1826–1835.
McDaniel, S., and McCammon, D. (**1987**). "Composite-roughness theory applied to scattering from fetch-limited seas," J. Acoust. Soc. Am. **82**, 1712–1719.
Melton, D. R., and Horton, C. W. (**1970**). "Importance of Fresnel correction in scattering from a rough surface. I. Phase and amplitude fluctuations," J. Acoust. Soc. Am. **47**, 290–298.
Nichols, R. H., and Senko, A. (**1974**). "Amplitude fluctuations of low-frequency underwater acoustic pulses reflected from the ocean surface," J. Acoust. Soc. Am. **55**, 550–554.
Plant, W. J. (**2002**). "A stochastic, multiscale model of microwave backscatter from the ocean," J. Geophys. Res. **107**(C9), 3120–3140.
Romond, R. A. (**2004**). "The effect of near-surface ocean currents on high-frequency underwater acoustic signal intensity," Master's thesis, The Pennsylvania State University.
Scrimger, J. A. (**1961**). "Signal amplitude and phase fluctuations induced by surface waves in ducted sound propagation," J. Acoust. Soc. Am. **33**, 239–247.
Stroud, J. S., Marston, P. L., and Williams, K. L. (**1997**). "Intensity moments of underwater sound scattered by a Gaussian spectrum corrugated surface: Measurements and comparison with a catastrophe theory approximation," in *High Frequency Acoustics in Shallow Water*, edited by N. G. Pace, E. Pouliquen, O. Bergem, and A. P. Lyons (NATO SACLANT Undersea Research Centre, La Spezia, Italy), pp. 525–532.
Thorsos, E. I. (**1988**). "The validity of the Kirchhoff approximation for rough surface scattering using a Gaussian roughness spectrum," J. Acoust. Soc. Am. **83**, 78–92.
Thorsos, E. I. (**1990**). "Acoustic scattering from a "Pierson-Moskowitz" sea surface," J. Acoust. Soc. Am. **88**, 335–349.
Yang, C. C., Fennemore, G. C., and McDaniel, S. T. (**1992**). "Scintillation index of the acoustic field forward scattered by a rough surface for two- and three-dimensional scattering geometries," J. Acoust. Soc. Am. **91**, 1960–1966.
Yang, C. C., and McDaniel, S. T. (**1991**). "Fourth moments of acoustic waves forward scattered by a rough ocean surface," Waves Random Media **1**, 419–439.

Near-resonance scattering from arrays of artificial fish swimbladders

R. W. Nero,[a] C. Feuillade, and C. H. Thompson[a]
Naval Research Laboratory, Stennis Space Center, Mississippi 39529-5004

R. H. Love
BayouAcoustics, Fort Worth, Texas 76179

(Received 2 May 2006; revised 5 September 2006; accepted 1 October 2006)

The air-filled swimbladders of fish resonate like damped air bubbles, and are very efficient acoustic scatterers at low to mid frequencies (typically <20 kHz). Scattering experiments were performed on an artificial "fish school" constructed from polyethylene bubbles. A mathematical model, developed to describe near-resonance backscattering from schooling fish [J. Acoust. Soc. Am. **99**, 196–208 (1996)], was used to analyze the physical behavior for three different arrays of these bubbles. The measurements gave excellent agreement with the model, showing that coupled-resonance and interference effects cause the frequency response of tightly packed arrays, with spacing corresponding to the order of a body length for fish, to differ significantly from those of more dispersed arrays. As the array spacing is increased to the equivalent of several body lengths, these effects rapidly diminish. The results of this comparison demonstrate that, at low to mid frequencies, coupled resonance and interference effects are likely in schooling fish, and need to be considered in applications of underwater acoustic methods to the study of fish populations. [DOI: 10.1121/1.2382277]

PACS number(s): 43.30.Sf, 43.30.Ft, 43.30.Gv [KGF] Pages: 132–143

I. INTRODUCTION

Low to mid frequency (<20 kHz) acoustic backscattering within the ocean volume is usually dominated by scattering from fish swimbladders.[1] Analytical models of low frequency scattering from individual swimbladders[2,3] have been used successfully to explain experimentally measured reverberation levels from layers containing many dispersed (i.e., widely spaced) fish in a number of oceanic regions.[4–11] In each of these experiments, it was shown that scattering strengths over broad frequency bands could be directly attributed to swimbladder resonance scattering from fish populations.

Schooling fish generally space themselves from one to several body lengths apart.[12] Low to mid frequency backscattering measurements on fish schools have shown that scattering models for individual fish swimbladders are not applicable to fish ensembles in which the fish are swimming so closely together.[13–15] The results of these measurements indicate that acoustic interactions between the schooling fish modify the character of the backscatter from an individual fish embedded in the ensemble. A semiempirical model of scattering from schools of fish, that includes multiple scattering and attenuation effects, exists for frequencies well above swimbladder resonance.[16] However, since earlier scattering models[2,3,16] are not applicable to schooling swimbladder-bearing fish in the frequency range of the swimbladder resonance, an analytical model of scattering from small schools of swimbladder fish in this regime has been developed.[17] This model uses scattering from an individual fish swimbladder as its kernel,[5] and incorporates coupled-resonance and coherent interaction effects, but not absorption. Predictions made using the model indicate that fish in tightly packed schools scatter sound differently from an individual fish in two ways. First, coupled-resonance interactions between swimbladders (i.e., where the resonance motion of an individual swimbladder is modified by acoustic interactions with its similarly resonating neighbors) will typically cause the frequency and level of the resonance peak to be shifted downward. Second, destructive and constructive interference effects (i.e., coherent interactions between the acoustic fields scattered back to the receiver by different swimbladders) will cause additional peaks and troughs to appear in the scattering levels at frequencies near resonance and above. Both of these effects are strongly dependent on the size of the swimbladders, their depths, and the distance and regularity of the spacings among fish in the school.

Verification of this school scattering model using available acoustic backscatter data from oceanic fish schools is difficult, because most of the biological parameters the model requires as inputs are typically not obtained. In two reported cases, fish schools were encountered during the course of other acoustic experiments, and no biological information was recorded.[13,14] In a third case, the measurement of scattering from fish schools was the objective, and each target school was sampled to determine fish species and sizes, and estimates of school depth were also made.[15] However, spacings among the fish could not be determined. Recently, a transmission loss experiment detected a decrease in the resonance frequency for schooling fish,[18] but again exact school parameters were uncertain.

[a]Present address: National Marine Fisheries Service, Stennis Space Center, Mississippi, 39529-0001.

Because fish spacings are critical to the school scattering model, and simultaneous measurement of school backscatter and fish sizes and spacings within the school targets are difficult to obtain, the Naval Research Laboratory (NRL) decided to conduct low to mid frequency backscatter measurements on several arrays of artificial swimbladders (referred to in this present work as "bubble schools"). The assumption was made that many of the parameters, such as scatterer number, size, depth, and spacing, that are typically undetermined in field measurements made on oceanic fish schools, could be controlled and known using artificial bubble schools. Analysis of the scattering properties of these known bubble school targets could be used to verify the school scattering model and, subsequently, facilitate the interpretation of scattering measurements from oceanic schools.

This paper briefly reviews the small school model, describes the experimental measurements made using the bubble schools, presents the results of those measurements, details the model-data comparison, and discusses the validity and limitations of the model.

II. SCHOOL SCATTERING MODEL

Only a brief summary of the small school scattering model used in this work is provided here. A detailed explanation is available in Ref. 17.

The level of resonant acoustic scattering from an individual gas filled swimbladder, which is assumed to be spherical, is modeled on the closely related work of Devin (who considered scattering from air bubbles in water)[19] by solving the equation of motion

$$m\ddot{v} + b\dot{v} + \kappa v = -Pe^{i\omega t}, \tag{1}$$

where v is the differential volume (i.e., the difference between the instantaneous and equilibrium swimbladder volumes). The coefficient $m(=\rho/4\pi a)$ is termed the inertial "mass," where a is the swimbladder radius, ρ the water density; and $\kappa(=3\gamma P_A/4\pi a^3)$ is the "adiabatic stiffness," where γ is the ratio of gas specific heats and P_A is the ambient pressure. The coefficient b describes the damping of the swimbladder motion, while P and ω represent the amplitude and frequency, respectively, of the periodic external pressure field applied to the swimbladder (P is the difference between the total instantaneous pressure field at the swimbladder and P_A). If a harmonic steady state solution of Eq. (1) of the form $v=\bar{v}e^{i\omega t}$ is assumed, substitution gives the resonance response

$$\bar{v} = \frac{-P}{\kappa - \omega^2 m + i\omega b} = \frac{-(P/m\omega^2)}{\left[\frac{\omega_0^2}{\omega^2} - 1\right] + i\frac{b}{m\omega}}, \tag{2}$$

where $\omega_0=\sqrt{\kappa/m}=(\sqrt{3\gamma P_A/\rho})/a$ is the resonance frequency. The factor $(b/m\omega)$ in the denominator can be identified with a damping constant δ for the swimbladder, consisting of radiative, viscous and thermal terms, i.e.,

$$\frac{b}{m\omega} = \delta = \delta_r + \delta_v + \delta_t. \tag{3}$$

In the theory of resonant acoustic scattering by fish swimbladders developed in Ref. 3, the quantity δ is replaced by a factor $(\omega_0/\omega H)$. The frequency dependent parameter H, which is analogous to the "quality factor" Q for the monopole resonance of an air bubble in water,[20] also consists of three components, which are combined as follows:

$$\frac{1}{H} = \frac{1}{H_r} + \frac{1}{H_v} + \frac{1}{H_t}. \tag{4}$$

In the case of fish, the damping due to thermal conductivity effects is generally negligible compared to radiative and viscous damping. The values of H_r and H_v are given by

$$H_r = \frac{\omega_0 c}{\omega^2 a}; \quad H_v = \frac{\omega_0 \rho a^2}{2\xi}, \tag{5}$$

where c is the speed of sound in water and ξ is the viscosity of fish flesh surrounding the swimbladder. The differences between the densities and sound speeds of water and fish flesh is typically very small and may be neglected.

The steady state level of acoustic scattering from a swimbladder embedded within a fish school, when ensonified by an external field and also radiatively coupled with the other swimbladders in the school, is determined by solving the matrix equation $\mathbf{Mv}=\mathbf{p}$; where $\mathbf{v}=\{\bar{v}_1,\ldots,\bar{v}_n,\ldots,\bar{v}_N\}$ and $\mathbf{p}=\{-P_1e^{i\phi_1},\ldots,-P_ne^{i\phi_n},\ldots,-P_Ne^{i\phi_N}\}$ are column vectors containing the steady state volume oscillation amplitudes and external fields, respectively, for the N swimbladders in the school, and $\mathbf{M}$ is an $N\times N$ matrix with elements:

$$\begin{aligned} M_{nn} &= \kappa_n - \omega^2 m_n + i\omega b_n, \\ M_{nj} &= \frac{-\omega^2 \rho e^{-ikr_{nj}}}{4\pi r_{nj}} \quad (n \neq j). \end{aligned} \tag{6}$$

Each diagonal term (e.g., M_{nn}) describes the resonance behavior of an individual swimbladder, as if it were uncoupled from all the others. Variations in size, damping, depth, etc., are incorporated into these diagonal elements. Every off-diagonal element (e.g., M_{nj}) describes the radiative coupling between two of the swimbladders, which is expected to be particularly important for ensonification frequencies close to the resonance frequencies of the individual swimbladders. The solution of the matrix equation (i.e., $\mathbf{v}=\mathbf{M}^{-1}\mathbf{p}$) enables the description of steady state scattering from the whole ensemble of swimbladders as a function of the external field amplitude and frequency. Once the solutions $\bar{v}_n$ are found, the scattered pressure field (and hence target strength) for the whole school is obtained using coherent summation, which incorporates interference effects.

The school scattering model is intended to apply to small schools of fish, where it may be reasonably assumed that absorption of sound within the school is negligible, and therefore that all the fish experience the same incident field amplitude (although not the same phase) such that $P_1=\cdots=P_n=\cdots=P_N=P$. Absorption of sound within the school is not considered in the model. The school target strength for

one configuration (i.e., one specific geometrical arrangement of fish locations) of fish in a school is given by

$$\mathrm{TS} = 10\log_{10}\left(\frac{\sigma_s}{4\pi}\right) = 10\log_{10}\left(\frac{\omega^4\rho^2\left|\sum_{n=1}^{N}\bar{v}_n e^{i\phi_n}\right|^2}{(4\pi)^2 P^2}\right), \tag{7}$$

where σ_s is the back scattering cross-section for the school, and the factors $e^{i\phi_n}$ appearing in the numerator of the right-hand side (rhs) take into account the different phases of the external field incident on the various swimbladders, according to their locations within the school configuration.

When Eq. (7) is used to simulate scattering from an ensemble of swimbladders, several assumptions and steps can be taken to provide as realistic a simulation as possible. The off-diagonal elements of matrix **M** incorporate swimbladder positions relative to each other, and the column vector **p** includes the phases of the external acoustic field. Hence, the simulations can include exact descriptions of variations in swimbladder position within the ensemble, and the position of the ensemble relative to the source and receiver. Therefore, parameters of variable nature, such as swimbladder position and target orientation, can be randomized about their expected values for a number of simulations, producing a realistic average target strength for the ensemble.

Figure 1 illustrates examples of using the small school model to calculate scattering from groups of $N=27$ fish at different body length spacings. In each case the fish are assumed to be 10 cm long with spherical swimbladders 1 cm in diameter. They are placed at a depth of 100 m. Two values of viscosity, one realistic (20 Pa s), and the other unrealistically high (2000 Pa s), were selected to illustrate different features of the scattering. (N.B. The range of viscosity values typically used for analyzing volume scattering data for fish is ~10–50 Pa s. The high value of 2000 Pa s is used in Fig. 1 to emphasize the scattering phenomena predicted by the model when the individual fish resonance is suppressed by heavy damping.) As described previously in Ref. 17, the curves shown in Fig. 1 were produced by simulating fish schools constructed from a cubic-lattice, in a manner which is intended to approximate the formations fish adopt when swimming closely together. The positions of the fish are allowed to vary slightly, so that the structure is not completely uniform.

In Fig. 1(a) it is assumed that there are no acoustic (i.e., radiative) interactions between the fish (i.e., the off-diagonal elements $M_{nj}=0$, for all n, j), which is equivalent to assuming that the fish are spaced an "infinite" number of fish lengths apart. For this case also, as described in Ref. 17, the target strength of the group is given by incoherently summing the target strengths of the 27 fish. This is calculated as the target strength of a single fish (obtained using the individual swimbladder scattering model[3]) plus $10\log(N) = 10\log(27)$. For the case where fish flesh viscosity is 20 Pa s, a strong swimbladder resonance peak is seen at about 2100 Hz. However, when the viscosity is increased to 2000 Pa s, the swimbladder is heavily damped and the resonance peak is eliminated.

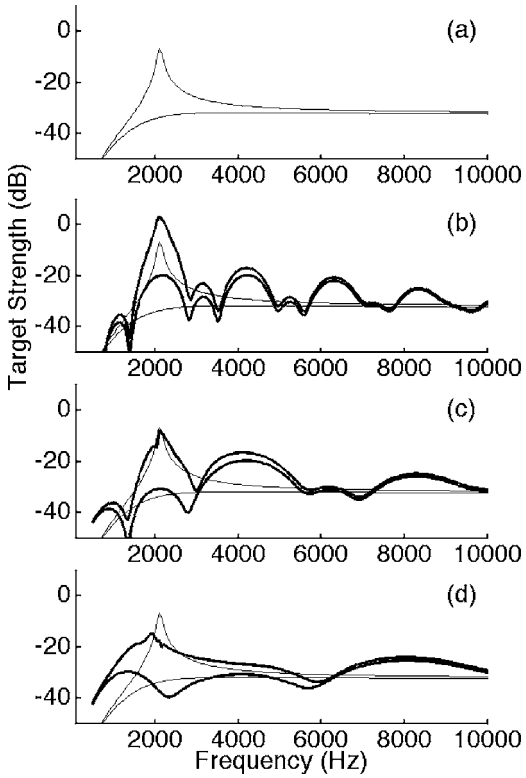


FIG. 1. Small school model simulations of scattering from groups of 27 fish at different body length spacings. (a) No acoustic (radiative coupling) interactions between swimbladders (i.e., the fish are spaced "infinitely" far apart from each other). Upper thin line: viscosity is 20 Pa s. The swimbladder resonance is clearly seen. Lower thin line: Viscosity is 2000 Pa s. The resonance is eliminated by heavy damping. These two thin lines are also superimposed on panels (b)–(d) for comparison. (b) Average fish spacing of four body lengths. Upper thick line: With resonance. Lower thick line: Resonance damped. (c) Average fish spacing of two body lengths. Upper thick line: With resonance. Lower thick line: Resonance damped. (d) Average fish spacing of one body length. Upper thick line: With resonance. Lower thick line: Resonance damped.

Figures 1(b)–1(d), illustrate how the scattering behavior is modified as the fish in the school are spaced progressively closer together. Ten simulations with randomly selected positions were run, and an average calculated, for each school spacing. The two curves for the "infinitely" spaced fish [Fig. 1(a)] are also included in each of these figures.

In each of Figs. 1(b)–1(d) for the 2000 Pa s viscosity case, a series of peaks and troughs is seen to replace the smooth curve in Fig. 1(a). These peaks and troughs are indicative of frequency dependent coherent interactions (i.e., constructive and destructive interference) of the acoustic fields scattered from the different fish in the school. The variations allowed in the positioning of the fish within the lattice structure tend to broaden the peaks slightly, and also remove the unnaturally deep troughs that have been observed to occur if a strictly uniform configuration is assumed. As the fish are brought closer together, from four to two body lengths apart, and then from two to one body length(s) apart, the interference peaks and troughs are seen to broaden and spread apart. For example, analysis shows that the peak seen

at around 2000 Hz at four body length spacing [Fig. 1(b)] shifts upwards in frequency as the fish spacing is reduced, and is identified with the peak seen at about 4000 Hz at two body lengths [Fig. 1(c)], and with that seen at about 8000 Hz at one body length [Fig. 1(d)]. Model results at intermediate spacings (not presented here) show that the peaks do undergo a shift in frequency, as opposed to various stationary peaks being enhanced or decreased by interference effects.

In the lower viscosity (i.e., 20 Pa s) case, the effect of the swimbladder resonance on scattering is clearly seen. At the four body length spacing [Fig. 1(b)], the swimbladder resonance peak, and a strong interference peak, coincide, resulting in an apparently very high resonance peak level for the school target strength at this frequency. At the two body length spacing [Fig. 1(c)], the swimbladder resonance peak coincides with a weak interference peak, and the resonance peak of the school differs only slightly from seen in Fig. 1(a). At one body length spacing [Fig. 1(d)], the swimbladder resonance peak coincides with an interference trough and, in addition, the resonance peak of the school decreases in amplitude and frequency due to coupled-resonance effects. The combination of these two effects causes the resonance peak to be lowered in frequency and target strength, and also broadened. In the low viscosity cases, the individual swimbladder resonance plays a major role in scattering only in the vicinity of resonance peak. An octave above or below the resonance frequency, the scattering level is increased by only 2–4 dB compared to the high viscosity case. Another remark that may be made about Figs. 1(b) and 1(c) is that towards the high end of the frequency range the target strength of the school tends to approach that of a comparable number of nonacoustically interacting fish. As we shall see, this feature is also discernible in the experimental data, and in other simulations using the small school model.

The purpose of the bubble school measurements described here was to determine whether the scattering behavior predicted by the small school scattering model, as exemplified in Fig. 1, could be confirmed by a controlled experiment.

III. MEASUREMENTS

During November 1997, NRL conducted a low to mid frequency volume scattering experiment in the northern Gulf of Mexico aboard the R/V GYRE. In conjunction with volume scattering measurements on actual fish, target strength measurements of bubble schools were made during the afternoon of 9 November 1997. Measurements were made in the vicinity of 29.0 °N, 87.4 °W, in water depths of about 1600 m. Weather conditions during the measurements were benign: Winds blew at 2–4 m/s; waves were 0–0.3 m; and swell was 0.5–1 m. During the course of the 6 h measurement sequence, the ship drifted about 2.5 nmi in a southeasterly direction.

A temperature-salinity profile obtained at the end of the measurement sequence was used to determine the sound speed profile. From the sea surface to about 50 m depth the water column was well mixed with a water temperature of 23.5 °C and a sound speed of 1534 m/s. From the bottom of the mixed layer to 100 m the temperature and sound speed linearly decreased to 18.5 °C and 1520 m/s, respectively.

FIG. 2. (Color online) Detail of the bubble school, showing supporting multifilament netting and bubbles glued into position.

Three bubble schools were deployed, and experimental data recorded: A "full" school; a "medium" school; and a "sparse" school. The full school was constructed with individual bubbles cut from a sheet of polyethylene closed cell packaging film (commonly known by the trade name "Bubble Wrap®") glued to the interstices of 9 cm nylon multifilament-gill netting to form a $9\times9\times9$ cube of 729 bubbles (Fig. 2). The spatial separation between bubbles was, therefore, 9 cm for the full school. The netting was attached to a 1 m cubic frame constructed of PVC pipes (Fig. 3). The medium school was a $5\times5\times5$ cube of 125 bubbles formed by removing every other plane of bubbles from the full school. The spatial separation between bubbles was

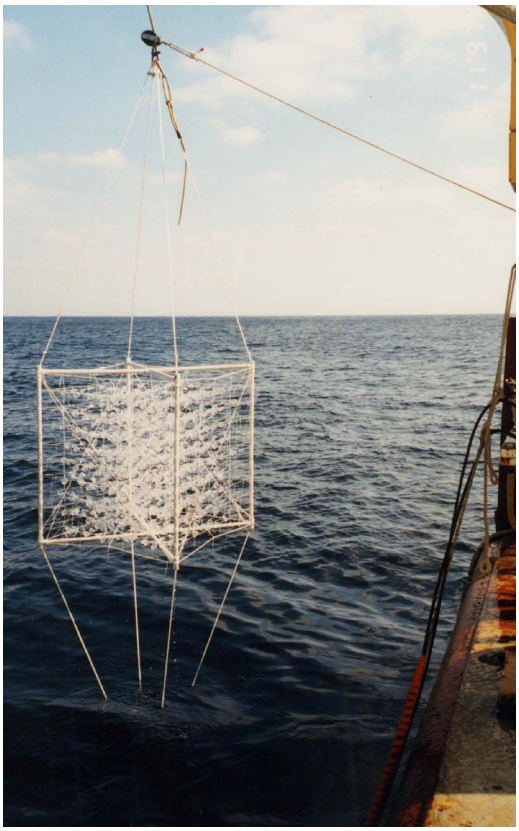


FIG. 3. (Color online) Dense bubble school being retrieved at sea. The $9\times9\times9$ array of bubbles, and the PVC frame, are clearly displayed.

18 cm for the medium school. The sparse school was a 3 $\times 3 \times 3$ cube of 27 bubbles formed by removing every other plane of bubbles from the medium school. The spatial separation between bubbles was 36 cm for the sparse school. All three schools were 72 cm on a side. Two weeks prior to going to sea, measurements on a sample of bubbles using their displacement in water gave an average volume of 8.6 ml (equivalent spherical radius=1.27 cm). The bubbles were very uniform in size, with a coefficient of variation (standard deviation/mean) of the bubble radius of only 0.0064. During deployment, water motion displaced small unwanted air bubbles from the surfaces of the Bubble Wrap® material, and from the PVC frame.

The acoustic measurement technique employed explosive sources and a downward-looking receiver to measure target strengths over a wide frequency range. Sources were 0.23 kg blocks of TNT, electrically detonated 0.5 m below the surface. This shallow detonation depth allowed the gas bubble created by the explosion to vent, and eliminated the multiple sound pulses caused by bubble oscillations associated with detonations at greater depths providing a single sharp sound pulse free of surface reflections.

The receiver was an NRL USRD Model F78, which consists of a 48-element line hydrophone, 1.4 m long, mounted along the axis of a 2.7 m diameter 45° conical reflector. By grouping the upper 8, 16, 32, or all 48 hydrophone elements, four different apertures can be formed, and the 3 dB beamwidth can be maintained between 10° and 20°, between 1.7 and 20 kHz. The beamwidth obtained using all 48 elements increases to about 40° at 1 kHz, and to 50° at 500 Hz.

The F78 was deployed from a stern U-frame to a depth of 7 m, just below the ship's hull. The bubble schools were also deployed from the stern U-frame, about 2 m to port of the F78. A dacron line was used to lower the bubble schools to a nominal depth of 96.6 m. A 27 kg lead ball was suspended about 5.5 m below the bubble frame, to keep the bubble schools near the axis of the F78 beam. The TNT blocks were deployed from the starboard side of the stern and allowed to drift aft and starboard away from the ship before being detonated. The horizontal distances between the detonations and the F78 were estimated at 20–30 m. Individual shots in a sequence were typically 6–9 min apart.

Mounted on the outside of the base of the F78 was a 200 kHz transducer that was connected to an echo sounder with a color monitor. This system was used to monitor the position of the bubble schools in the F78 beam. A spectrum of colors, from white to yellow to red to dark red, corresponded in a qualitative way to increasing echo strengths. During the target strength measurements, variations seen in the colors indicated that the bubble school was swinging slowly in the echo sounder beam and, therefore, in the F78 beam. Thus, the weight of the lead ball was not sufficient to overcome drag on the school. The drag was, of course, greatest on the full school, and the echo colors seen varied from yellow to red. The drag was less for the medium and sparse schools, when the echo colors observed varied from red to dark red. The TNT blocks were detonated at the instant when the echo colors observed were darkest.

Volume scattering measurements conducted earlier in the experiment showed that biological scatterers near the bubble school measurement site were well below 100 m during the daytime. At the start of the target strength measurement sequence for the bubble schools, two shots were fired to confirm that the background scattering at 100 m was negligible. The full bubble school was then deployed, and target strength data were obtained from ten shots. The school was brought back on deck, and bubbles were removed to produce the medium school. The medium school was deployed, and data were obtained from eight shots. This school was brought back on deck and additional bubbles were removed to produce the sparse school. The sparse school was deployed, and data were obtained from six shots. This school was brought back on deck, all remaining bubbles were removed, and then the frame was re-deployed. Six shots were fired to obtain scattering data for the frame. The first shot was fired with the frame at the same depth as the bubble schools. However, the echo signal was weak, so the frame was raised to about 58 m for the final five shots. Measurements on the frame were completed 30 min after sunset, and concurrent observations with a hull mounted 38 kHz echosounder showed the measurements were complete shortly before the biological scatterers rose to their shallow nighttime depths.

The F78 signals received from each shot were amplified, high- and low-pass filtered at 400 Hz and 12 kHz, digitized at a 40 kHz sampling rate, and stored. Digitally stored data were subsequently filtered into 1/6-octave bands, and amplitude versus time envelopes calculated for each band. Data samples collected just prior to source detonation were used to calculate average noise levels, which were used for noise level corrections to the data. Noise level corrected pressure amplitudes (p) were used to calculate target strengths (TS) for each band, i.e.,

$$\mathrm{TS} = 10\log\left[\int_{t_1}^{t_2} p^2(t)dt\right] - 10\log(\rho cE) + 20\log(d_s) + 20\log(d_r) + \alpha ct - 10\log(BW), \quad (8)$$

where t is time in seconds after source detonation; t_1 and t_2 are the times of the beginning and end of the echo, respectively, from the bubble school; ρ is the density of sea water in kg/m^3; c is the sound speed in m/s; E is the source energy flux density in joules/m^2/Hz in the 1/6-octave band measured at 100 m and corrected to 1 m; d_s and d_r are the distances in m from the explosive source to the bubble school, and from the bubble school to the F78 receiver, respectively; α is the absorption in dB/m; and BW is the bandwidth in Hz of each 1/6-octave band.

For the 48-, 32- and 16-element apertures, target strengths were calculated for 1/6-octave bands centered at 500 Hz to 3.15 kHz, 1–6.3 kHz, and 4–10 kHz, respectively. Since 10 kHz was chosen as the upper frequency limit for these measurements, data from the eight-element aperture were not processed independently.

Equation (8) provides an accurate TS only if the bubble school was centered on the maximum response axis (MRA) of the F78. However, the 200 kHz echo sounder indicated that the school was not always on axis. Therefore, overlap-

ping frequency bands of the 32- and 16-element apertures (4, 4.5, 5, 5.6, and 6.3 kHz) were used to determine corrected target strengths. The beam patterns for each aperture were compared, and a curve showing differences in dB down for the two apertures, at a given beam angle, was developed for each frequency. Target strengths were then calculated using Eq. (8), assuming that the school was centered on the F78 MRA. Differences between calculated target strengths for the two apertures were determined for each frequency. These differences were compared to the dB down difference curve to estimate the angular position of the school in the beam for each frequency. The results for the five frequencies were averaged to determine a final adjusted beam angle. Lastly, target strengths at all apertures and frequencies were corrected based on the school being positioned at the adjusted beam angle.

IV. RESULTS

A. Acoustic results

The initial acoustic results were the positions of the bubble frame relative to the MRA of the F78. Beam angle calculations show that the dense school was 8°–14° off axis for individual shots, with a mean of 10.8° for all shots. The medium school was 2°–5° off axis for all shots but one, which was 8° off axis, with a mean of 3.5° for all shots. The sparse school was 2°–4° off axis, with a mean of 2.9°. The empty bubble frame was 2°–5° off axis, with a mean of 3.0°. The individual angles were used in calculating measured target strengths, and also in the model simulations.

Figure 4 shows the target strengths obtained for each individual shot for the dense, medium, and sparse bubble schools; and also for the empty frame. The target strength of the empty frame is fairly flat, averaging about −30 dB between 1400 and 10 000 Hz, except for a noticeable peak of −24 dB at 4000 Hz. Below 1400 Hz the target strength of the frame apparently increases rapidly with decreasing frequency. It is highly probable that this is not due to any acoustical phenomenon of the frame itself, but is actually caused by weak scattering from the sea surface, whose effect becomes stronger as the F 78 beam becomes broader with decreasing frequency. This effect has been observed on many occasions, by NRL investigators, in data obtained using the same, or very similar, equipment to perform fish survey exercises in the ocean. Due to this contamination by surface scatter, no data below 1400 Hz is used in the analysis presented in this work.

Interactions between the bubble schools and the PVC frame used to deploy and support them could systematically bias the received data and thus affect its subsequent analysis. There are three ways in which interactions between the bubble schools and the frame could occur. First, by radiative coupling, through the water, of the acoustic resonances of the bubbles in the school and that of the frame itself. Second, by mechanical coupling via the physical structure of the bubble school apparatus. Third, coherent interference effects between the scattered fields of the bubbles, and of the frame.

The only physical action of the frame which is likely to interact with the acoustical response of the bubbles is the

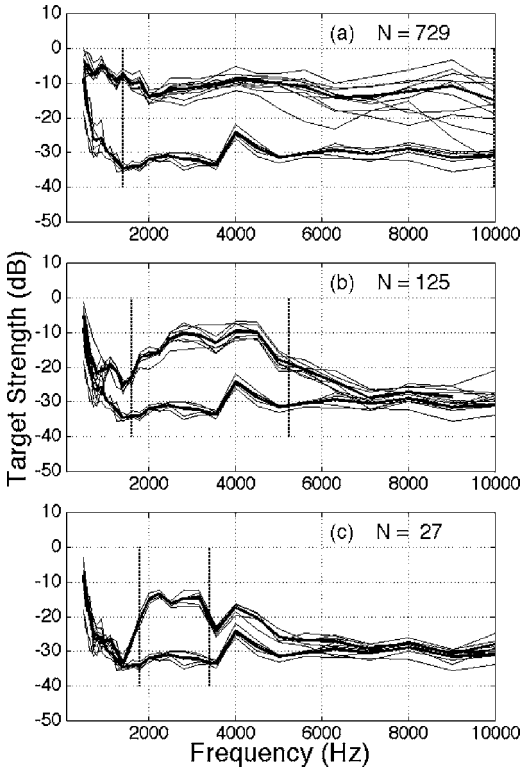


FIG. 4. Measured target strength of the bubble school. In each of the three panels the lower set of curves represent the measured target strength values for the empty frame, and are included for reference and comparison purposes. *N* is the number of bubbles in the school. Heavy lines denote mean values of the measured target strengths. The two dashed vertical lines indicate the frequency range within which the 12 dB criterion mentioned in the text is satisfied. (a) The dense bubble school. (b) The medium bubble school. (c) The sparse bubble school.

frame resonance peak at 4000 Hz. As will be shown below, modeling results presented here for the three different bubble schools indicate that the resonance frequency of an individual noninteracting bubble deployed during the experiment ranges between about 2000 and 3000 Hz. This suggests only minimal frequency overlap between the bubble resonances and the frame resonance at 4000 Hz, such that coupling between the two phenomena is probably very weak, either acoustically through the water or mechanically through the frame. It is almost certainly weaker than that of the bubbles among themselves (all of which have practically identical resonance frequencies) which, in the case of the sparse school represented in Fig. 4(c), where interaction with the frame is expected to be most problematic, produces only a small shift in the bubble resonance frequency. In accordance with this reasoning, it is assumed here that the effects of acoustical or mechanical coupling between the bubbles and the frame on the overall scattering properties of the bubble schools are negligible.

Coherent interference effects between the scattered fields of the bubbles, and of the frame, are of much greater importance. The frame resonance at 4000 Hz is especially problematic, especially in the case of the sparse school. The sparse school consists of three layers of bubbles, 36 cm apart; while the distances from the uppermost and lowest of these bubble layers, to the upper and lower sections of the

frame, respectively, is about 14 cm. Since the acoustic wavelength at 4000 Hz is about 37.5 cm, there could potentially be strong constructive interference between the scattered fields from the frame and the bubbles, leading to an apparently increased target strengths at this frequency. However, strong constructive interference is also expected from the three layers of bubbles *themselves* (which is an effect which is incorporated into the modeling using the smallschool scattering model). The practical effect of these phenomena is that, at 4000 Hz and above, the target strength of the sparse bubble school, and of the frame, cannot be readily separated.

In order to account for coherent effects, a criterion was adopted in order to specify which data could be considered unaffected by interference from the frame. The criterion was that there be a minimum 12 dB difference between the mean target strength level of the bubbles+frame, and of the target strength for the frame alone. The three panels in Fig. 4 each show two parallel dashed vertical lines. These indicate the frequency range within which this criterion is satisfied and thus where the data may be considered analyzable. Curves delineating the mean value of the data shots for each school, found by averaging the pressure field amplitude values, are included.

The argument for choosing this criterion is that, in the worst case, if the observed difference between the levels is 12 dB, and the scattered pressure field from the bubbles adds constructively to (i.e., in perfect phase with) that from the frame, then that would mean that the actual difference between the two TS would be about 9.5 dB. Conversely, if the observed difference between the levels is 12 dB, and the scattered pressure field from the bubbles adds destructively to (i.e., in perfect antiphase with) that from the frame, then that would mean that the actual difference between the two TS would be about 14 dB. This indicates a possible 4.5 dB range of systematic error due to coherent interaction effects. While not insignificant in itself, this range is less than the observed cumulative scatter of the different data shots for the bubbles+frame, and frame alone, which is at least about ±2.5 dB, as may be observed from inspection of Fig. 4. This implies that, even if there should be an error due to interference effects, this could not be discerned within the statistical scatter of the experimental data.

For the sparse school, the analyzable frequency range of the data is from about 1800 to 3400 Hz. This includes the two most prominent peaks in the data, which were then used in the fitting procedure. For the medium school, the frequency range is from about 1600 to 5250 Hz. For the dense school, the frequency range is from 1400 Hz (set by the lower limitation due to the surface scattering effect discussed earlier) to 10 000 Hz.

The target strengths of the dense school were generally flat between 2000 and 7000 Hz, with evidence of a small peak at 4000 Hz. The average target strength for all shots between 2000 and 7000 Hz was −9 dB. Below 2000 Hz, the target strengths increased somewhat with decreasing frequency. The target strengths for nine of the ten shots were within a few dB at all frequencies below 7000 Hz. Inaccuracies in the calculation of the beam pattern target strength correction cause the apparent large shot-to-shot variations above 7000 Hz. At angles of 8°–14° off the MRA, the dense school was near the edge of the F78 main lobe at the higher frequencies, so that small errors in the uncorrected target strengths lead to errors in calculated beam angles, which can magnify the errors in the beam corrected target strengths.

The target strengths of the medium school had a double peak between 2800 and 4500 Hz, and also a small shoulder at around 2000 Hz. The average target strength for all shots between 2800 and 4500 Hz was −11 dB.

The target strength of the sparse school shows a double peak between 2000 and 3150 Hz, as mentioned above. The average target strength for all shots between 2000 and 3150 Hz was −15 dB. Overall, a significant feature to note when comparing the results for the three schools is that the relatively flat character of the target strength curve for the dense school differs from that of the medium and sparse schools.

B. Physical observations

In an ideal experiment, the bubbles would have been identical perfect spheres, whose known locations were fixed in a perfect lattice that was normal to a co-located source and receiver. None of these ideal conditions was achieved in this experiment. The actual conditions are described below. However, much more is known, or can be inferred, about the physical conditions of the bubble schools than can be known or inferred about an actual fish school. Thus, although the conditions of the bubble schools are not fully known, they are known sufficiently well to provide a reasonable evaluation of the small school scattering model.

Figure 2 shows that the individual bubbles in air are not spherical. In water, buoyancy causes the bubbles to float above the netting with the air in the upper portion of the polyethylene shell. At depth the bubbles were most likely not spherical, but were probably crescent shaped pockets of air. This shape could be approximated by (possibly bent) prolate spheroids, or cylinders with hemispherical end caps. The monopole scattering from spheroids and cylinders has been shown to be very similar to that from spheres, the major difference being that the resonance frequencies increase slowly as the shape becomes less spherical.[21–23] For example, given spherical and prolate spheroidal bubbles of the same volume, the resonance frequencies of spheroidal bubbles with major-to-minor axis ratios of 5 and 10 are only 12% and 24% higher than that for the spherical bubble.[22]

Unfortunately the polyethylene bubble wall material was found to be permeable to gas at depth, so that the bubble volumes decreased during the course of the experiment. Each time the bubble school was retrieved, it was observed that the individual bubbles were smaller and their sizes were less consistent. At the end of the target strength measurements, the bubbles were estimated visually to have shrunk to about 1/2 of their initial volume (i.e., from 8.6 to about 4.3 ml), and there were noticeable size differences among the 27 bubbles in the sparse school. Exact measurements were not possible, because the equipment needed to perform a careful measurement was not available aboard the ship. Accurate measurements on the 27 bubbles comprising the sparse

school were conducted back in the laboratory five days after the target strength measurements were completed. By that time the bubbles had resorbed air and their average volume was 7.3 ml, which was 85% of what it had been a month earlier, before the experiment. The coefficient of variation of bubble radius was 0.043, about seven times higher than it had been before the experiment. Interestingly, volume measurements on these bubbles one year later (November 1998) showed little change; the average bubble volume was then 7.4 ml, and the coefficient of variation of bubble radius was 0.040.

Figure 2 shows that each individual bubble has a tail about 2 cm long by which it is glued to the netting. Hence, every bubble is free to sway about its tether, and its position on the netting could vary by up to 2 cm in all three planes.

The multifilament netting was strung as tightly as possible, but forces on the frame, and the buoyant force of the bubbles in the netting, caused the netting to bow, or warp, over its 72 cm span. Figure 3 shows that the dense bubble school is bowing downward in air as it is being retrieved. In water, the bubble buoyancy would cause the netting to bow upward to a somewhat greater degree. It was estimated from observations of the bubble schools hanging in the air while being deployed and retrieved, with the lead ball in the water, that the center of the dense bubble school would have been 10–15 cm higher than the sides while at depth. The warp for the sparse school was estimated to be about half that of the dense school, and the warp for the medium school was between that for the other two.

The echosounder mounted on the frame of the F78 receiver indicated that slant ranges from the surface to the bubble school were about 96 m. Comparison of this range to the nominal depth of 96.6 m estimated from the length of line deployed indicates there was virtually no catenary in the line. Hence the center of the dense bubble school was assumed to be at 95 m depth, and the centers of the medium and sparse schools were assumed to be at 96 m.

V. SIMULATIONS

For each shot in a sequence, a mathematical description of the school, including the positions of each bubble in the target, and the angles to the source and receiver, was set up and a simulation run to produce a predicted target strength. The school was placed at the calculated beam angle, and the declination angle to the source estimated from the deployment geometry. The source, target, and receiver, were assumed to be in the same plane, and the drift of the ship was assumed to be in the same direction along this plane. Values for the bubble shape and viscosity, and school warp and tilt, were the same during a sequence. Values for the coefficient of variation of bubble radius and sway were set, but the volumes and positions of each individual bubble were randomized and differed for each shot. The outputs of a sequence of simulations were averaged to produce a single curve, which was then compared to the measured target strength curves for the school of interest.

There are eight parameters in Table I whose values had to be set before results from the small school scattering model could be compared to measured bubble school target strengths. These parameters include bubble volume (mean and coefficient of variation of equivalent spherical radius); bubble shape; stiffness of the polyethylene bubble wall; relative positions of the bubbles in the schools, including sway and warp; and depth and tilt of the bubble frame.

TABLE I. Model parameters requiring iterative selection to obtain best fits to the bubble school measurements.

Variable	Sparse (initial)	Sparse (final)	Medium	Dense
Bubble Radius (ml)	0.48	0.48	0.60	0.77
C.V. of Radius	0.043	0.086	0.067	0.052
Shape	1/1	16/1	14/1	12/1
Stiffness (Pa s)	20	200	200	200
Sway (cm)	2	2	2	2
Warp (cm)	6	6	9	12
Depth (m)	96	96	96	95
Tilt (deg)	2.9	2.9	3.5	10.8

Each of these eight parameters, except for the stiffness of the polyethylene bubble wall, has been discussed in the previous section. The bubble wall actually consists of an elastic material. However, for both physical and mathematical reasons that are discussed in Ref. 3, fish flesh, in the small school model, is approximated as a viscous material. Therefore, since it is the small school model that is being evaluated in the present work, the bubble wall has been represented phenomenologically by a highly viscous material.

Setting the values for most of the eight parameters was a two-step process. The first step was to set limits or initial values for each parameter based on the physical observations and other information (Table I, first column). The second step was to run the small school scattering model for the sparse school of 27 bubbles using values chosen in the first step. Values for one parameter were varied within the set limits until the best fit to the measured target strengths of the sparse school was obtained within the analyzable frequency range. The parameters were varied in turn, starting with those in which there was the most confidence. The process was iterated until the best match to the data was obtained and a final set of parameter values chosen (Table I, second column). These values provided the basis for modeling of the dense and medium schools.

The sparse school was chosen to set values of the parameters because it was assumed that interactions between the bubbles would be least for this school. Interactions between the bubbles would have been even less if the target strengths of an eight (i.e., $2\times2\times2$ cube) bubble school, or even a single bubble, could have been measured. Time constraints did not allow such measurements but, more significantly, scattering from the bubble frame would probably have rendered such measurements invalid.

The values of two of the eight parameters were fixed prior to running the model. The bubble school depth was set based on beam angle and echosounder slant range calcula-

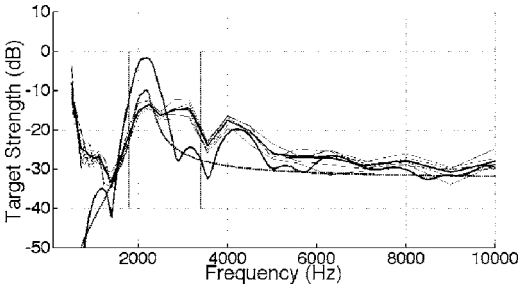


FIG. 5. Comparison of target strength data and model simulation. The initial values of the bubble and school parameters for the sparse school (first column of Table I) are used. Data curves, individual shots and the average, as in Fig. 4. Model curves: Thick dotted curve, model prediction assuming 27 noninteracting bubbles; and thick solid curve, model prediction including radiative coupling interactions.

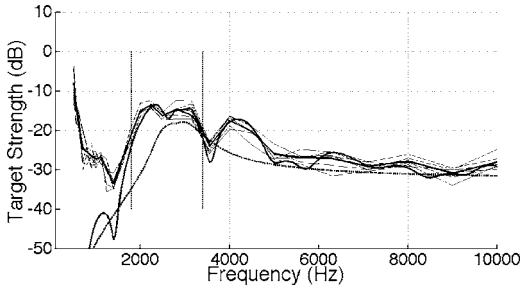


FIG. 6. Comparison of target strength data and model simulation. The final values of the bubble and school parameters for the sparse school (second column of Table I) are used. Curves as in Fig. 5.

tions. The average bubble volumes for each school were determined by assuming that the volumes at the completion of the measurements were half of their original value, and that the gas in the bubbles was lost uniformly with time while the schools were at the measurement depth. The resulting average bubble volumes at depth were estimated as 0.77, 0.60, and 0.48 ml, for the 729, 125, and 27, bubble cases, respectively. Although the bubble volumes changed during a measurement sequence, resulting changes in the resonance frequency of an individual bubble from the first to last shot in a sequence were calculated to be significantly less than the 1/6-octave resolution of the data analysis. Therefore, the volume changes during a sequence had a minimal effect on measured school target strengths, and the average bubble volumes were adequate for modeling purposes.

Initial values for the remaining six parameters were chosen as described below. The initial value for the coefficient of variation of bubble radius was set at 0.043, the value obtained from the laboratory measurements made right after the experiment. It was expected that this value would probably be low. The initial bubble shape at depth was assumed to be a sphere, which was progressively deformed into a prolate spheroid of increasing aspect ratio to obtain a final correction factor for the resonance frequency.[22] The initial choice of viscosity of the bubble wall was 20 Pa s, which was based on values used for fish, which range from 10 to 20 Pa s for small mid-water fishes, to 50 Pa s for commercial size fishes.[5–11] The initial value chosen for the sway of the bubbles was 2 cm, the length of the polyethylene tethers. The initial value chosen for the warp of the horizontal planes of bubbles was 6 cm, the mid-point of the range estimated by visual observation. The initial value chosen for the tilt of the bubble school was 2.9°, equivalent to the average angle the school was off the MRA of the F78. As stated above, final values for these parameters in Table I were determined by iteratively running the model for the sparse school.

The model result using the initial values of the bubble and school parameters for the sparse school (i.e., the values shown in the first column of Table I) is shown in Fig. 5. Also shown is the model result obtained for 27 noninteracting bubbles using the same set of parameters as the bubble school. As in Fig. 1, this curve gives the position of the resonance peak of individual bubbles, and implies that the other peaks are the result of coherent interference effects. (Similar curves will be shown in each of the subsequent figures.) The initial parameter values are seen to give rise to a peak at the individual bubble resonance frequency which has a much higher target strength than indicated by the data; while the peak, at about 3200 Hz, had a lower target strength than the data. Since the primary resonance peak was too high, some of the parameter values chosen initially had to be modified.

The best match obtained of model result to measured target strengths for the sparse school is shown in Fig. 6. This match was accomplished by: (a) Changing the shape of the bubbles from spheres to prolate spheroids with major-to-minor axis ratios of 16, which shifted the bubble resonance peak up to about 3000 Hz; (b) increasing the value of the viscosity of the bubble wall material to 200 Pa s (significantly higher than values typically used for fish), which lowered the target strength of the bubble resonance; and, (c) approximately doubling the value of the coefficient of variation of bubble radius to 0.086, which broadened and lowered all the peaks slightly. The final values for sway, warp, and tilt were kept the same as the initial values. Thus, it was found necessary to modify only the bubble parameters, the position parameters did not need to be changed.

The conditions for the sparse school were similar to those used in the initial example shown in Fig. 1(b). The bubble/swimbladder size and depth were about the same, while the ratio of individual scatterer spacing distance (D) to scatterer radius (a) was 80 for the example in Fig. 1(b), and 74 for the sparse school. This implies that the peak near 3000 Hz in the target strength curve for the sparse school can be attributed to the resonance effect of the individual bubbles. The other peaks are due to interference effects. The bubble parameters affect the bubble resonance peak, while the bubble position parameters affect the interference peaks and troughs.

After satisfactorily modeling the scattering from the sparse bubble school, the scattering results for the medium and dense schools were modeled. In these more densely populated schools, the target strengths, in the region of primary interest, were much greater than those of the frame, which could therefore safely be neglected. Of the six parameters that were varied to find the best fit of model to data for the sparse school, only two, the viscosity of the bubble wall, and the sway of the bubbles, were kept exactly the same for the medium and dense schools. For two other parameters,

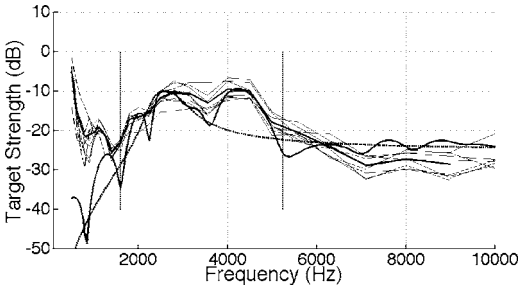


FIG. 7. Comparison of target strength data and model simulation. The final values of the bubble and school parameters for the medium school (third column of Table I) are used. Curves as in Fig. 5.

i.e., the tilt of the school and the warp of the horizontal bubble school planes, the values determined for the sparse school were applied to the medium and dense schools. The tilt angle for all three schools was chosen to be equivalent to the average angle the school was off the MRA of the F78: 3.5° for the medium school; and 10.8° for the dense school. Visual observations of the schools indicated that the warp of the dense school was about twice that of the sparse school, so the warp of the dense school was chosen to be 12 cm. The warp of the medium school was chosen to be the average of the other two, i.e., 9 cm. On-deck observations of the bubble schools showed that the bubbles got progressively smaller and more variable in size after each deployment. Therefore, it was assumed that the coefficient of variation of bubble radius was inversely proportional to bubble size; and was calculated to be 0.052 for the dense school and 0.067 for the medium school. Simulations of the sparse school showed that the coefficient of variation has a minor effect on the target strength, so that small errors in estimating its value do not cause noticeable errors in target strength. It was assumed that the bubbles became thinner as they decreased in volume, so that the minor axis of the prolate spheroid got smaller while the major axis remained constant. Calculations for the dense and medium schools give major-to-minor axis ratios of 12 and 14, respectively, which produce correction factors to the resonance frequency of 1.28 and 1.32, respectively. Target strength curves for the medium and dense bubble schools are shown in Figs. 7 and 8, respectively. Also, because of the large variation in measured target strengths for the dense school at high frequencies, an average target strength curve for the experimental data is shown in Fig. 8.

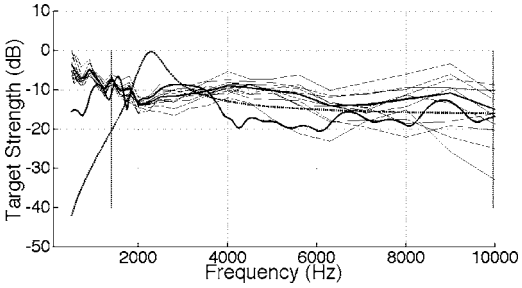


FIG. 8. Comparison of target strength data and model simulation. The final values of the bubble and school parameters for the dense school (fourth column of Table I) are used. Curves as in Fig. 5.

The simulated target strengths match the data very well for the medium bubble school. Figure 7 shows that both the frequencies and levels of the major peaks are in excellent agreement. The individual bubble resonance peak is near 2600 Hz. The conditions for the medium school were similar to those used in the example shown in Fig. 1(c). The bubble/swimbladder size and depth were about the same; and D/a was 40 for the example, and 34 for the medium school. It is evident that the peaks near 2000 Hz (cf. the shoulder in the data) and 4200 Hz, in the target strength curve for the medium school can be attributed to interference effects.

Figure 8 shows that the simulated target strengths for the dense bubble school match the data well from 1400 to about 3500 Hz, and are close to the average data curve at 9000–10 000 Hz. From 4000 to 8000 Hz, the simulated target strengths are well below the data, being about 10 dB below the average data curve from 4000 to 5600 Hz, and about 6 dB below from 6300 to 8000 Hz. The conditions for the dense school were somewhat similar to those used in the example shown in Fig. 1(d). The bubble size was slightly larger than the swimbladder size used in the example, but the depths were about the same; and D/a was 20 for the example, and 16 for the medium school. The difference in D/a ratios shifts the interference peaks and troughs shown in the example by a factor of 1.25 for the dense school. For example, the peak near 4000 Hz in the example should be near 5000 Hz for the dense school. The individual bubble resonance peak is near 2300 Hz. This peak has been effectively lowered and shifted to lower frequencies by interference and coupled-resonance effects. Similar effects are seen in Fig. 1(d).

VI. DISCUSSION

Examination of the small school model simulations shown in Figs. 1 and 5–8 shows three distinct frequency regimes within the analyzable frequency ranges. The first is the region close to the resonance of an individual swimbladder. The second is the regime of strong coherent interference, which can occur both above and below the resonance of an individual swimbladder. The third is the high frequency regime, where the target strength of a school approaches that of a comparable number of widely spaced (non acoustically interacting) fish.

Varying the parameters of the sparse school within rather well defined limits, that were generally based on physical observations of the schools, has been seen to produce model results in good agreement with the target strength data for the resonance and strong interference regimes (Fig. 6).

The parameters used for the medium school were based on those used for the sparse school. The model results also fit the school target strength data very well below and above resonance, and in the regime of strong coherent interference (Fig. 7).

The parameters used for the dense school were also based on those used for the sparse school. The model results fit the school target strength data very well near, and somewhat above, resonance, up to 3600 Hz (Fig. 8). The data

above 4000 Hz are quite variable, primarily due to drag on the school, which had two effects. First, different tilt angles were encountered from shot to shot. Second, the school was pushed away from the MRA of the F78 receiver. However, it would be expected that the average of ten measurements should provide a reasonably accurate estimate of the school target strength. The model and the average of the data agree in the high frequency regime, i.e., 9000–10 000 Hz. Agreement is poor in the vicinity of the first interference peak above resonance, which should be at about 5000 Hz. Figure 1(d) indicates that this peak should not be very strong, but it would not be expected to fall below the curve for 729 widely spaced bubbles. Thus, the combination of parameters used for modeling the dense school has produced a broad trough where a broad, albeit weak, peak would be expected. Simulations with other combinations of parameters were tried, and it was found possible to raise the modeled target strengths to the level of the data in the 4000–8000 Hz region for some cases, e.g., by zeroing the tilt angle and removing any variation in the bubble positions. There is no clear physical justification for using these combinations, but the drag on the bubble school possibly gave rise to some effects that can not be accounted for. For example, drag could have reduced the sway of the individual bubbles or changed the warp of the bubble planes, possibly making them asymmetrically concave, rather than symmetrically convex, as has been assumed. Such hypothetical combinations were not explored further. However, the fact that some combinations of parameters could raise the modeled school target strength in the 4000–8000 Hz range indicates that the differences between model simulation and measurement in this frequency range may not be caused by inherent faults in the model but rather, by the inability to accurately determine the exact parameters of the dense school.

Target strength measurements from the three bubble schools demonstrate that using a scattering model which assumes incoherent addition of the pressure fields from individual bubbles (or swimbladders) is inadequate to estimate target strengths of small closely spaced schools in frequency regimes near the individual bubble (or swimbladder) resonance frequency, and when strong interference effects occur. Model simulations show that the small school model provides reasonable approximations of small school target strengths in these regimes, as long as good values for the school parameters can be determined. Since the small school model incorporates the individual swimbladder model, it is applicable at high frequencies and large scatterer spacings. However, this present work indicates that scattering levels of closely spaced schooling fish in the mid-frequency range (5–10 kHz) are likely to be complicated by coherent interaction effects. Future modeling efforts aimed at addressing the problem of acoustic identification of fish *in situ* will need to consider these effects. In addition, scattering models which incorporate both fish swimbladders and fish bodies as complex objects,[24,25] when they are applied in this mid-frequency regime, will need to consider the complications introduced by coherent interactions.

The model-data comparisons analyzed in this work corroborate previous experimental observations of acoustic scattering from oceanic schooling fish.[13–15] In particular, the fluctuations seen in the scattering levels as a function of frequency appear to be real phenomena which result from mechanisms that are reproducible using physical modeling, i.e., coupled resonance and coherent interaction effects.

Although the present measurements on small bubble schools could not address the issue, previous experience in the analysis of scattering measurements from large oceanic schools of fish[13–15] leads us to expect that, where attenuation of sound through the school is significant, the small school model should accurately predict the variability of the scattering, but likely overestimate the overall level.

VII. SUMMARY

Broad band target strength measurements on three "schools" of polyethylene bubbles were made at sea to test the validity of a theoretical model of scattering from small schools of swimbladder-bearing fish.[17] A "dense" school was constructed using 729 bubbles spaced approximately 9 cm apart in a cubic lattice. A "medium" school was constructed of 125 bubbles approximately 18 cm apart, and a "sparse" school was constructed of 27 bubbles approximately 36 cm apart. Bubbles in the schools were cut from sheets of "bubble wrap." At the measurement depth of about 96 m, the bubbles had an equivalent spherical diameter of approximately 1 cm. Measurements were made using explosive sources detonated just below the sea surface, and a downward looking directional receiver.

The small school model predicts that scattering from individual or widely spaced fish, which is typically dominated by resonance scattering from the swimbladder, will be modified by both interference and coupled-resonance effects, as the fish swim closely together. Spacings between fish in a school are generally one to several body lengths apart. Spacings in the bubble schools were equivalent to fish spacings of about 0.8–4 body lengths.

Theoretically, the ideal bubble school would contain spherical bubbles, all the same size, in a rigid geometrically uniform lattice, placed normal to a co-located source and receiver. Practically, in the at-sea measurements, the bubbles had a range of sizes, and were not spherical. Their positions varied from a rigid uniform lattice, and the source and receiver were not co-located. The effects of variable size and position are not all negative, e.g., they cause the bubble schools to more closely resemble real schools of fish. Because of the variability, several parameters of the individual bubbles and their positions had to be determined. Limits were placed on these parameters, usually based on physical observations, and iterative model simulations were run on the sparse school to obtain the set of parameters which gave the best fit of model to target strength data. These parameters were then used as a basis for the model simulations of the medium and dense schools.

Comparisons of model simulations to target strength data show that, if the proper school parameters can be determined, the small school scattering model appropriately accounts for the resonance action of the individual swimbladders, coupled-resonance interactions between the scatterers,

and the coherent interactions of the scattered fields. The model is also applicable at high frequencies, and for widely dispersed fish ensembles.

ACKNOWLEDGMENTS

The authors thank the Captain and crew of the R/V Gyre for their assistance at sea. This work was supported by the Office of Naval Research, Program Element 62435N.

[1]*Proceedings of an International Symposium on Biological sound Scattering in the Ocean*, Rep. 005, edited by G. B. Farquhar (Maury Center for Ocean Science, Washington, DC, 1970).
[2]I. B. Andreeva, "Scattering of sound by air bladders of fish in deep sound scattering layers," Sov. Phys. Acoust. **10**, 17–20 (1964).
[3]R. H. Love, "Resonant acoustic scattering by swimbladder-bearing fish," J. Acoust. Soc. Am. **64**, 571–580 (1978).
[4]R. H. Love, "Predictions of volume scattering strengths from biological trawl data," J. Acoust. Soc. Am. **57**, 300–306 (1975).
[5]R. H. Love, "A comparison of volume scattering strength data with model calculations based on quasisynoptically collected fishery data," J. Acoust. Soc. Am. **94**, 2255–2268 (1993).
[6]R. W. Nero and M. E. Huster, "Low-frequency acoustic imaging of Pacific salmon on the high seas," Can. J. Fish. Aquat. Sci. **53**, 2513–2523 (1996).
[7]R. W. Nero, C. H. Thompson, and R. H. Love, "Abyssopelagic grenadiers: the probable cause of low frequency scattering at great depths off the Oregon and California coasts," Deep-Sea Res., Part I **44**, 627–645 (1997).
[8]R. W. Nero, C. H. Thompson, and R. H. Love, "Low frequency acoustic measurements of Pacific hake, *Merluccius productus*, off the west coast of the United States," Fish. Bull. **96**, 329–343 (1998).
[9]R. H. Love, C. H. Thompson, and R. W. Nero, "Changes in volume reverberation from deep to shallow water in the eastern Gulf of Mexico," J. Acoust. Soc. Am. **114**, 2698–2708 (2003).
[10]R. H. Love, R. A. Fisher, M. A. Wilson, and R. W. Nero, "Unusual swimbladder behavior of fish in the Cariaco Trench," Deep-Sea Res., Part I **51**, 1–16 (2004).
[11]R. W. Nero, C. H. Thompson, and J. M. Jech, "*In situ* acoustic estimates of the swimbladder volume of Atlantic herring (*Clupea harengus*)," ICES J. Mar. Sci. **61**, 323–337 (2004).
[12]T. J. Pitcher and J. K. Parrish, "Functions of shoaling behavior in teleosts," in *Behavior of Teleost Fishes*, edited by T. J. Pitcher (Chapman and Hall, New York, 1993), 2nd ed., pp. 363–439.
[13]B. S. McCartney, "Underwater sound in oceanography," in *Underwater Acoustics*, edited by V. M. Albers (Plenum, New York, 1967), Vol. **2**, Chap. 10, pp. 185–201.
[14]J. A. Scrimger, R. G. Turner, and G. J. H. Heyd, "Backscattering of underwater sound in Saanich Inlet, British Columbia, including observations of scattering from a fish school," J. Acoust. Soc. Am. **51**, 1098–1105 (1972).
[15]D. V. Holliday, "Resonance structures in echoes from schooled pelagic fish," J. Acoust. Soc. Am. **51**, 1322–1332 (1972).
[16]R. H. Love, "A model for estimating distributions of fish school target strengths," Deep-Sea Res., Part A **28A**, 705–725 (1981).
[17]C. Feuillade, R. W. Nero, and R. H. Love, "A low-frequency acoustic scattering model for small schools of fish," J. Acoust. Soc. Am. **99**, 196–208 (1996).
[18]O. Diachok, "Effects of absorptivity due to fish on transmission loss in shallow water," J. Acoust. Soc. Am. **105**, 2107–2128 (1999).
[19]C. Devin, "Survey of thermal, radiation, and viscous damping of pulsating air bubbles in water," J. Acoust. Soc. Am. **31**, 1654–1667 (1959).
[20]C. S. Clay and H. Medwin, *Acoustical Oceanography* (Wiley, New York, 1977), p. 198.
[21]M. Strasberg, "The pulsation frequency of nonspherical gas bubbles in liquids," J. Acoust. Soc. Am. **25**, 536–537 (1953).
[22]D. E. Weston, "Sound propagation in the presence of bladder fish," in *Underwater Acoustics*, edited by V. M. Albers (Plenum, New York, 1967), Vol. **2**, Chap. 5, pp. 55–88.
[23]C. Feuillade and M. F. Werby, "Resonances of deformed gas bubbles in liquids," J. Acoust. Soc. Am. **96**, 3684–3692 (1994).
[24]C. S. Clay, "Composite ray-mode approximations for backscattered sound from gas-filled cylinders and swimbladders," J. Acoust. Soc. Am. **92**, 2173–2180 (1992).
[25]C. S. Clay and J. K. Horne, "Acoustic models of fish: The Atlantic cod (Gadus morhua)," J. Acoust. Soc. Am. **96**, 1661–1668 (1994).

Acoustic positioning using a tetrahedral ultrashort baseline array of an acoustic modem source transmitting frequency-hopped sequences

Pierre-Philippe J. Beaujean[a)]
Department of Ocean Engineering, Florida Atlantic University, SeaTech Campus, 101 North Beach Road, Dania Beach, Florida 33004

Asif I. Mohamed[b)] and Raphael Warin[c)]
Florida Atlantic University, 101 North Beach Road, Dania, Florida 33004

(Received 19 July 2006; revised 18 October 2006; accepted 22 October 2006)

Acoustic communications and positioning are vital aspects of unmanned underwater vehicle operations. The usage of separate units on each vehicle has become an issue in terms of frequency bandwidth, space, power, and cost. Most vehicles rely on acoustic modems transmitting frequency-hopped multiple frequency-shift keyed sequences for command-and-control operations, which can be used to locate the vehicle with a good level of accuracy without requiring extra signal transmission. In this paper, an ultrashort baseline acoustic positioning technique has been designed, simulated, and tested to locate an acoustic modem source in three dimensions using a tetrahedral, half-wavelength acoustic antenna. The position estimation is performed using the detection sequence contained in each message, which is a series of frequency-hopped pulses. Maximum likelihood estimation of azimuth and elevation estimation is performed using a varying number of pulse and various signal-to-noise ratios. Simulated and measured position estimation error match closely, and indicate that the accuracy of this system improves dramatically as the number of pulses processed increases, given a fixed signal-to-noise ratio. © *2007 Acoustical Society of America.*
[DOI: 10.1121/1.2400616]

PACS number(s): 43.30.Tg, 43.30.Wi [EJS] Pages: 144–157

I. INTRODUCTION

Ultrashort baseline (USBL) acoustic positioning systems (APSs) are well-suited for unmanned underwater vehicle (UUV) operations because of their small space requirement and ease of deployment. This type of positioning system is based on the estimation of the relative phases between signals measured by a set of transducers, each a half-wavelength from the others. The accuracy in estimating these relative phases is a function of several limiting factors. Reverberation may cause major signal distortion in the acoustic signals received by the array and limit the accuracy of time delay estimation during position estimation.[1] Another limiting factor is the ambient noise present in shallow and deep waters generated by waves, rain, boat traffic, and biological life.[2]

Although USBL APS technology is a relatively mature field,[3,4] new challenges have appeared due to the limitation of frequency bandwidth available during a typical mission. Indeed, most APSs operate between 5 and 50 kHz, a frequency band also used by subbottom profilers and most acoustic modems. Also, acoustic communication systems (ACOMSs) tend to use an ever increasing frequency bandwidth, as the need for data transfer and networking keeps growing.[5,6] Finally, UUV technology includes an increasing number of small vehicles with limited energy capacity, which can accommodate only a very limited number of sensors. Since positioning and communications are vital aspects of underwater robotics, technology capable of combining acoustic communications and positioning is of tremendous interest.[7]

An essential aspect of combining ACOMSs and APSs is in understanding how an acoustic modem typically operates. Ideally, the APS should remain transparent to the ACOMS: it should operate with little or no modification to the acoustic communication protocol and the modulation, and should not impair the performance of the ACOMS. The USBL-APS approach is well suited as it relies solely on the estimation of the time of travel between source and receiver, and the estimation of the azimuth and elevation of the source with respect to the receiver. ACOMS technology provides features to measure the time of travel between the source and the receiver. The challenging task of accurately estimating the azimuth and elevation angles can be completed using a multiple channel array and processing a known portion of the incoming signal.

Although many signaling techniques are used in ACOMSs, the most commonly encountered are known as frequency-hopped multiple frequency shift keying[8,9] (FH-MFSK) and direct-sequence spread spectrum[10,11] (DSSS). FH-MFSK has by far proven to be the most robust technique, although it does not allow fast data transmissions, and is commonly used in ACOMSs to monitor and control underwater vehicles. A very specific aspect of a FH-MFSK modu-

[a)]Electronic mail: pbeaujea@seatech.fau.edu
[b)]Electronic mail: amoham@seatech.fau.edu
[c)]Electronic mail: rwarin@hotmail.com

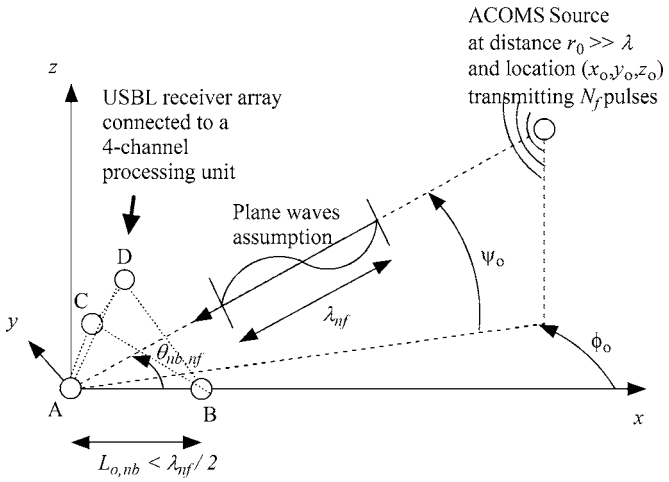


FIG. 1. Overview of the USBL-APS system.

lated signal is that it uses relatively narrow band tones transmitted sequentially and usually includes a known detection sequence at the beginning of each message.[8,9] This detection sequence can be used to estimate the location of the source with a USBL-APS, given the proper receiver array and signal processing technique. This paper introduces a technique to estimate the location of an acoustic modem transmitting FH-MFSK sequences, and is focused primarily on an approach of estimating the azimuth and elevation angles.

The method is based on the maximum-likelihood estimation[12,13] and uses some relationships developed by Quazi and Lerro.[14] Although previous work has been published on maximum likelihood estimation of source location,[15–17] no attempt has been made to estimate the location of a source from FH-MFSK sequences using a three-dimensional, half-wavelength USBL array. Synchronous GPS-based time-of-travel estimation is used to estimate the range from the source to the array.[9] Alternately, the positioning unit can interrogate each modem, in which case the range is estimated using the two-way travel time. A USBL-APS compatible with dual purpose acoustic modem (FAU-DPAM) operations[9] is introduced, along with a detailed analysis of experimental data.

II. OVERVIEW OF THE USBL APS

A. System description

Figure 1 shows an overview of the USBL-APS operation to estimate the position of a modem source. The source transmits $N_f=8$ pulses at the beginning of each message, at 13.54 ms intervals. These pulses last 620 μs each and use different frequency bands staggered between 16.2 and 27 kHz. The time-varying spectrum of the transmitted sequence is shown in Fig. 2. The pulses are transmitted sequentially at 24.6, 21, 17.4, 27, 23.4, 19.8, 16.2, and 25.8 kHz.

The choice of the USBL array is driven by frequency and equipment constraints: The system must process pulses centered between 16.2 and 27 kHz, while the acquisition system of the APS unit can only accommodate up to four channels. According to Ray and Mahajan,[18] when four hydrophones are used, a tetrahedral configuration is optimal for three-dimensional positioning so long as all four receivers do not to lie on a sphere. For this reason, the acoustic antenna is composed of four transducers (A–D) arranged as a tetrahedron (Fig. 1).

Three of the four transducers form three coplanar baselines (labeled n_b) of 0.025 m spacing ($n_b=1, 2, 3$) and three noncoplanar baselines ($n_b=4, 5, 6$) of 0.024 m spacing. The spacing is the center-to-center distance between two receivers. To prevent ambiguity in resolving the direction of arrival of an incoming signal, the length L_{0,n_b} of each baseline is slightly less than half the shortest wavelength $\lambda_{\min}$ contained in the detection sequence,

$$L_{0,n_b} < \frac{\lambda_{\min}}{2}, \tag{1}$$

$$\lambda_{\min} = c/f_{\max} = 1500/27\,000. \tag{2}$$

The variable c represents the speed of sound in the water in meters per second.

B. Bearing angle estimation

The real-time process used to estimate the bearing angles for each baseline and pulse is given in Fig. 3. An essential portion of the USBL-APS algorithm is the detection of the transmitted signal over noise and interferences. Consider the transmitted signal $s_{n_f}(t)$ of duration T,

$$s_{n_f}(t) = w(t)\cos(2\pi f_{n_f} t), \tag{3}$$

where $w(t)$ is a Tukey time window function,

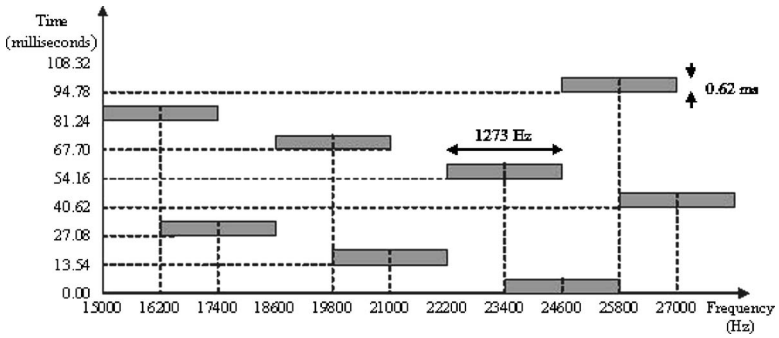


FIG. 2. Time-frequency plot of the transmitted sequence.

$$w(t)=\begin{cases}\dfrac{1}{2}\left(1+\cos\left(\pi\left(\left(t-\dfrac{T}{2}\right)-\dfrac{\beta_T T}{2}\right)\Big/\left((1-\beta_T)T\right)\right)\right) & \text{for } 0\leqslant t\leqslant(1-\beta_T)\dfrac{T}{2} \quad \text{and} \quad (1+\beta_T)\dfrac{T}{2}\leqslant t\leqslant T \\ 1 & \text{for } \dfrac{\beta_T T}{2}\leqslant t\leqslant T-\dfrac{\beta_T T}{2}.\end{cases} \tag{4}$$

The Tukey coefficient β_T is set to 0.25. The transmitted signal can also be expressed in analytical form using a Hilbert transform operation[19] $H\{\ \}$,

$$\tilde{s}_{n_f}(t)=s_{n_f}(t)+iH\{s_{n_f}(t)\}=w(t)e^{i2\pi f_{n_f}t}. \tag{5}$$

The acoustic modem transmits a detection sequence comprising a series of signals $s_{n_f}(t)$, $1\leqslant n_f\leqslant N_f$. The received signals are subject to noise and interferences. We assume that the impulse response of the acoustic channel is stationary, and that the source is located far enough from the receiver so that the acoustic waves are plane. Let j_1 and j_2 be the index of two hydrophones forming the n_bth baseline. For example, if $n_b=1$, j_1 and j_2 correspond to hydrophones A and B, respectively. The received signals at two hydrophones j_1 and j_2 of a baseline are

$$r_{n_f,j_1}(t)=s_{n_f}(t)\otimes h_{j_1}(t)+n_{j_1}(t),$$

$$r_{n_f,j_2}(t)=s_{n_f}(t)\otimes h_{j_2}(t)+n_{j_2}(t), \tag{6}$$

where $h_{j_1}(t)$ and $h_{j_2}(t)$ represent the impulse response of the acoustic channel measured at hydrophones j_1 and j_2, respectively. $n_{j_1}(t)$ and $n_{j_2}(t)$ represent the additive Gaussian noise measured at each hydrophone, of variance $\sigma^2_{\text{noise},j_1}$ and $\sigma^2_{\text{noise},j_2}$ respectively. We assume that the additive noise is isotropic. These signals are matched-filtered with $\tilde{s}_{n_f}(t)$,

$$\tilde{r}_{n_f,j_1}(t)=r_{n_f,j_1}(t)\otimes\tilde{s}_{n_f}(T-t),$$

$$\tilde{r}_{n_f,j_2}(t)=r_{n_f,j_2}(t)\otimes\tilde{s}_{n_f}(T-t). \tag{7}$$

If we assume that a direct path exists between the source and each receiver, and that the time difference between the direct path arrival and the next most immediate path is T_r, cT_r represents the difference in path length between the direct path (assumed to be the shortest) and the shortest interfering path. Over a short time duration, $h_{j_1}(t)$ and $h_{j_2}(t)$ become

$$h_{j_1}(t)=a_{j_1}\delta(t-T_{j_1}),\quad t<T_{j_1}+T_r,$$

$$h_{j_2}(t)=a_{j_2}\delta(t-T_{j_2}),\quad t<T_{j_2}+T_r. \tag{8}$$

T_{j_1} and T_{j_2} represent the time of travel from the source to receivers j_1 and j_2, respectively. The variables a_{j_1} and a_{j_2} represent the signal attenuation at receiver j_1 and j_2, respectively. Therefore, if a time window of duration T_r is applied to the received signal at time T_{j_1}, assuming that $T_{j_1}\leqslant T_{j_2}$ and that $T_{j_2}-T_{j_1}\leqslant T_r$, the time-gated signals received at j_1 and j_2 become

$$\tilde{x}_{n_f,j_1}(t)=a_{j_1}\tilde{s}_{n_f}(t-T_{j_1})+\tilde{n}_{n_f,j_1}(t),$$

$$\tilde{x}_{n_f,j_2}(t)=a_{j_2}\tilde{s}_{n_f}(t-T_{j_2})+\tilde{n}_{n_f,j_2}(t),T_{j_1}\leqslant t<T_{j_1}+T_r. \tag{9}$$

$\tilde{n}_{n_f,j_1}(t)$ and $\tilde{n}_{n_f,j_2}(t)$ represent the in-band complex additive white Gaussian noise at the output of the matched filter. In the experimental results section, $T_r=T=620$ μs, which corresponds to a traveled distance of 0.93 m if the sound speed c is equal to 1500 m/s. If the record does not contain any signal, as is assumed in the initial stage of the detection process, the in-band noise is estimated within the band used by the first pulse. The in-band noise variances $\sigma^2_{\text{noise},n_f,j_1}$ and $\sigma^2_{\text{noise},n_f,j_2}$ at hydrophones j_1 and j_2 are

$$\sigma^2_{\text{noise},n_f,j_1}=\frac{1}{T_n}\int_0^{T_n}|\tilde{n}_{n_f,j_1}(t)|^2dt,$$

$$\sigma^2_{\text{noise},n_f,j_2}=\frac{1}{T_n}\int_0^{T_n}|\tilde{n}_{n_f,j_2}(t)|^2dt. \tag{10}$$

T_n represents the time duration over which the in-band noise variance is estimated. In practice, the calculation is performed using a time-moving average of sequential noise variance estimates measured over a fixed duration. At a minimum, T_n is equal to 0.1 s. Once T_n seconds have passed, the APS units starts searching for an incoming signal on each receiver. The signal-to-noise ratio (SNR) measured at each hydrophone j_1 and j_2 is

$$\text{SNR}_{n_f,j_1}=\frac{\dfrac{1}{T_r}\displaystyle\int_0^{T_n}|\tilde{x}_{n_f,j_1}(t)|^2dt}{\sigma^2_{\text{noise},n_f,j_1}},$$

$$\text{SNR}_{n_f,j_2}=\frac{\dfrac{1}{T_r}\displaystyle\int_0^{T_n}|\tilde{x}_{n_f,j_2}(t)|^2dt}{\sigma^2_{\text{noise},n_f,j_2}}. \tag{11}$$

Signal detection occurs if the SNR, measured within the frequency band of the first pulse, exceeds a predefined threshold T_{SNR} at every receiver A, B, C, and D. The phase difference α_{n_b,n_f} between the signals measured at receiver j_1 and j_2 is

$$\alpha_{n_b,n_f}=\arctan\left(\frac{\mathcal{I}\left\{\displaystyle\int_0^{T_t}\tilde{x}_{n_f,j_1}(t)\tilde{x}^*_{n_f,j_2}(t)dt\right\}}{\mathcal{R}\left\{\displaystyle\int_0^{T_t}\tilde{x}_{n_f,j_1}(t)\tilde{x}^*_{n_f,j_2}(t)dt\right\}}\right). \tag{12}$$

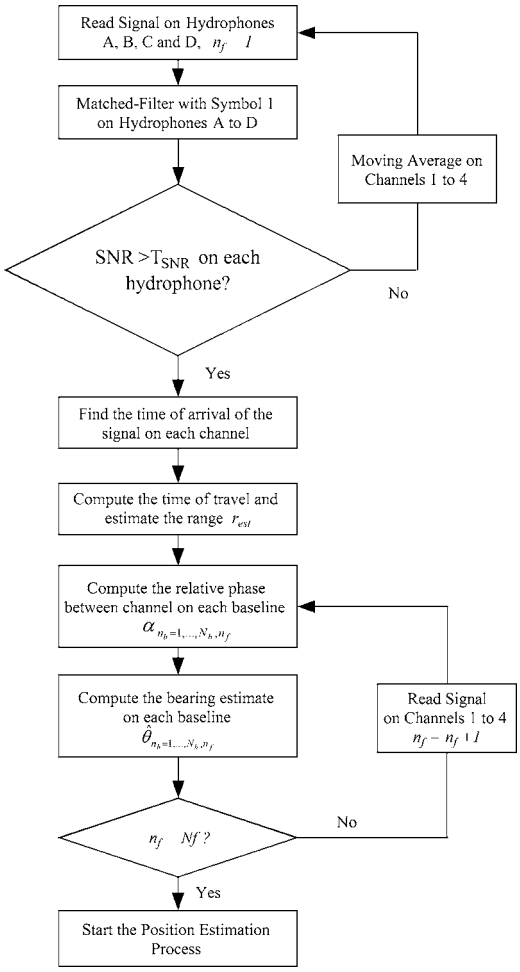


FIG. 3. Flow-chart of the bearing estimation process.

The same procedure is used to estimate the phase difference α_{n_b,n_f} between the signals recorded at each receiver of baseline n_b. If the incoming acoustic waves are assumed to be plane, the bearing angle $\hat{\theta}_{n_b,n_f}$ measured at the n_bth baseline for the n_fth pulse is obtained from α_{n_b,n_f},

$$\hat{\theta}_{n_b,n_f} = \arccos\left(\frac{c\alpha_{n_b,n_f}}{2\pi f_{n_f} L_{0n_b}}\right). \tag{13}$$

c represents the speed of sound and f_{n_f} is the central frequency of the pulse. With N_b=6 baselines for the tetrahedral USBL array and N_f=8 pulses, a total of 48 distinct bearing angles can be calculated. The next step is to determine the azimuth Φ_0 and elevation Ψ_0 of the source with respect to the USBL array given the 48 bearing angles.

C. Range estimation

The time of transmission of the pulse is assumed to be known, as both source and receiver operates with low-drift clocks, initially synchronized using a 1 μs GPS pulse.[9] Therefore, the time of transmission is known and the time of reception must be accurately estimated. The transmitted pulse is assumed to be of sufficient bandwidth W so that the direct path between source and receiver is only distorted by ambient noise. A classic maximum likelihood estimation technique is used to estimate the time of arrival on channel j_1 from the first pulse,[20]

$$\mathcal{R}\left\{\int_0^T (\tilde{x}_{1,j_1}(t) - \tilde{s}_{1,j_1}(t-\tilde{\tau}))\frac{\partial \tilde{s}^*_{1,j_1}(t-\tilde{\tau})}{\partial \tilde{\tau}}dt\right\}_{\tilde{\tau}=\tau_{\text{est}}} = 0. \tag{14}$$

At present, a single hydrophone and pulse are used, however the estimation from multiple hydrophones and pulses could be used in the future to obtain a more accurate time estimate. The range estimate r_{est} is obtained directly from τ_{est},

$$r_{\text{est}} = c\tau_{\text{est}}. \tag{15}$$

III. POSITION ESTIMATION

A. Variance of the bearing estimate

Azimuth and elevation of the transmitter with respect to the USBL are estimated based on the maximization of the log-likelihood function of estimating the proper source directions. We first assume that the 48 bearing angles estimates are biased Gaussian random variables,

$$\hat{\theta}_{n_b,n_f} = G(\theta_{n_b,t}, \sigma_{\theta_{n_b,n_f}}) + \theta_{\text{bias}_{n_b,n_f}} + \theta_{\text{curve}_{n_b}}. \tag{16}$$

$\theta_{n_b,t}$ is the true bearing angle for each baseline, and does not depend on frequency. The length of each baseline is a Gaussian random variable of mean value L_{0,n_b} and variance σ_L^2. $\theta_{\text{bias}_{n_b,n_f}}$ is a bias in the estimate resulting from the nonlinear relationship between L_{0,n_b} and $\hat{\theta}_{n_b,n_f}$. $\theta_{\text{curve},n_b}$ is a small error associated with the curvature of the wave front, as the bearing is computed under plane wave assumption. If we assume that the estimation error is solely due to signal noise and positional uncertainty of the sensors, the variance of the bearing angle estimates is given by,[14]

$$\sigma^2_{\theta_{n_b,n_f}} = \left(\cot^2(\theta_{n_b,t})\frac{\sigma_L^2}{L^2_{o,n_b}}\right) + \left(\frac{c^2\left(\frac{1+\text{SNR}_{j_1,n_f}+\text{SNR}_{j_2,n_f}}{\text{SNR}_{j_1,n_f}\text{SNR}_{j_2,n_f}}\right)}{8\pi^2 L^2_{o,n_b}(TW)\left(f^2_{n_f}+\frac{W^2}{12}\right)\sin^2(\theta_{n_b,t})}\right). \tag{17}$$

The bias $\theta_{\text{bias}_{n_b,n_f}}$ is given by[14]

$$\theta_{\text{bias}_{n_b,n_f}} = -\left((2\cot(\theta_{n_b,t}) + \cot^3(\theta_{n_b,t}))\frac{\sigma_L^2}{2L^2_{o,n_b}}\right) - \left(\frac{c^2\left(\frac{1+\text{SNR}_{j_1,n_f}+\text{SNR}_{j_2,n_f}}{\text{SNR}_{j_1,n_f}\text{SNR}_{j_2,n_f}}\right)\cos(\theta_{n_b,t})}{16\pi^2 L^2_{o,n_b}(TW)\left(f^2_{n_f}+\frac{W^2}{12}\right)\sin^3(\theta_{n_b,t})}\right). \tag{18}$$

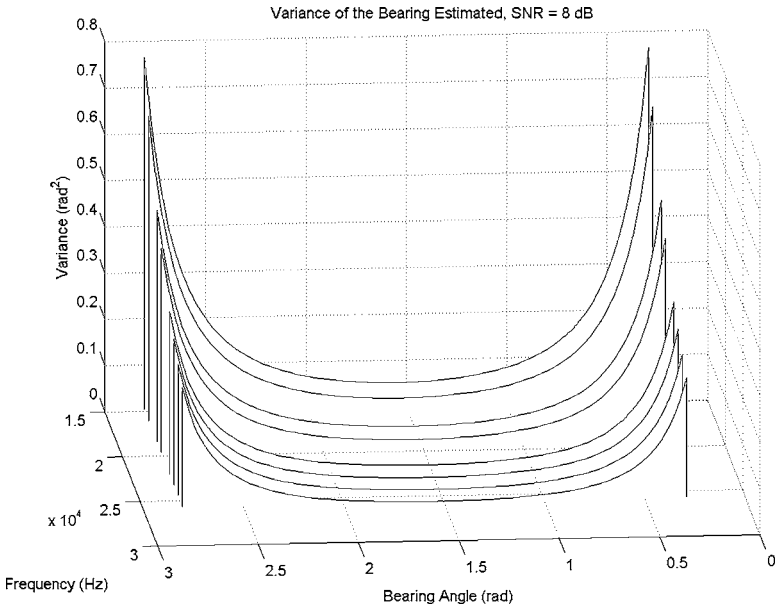


FIG. 4. Variance of the bearing estimate, for $\sigma_L^2=0$ m^2 and SNR=8 dB.

Equations (17) and (18) assume that the true length L_{0,n_b} of each baseline is known. For each pulse, W is the bandwidth and T is the pulse duration. SNR_{j_1,n_f} and SNR_{j_2,n_f} are measured at each hydrophone j_1 and j_2 of baseline n_b for the n_fth incoming pulse. It is assumed that, for all n_f,

$$L_{0,n_b} < \frac{c}{2f_{n_f}}. \tag{19}$$

The positional uncertainty of the sensors is by no means trivial to estimate, as it is a function of two error components, represented as Gaussian random variables of zero mean: (a) the distance inaccuracy between sensor centroids of variance σ_{pos}^2, and (b) the dimension and shape of each sensor of variance σ_{geom}^2. Under the assumption that a large number of trials are completed, σ_L^2 can be rewritten using the central limit theorem as

$$\sigma_L^2 = \sigma_{\mathrm{pos}}^2 + \sigma_{\mathrm{geom}}^2. \tag{20}$$

σ_{pos}^2 and σ_{geom}^2 are constants determined by the quality of the assembly and the shape of the sensor. Figures 4 and 5 show the variance of the bearing estimate, assuming that each sensor is a point of perfectly known location, $\sigma_L^2=0$ m^2, for a SNR of 8 and 40 dB, respectively. The variance is plotted as a function of the true bearing angle $\theta_{n_b,t}$ and the center frequency f_{n_f}. At a given frequency, the variance increases sharply as $\theta_{n_b,t}$ nears 0 and π. The variance of the bearing estimate also increases noticeably as frequency decreases, and as SNR decreases. Overall, given a fixed SNR, the best bearing estimates are obtained near normal incidence and at higher frequencies.

Figures 6 and 7 show the variance of the bearing estimate, taking into account the uncertainty for the sensors location and geometry, $\sigma_L^2=6.25e-6$ m^2, for a SNR of 8 and 40 dB, respectively. The comparison between Figs. 4 and 6 indicates that the impact of the sensor uncertainty is not sig-

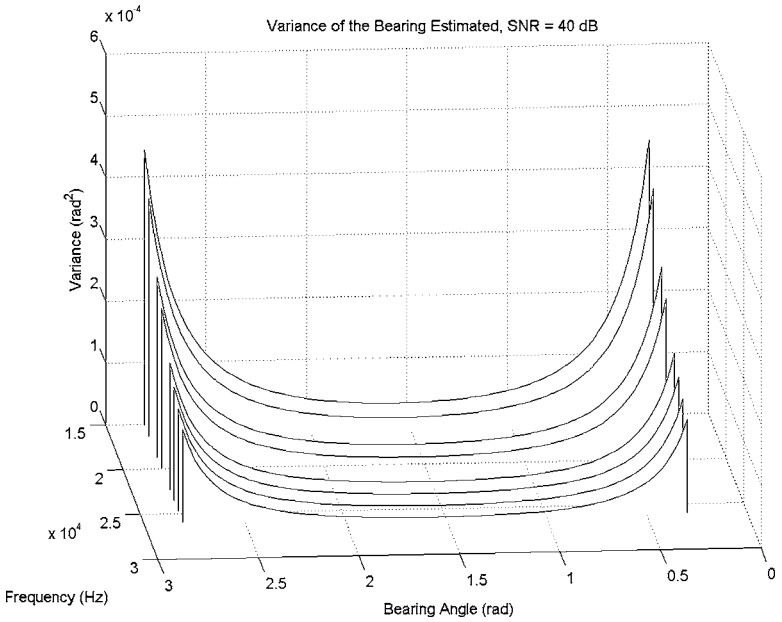


FIG. 5. Variance of the bearing estimate, for $\sigma_L^2=0$ m^2 and SNR=40 dB.

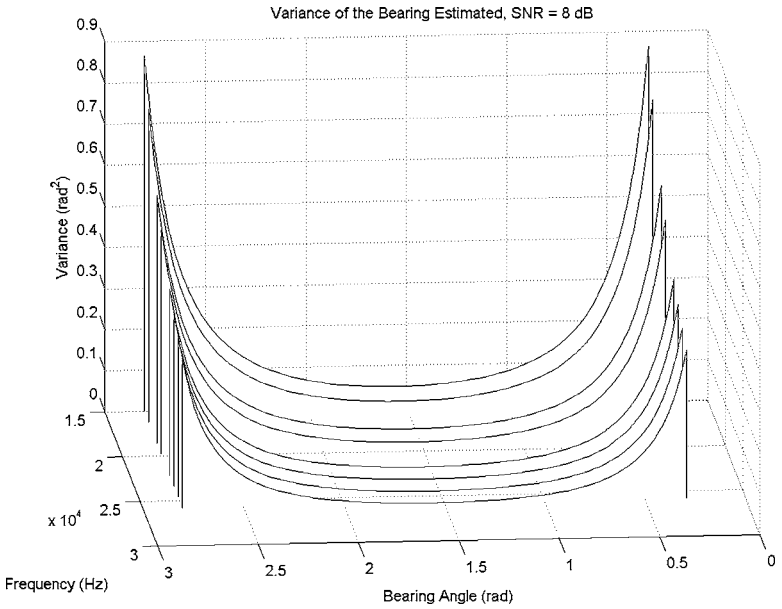


FIG. 6. Variance of the bearing estimate, for $\sigma_L^2 = 6.25e-6$ m^2 and SNR=8 dB.

nificant when the SNR is moderate (8 dB). However, Figs. 5 and 7 show that, at high SNR (40 dB), the geometrical uncertainty is a dominant factor.

Figures 8 and 9 show the bias of the bearing estimate, with $\sigma_L^2 = 6.25e-6$ m^2, for a SNR of 8 and 40 dB, respectively. With the exception of normal incidence, the bias on the estimate is significant when compared with the standard deviation of the bearing estimate, for any value of the bearing angle.

The curvature error $\theta_{\text{curve}_{n_b}}$ can be estimated geometrically. Under the assumption that a point source is used and that $L_{0,n_b} \ll r_0$, $\theta_{\text{curve}_{n_b}}$ becomes

$$\theta_{\text{curve}_{n_b}} = -\theta_{n_b,t} + \arccos\left(\frac{\sin(\theta_{n_b,t})}{\sin\left(\arctan\left(\frac{L_{0,n_b}\sin(\theta_{n_b,t})}{r_0 + L_{0,n_b}\cos(\theta_{n_b,t})}\right)\right)} - \frac{r}{L_{0,n_b}}\right). \quad (21)$$

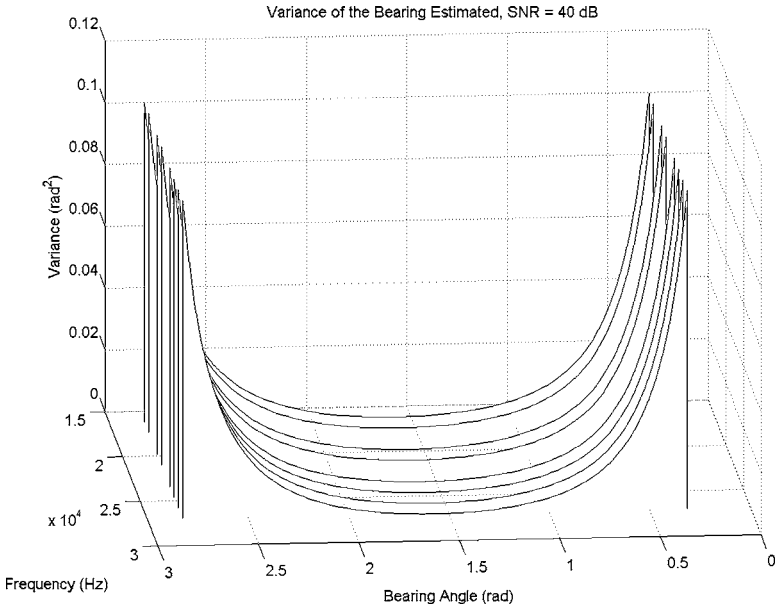


FIG. 7. Variance of the bearing estimate, for $\sigma_L^2 = 6.25e-6$ m^2 and SNR=40 dB.

Figure 10 shows the error in bearing estimate due to the curvature of the wave front, assuming that the source is located at $r_0 = 0.62$ m, which is the distance from the source to the center of the array in the calibration tank presented in Sec. IV. As expected, the largest discrepancy occurs near normal incidence. The error remains within 0.02 rad, which indicates that plane wave approximation is a reasonable assumption under the experimental conditions.

B. Variance of the range estimate

The range is estimated from the sensor labeled $A(j_1 = A)$, as shown in Fig. 1. The accuracy of the time of travel estimate from the source to the sensor is a function of the SNR and pulse characteristics. Since the APS is a passive system, and since the SNR is assumed to be larger than unity, the range estimate $\hat{r}_{n_f}$ for pulse n_f can be approximated as a Gaussian random variable of true mean r_0 and variance[21]

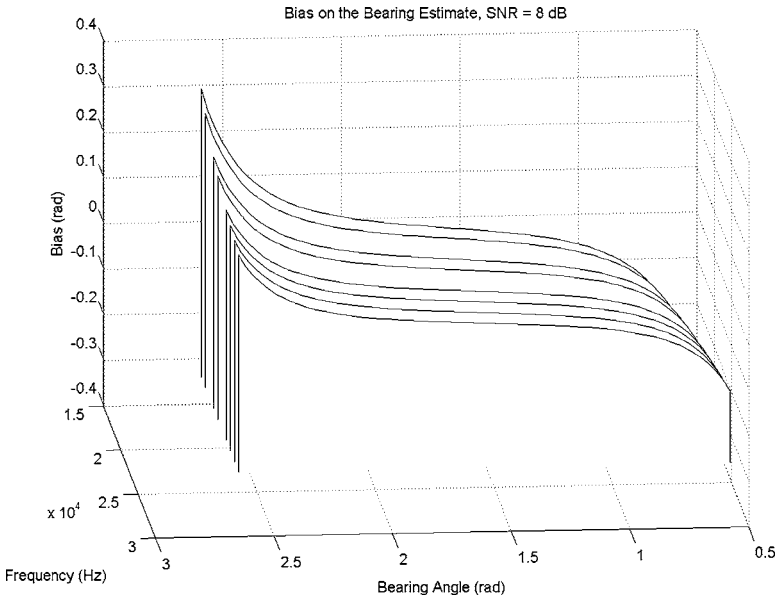


FIG. 8. θ_{bias} for $\sigma_L^2=6.25e-6$ m^2, SNR=8 dB.

$$\sigma_{r_{n_f}}^2 = \frac{c^2}{8\pi^2(TW)\text{SNR}_{j_1=A,n_f}\left(f_{n_f}^2 + \frac{W^2}{12}\right)}. \tag{22}$$

Equation (22) assumes that the range is obtained from maximum likelihood estimation [Eqs. (14) and (15)]. The variance of the estimate given in Eq. (22) is the Cramer-Rao bound.[22] In practice, the first pulse ($n_f=1$) is used for range estimation, noted $r_{\text{est}}=\hat{r}_{n_f=1}$. Figure 11 shows the fluctuation of $\sigma_{r_{\text{est}}}^2$ as a function of the SNR. The Cramer-Rao bound is only true under ideal circumstances, so that $\sigma_{r_{\text{est}}}^2$ is likely to be larger in practice. Nevertheless, Fig. 11 indicates that the range estimation error is most likely to be negligible as compared with the bearing estimation error.

C. Variance threshold of the bearing estimate

Given N_f pulses and N_b baselines, a total of $N_f N_b$ bearings are measured. However, only a limited number of bearings are retained to determine the source location, based on a series of decision rules. For each pulse, a subset of bearing estimates is retained from all N_b bearings measured. The N_f resulting subsets are combined to estimate the azimuth and elevation of the source by maximizing a log-likelihood function, as described in Sec. III D. The following procedure explains how each subset is created, and is repeated for each individual pulse.

First, the bearing $\hat{\theta}_{n_{b\min},n_f}$ with the lowest variance is retained. Next, the algorithm compares the variance of the bearing estimate measured at each baseline coplanar with baseline $n_{b\min}$, and the estimate with the lowest variance is also retained. Finally, the algorithm compares the variance of the bearing estimate measured at each baseline noncoplanar with baseline $n_{b\min}$, and the estimate with the lowest variance is also retained. This operation provides two coplanar and one noncoplanar bearing estimate, which is theoretically sufficient to obtain an unambiguous estimate of the source lo-

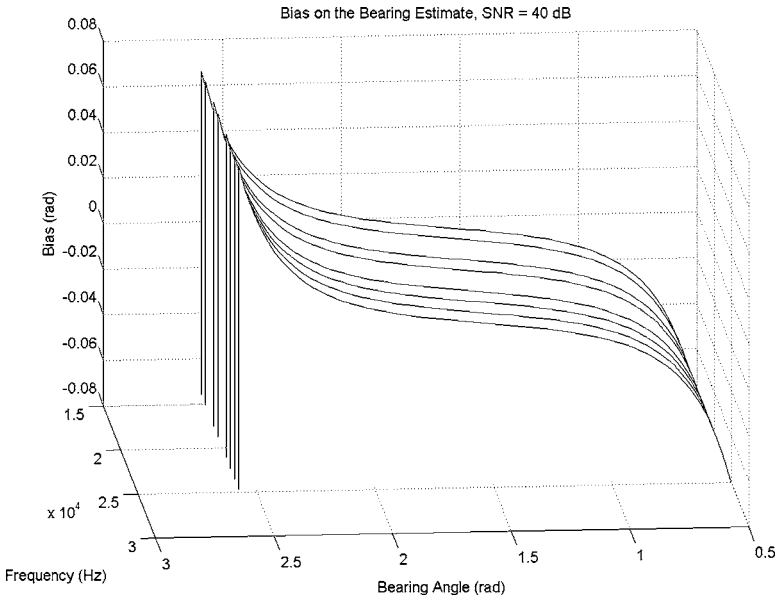


FIG. 9. θ_{bias} for $\sigma_L^2=6.25e-6$ m^2, SNR=40 dB.

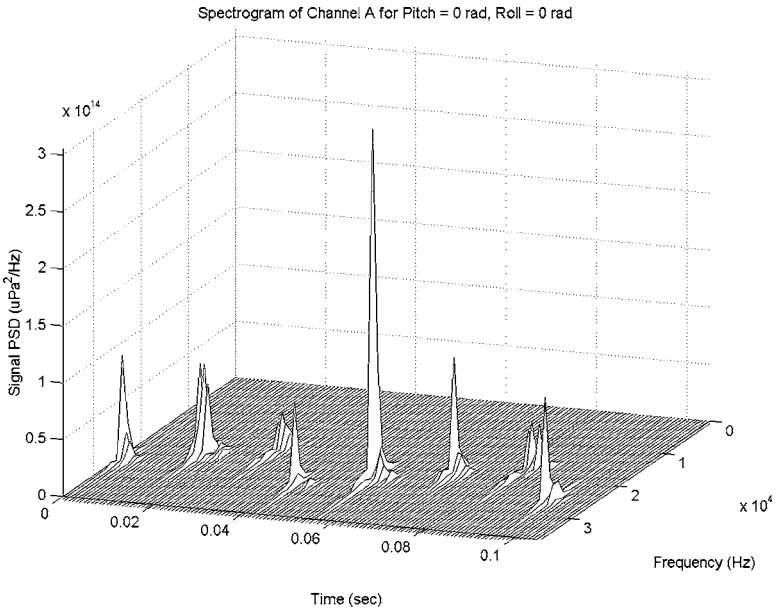


FIG. 10. Curvature error to true bearing ratio, $\theta_{\text{curve}_{n_b}}$ for r_0=0.62 m.

cation. Once three bearing estimates have been retained for each pulse, the mean variance σ^2_{avg} of all N_bN_f bearing estimates is computed,

$$\sigma^2_{\theta_{\text{avg}}} = \frac{\sum_{n_b=1}^{N_b}\sum_{n_f=1}^{N_f}\sigma^2_{\theta_{n_b,n_f}}}{N_bN_f}. \tag{23}$$

Also retained are the bearing estimates whose variance falls within a threshold σ^2_T, linearly proportional to the mean $\sigma^2_{\theta_{\text{avg}}}$ of the bearing variances. The results presented in this paper are obtained using $\sigma^2_T=\sigma^2_{\theta_{\text{avg}}}$. The bearing angles whose variance is larger than the mean variance are placed in the P_{rej} vector and are not included in the log-likelihood summation presented in the following.

D. Log-likelihood function

If we assume that each selected bearing angle is mostly Gaussian, $\hat{\theta}_{n_b,n_f}\approx G(\theta_{n_b,t},\sigma_{\theta_{n_b,n_f}})$, the probability that the set of selected bearing angles $[\hat{\theta}]$ be representative of the true azimuth Φ_0 and elevation Ψ_0 is[12]

$$\Pr([\hat{\theta}]|(\Phi_0,\Psi_0)) = \prod_{\substack{n_b=1,n_f=1\\(n_b,n_f)\notin P_{\text{rej}}}}^{N_b,N_f}\left(\frac{\exp\left(-\frac{(\hat{\theta}_{n_b,n_f}-\theta_{n_b,t})^2}{2\sigma^2_{\theta_{n_b,n_f}}}\right)}{\sigma_{\theta_{n_b,n_f}}\sqrt{2\pi}}\right). \tag{24}$$

The log-likelihood function is the natural logarithm of $\Pr([\hat{\theta}]|(\Phi_0,\Psi_0))$,

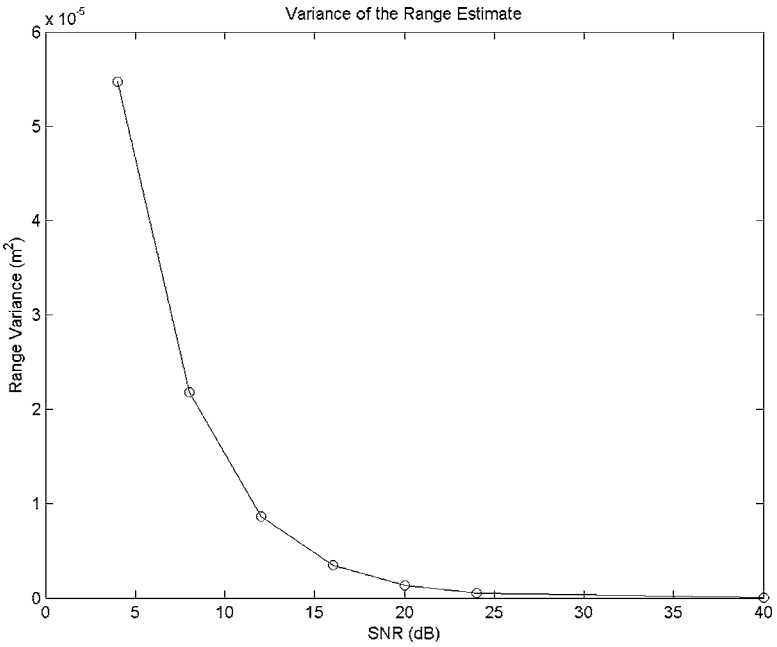


FIG. 11. Range estimation variance $\sigma^2_{r_{\text{est}}}$ as a function of the SNR.

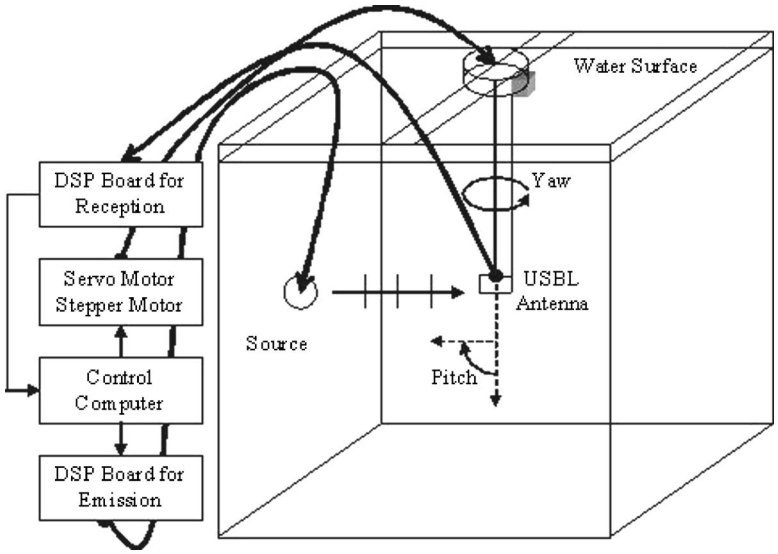


FIG. 12. Test setup.

$$L([\hat{\theta}]|(\Phi_0,\Psi_0)) = -\sum_{\substack{n_f=1,n_b=1 \\ (n_f,n_b)\notin P_{\text{rej}}}}^{N_b,N_f} \left(\ln(\sigma_{\theta_{n_b,n_f}}\sqrt{2\pi}) + \left(\frac{(\hat{\theta}_{n_b,n_f} - \theta_{n_b,t})^2}{2\sigma^2_{\theta_{n_b,n_f}}} \right) \right). \quad (25)$$

The estimated azimuth and elevation pair with the highest probability of being the true azimuth and true elevation corresponds to the maximum of the log-likelihood function. The downhill simplex method is used to find this maximum.[23] However, the log-likelihood function may also have local maxima. Since the reliability of this search method is contingent upon the initial starting region for three search points, it is imperative that these search points be in the vicinity of the true maximum of the log-likelihood function. A first estimate of the azimuth and elevation is computed using the three bearing angles with the lowest variance corresponding to three noncoplanar baselines. Three points are selected around the rough estimate of the azimuth and elevation and used in the maximum search.[12] The outcome of this search process is the estimated azimuth Φ_{est} and the elevation Ψ_{est}.

E. Position estimation error

The performance evaluation of the APS is better described in terms of statistical moments. N_t is the total number of estimation attempts, and n_t is the attempt number. Similarly, N_a is the total number of source locations tested, and n_a is the location index. The positioning accuracy of USBL systems is typically given in terms of position estimation error, which represents the relative error between the estimated source location and the true source location. In Cartesian coordinates, the source location for a true azimuth Φ_0, a true elevation Ψ_0, and a true range r_0 is given by

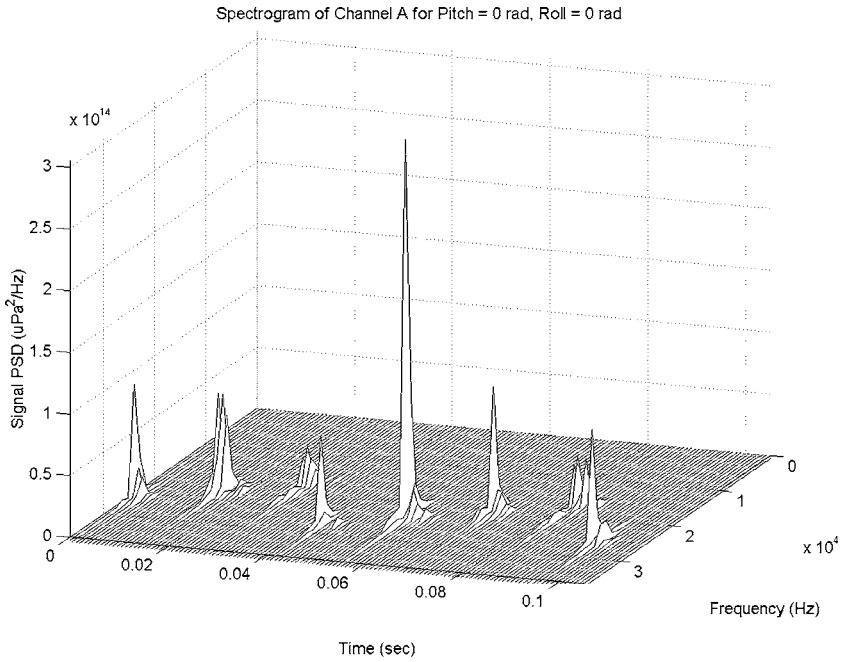


FIG. 13. Time-frequency response of the calibration tank.

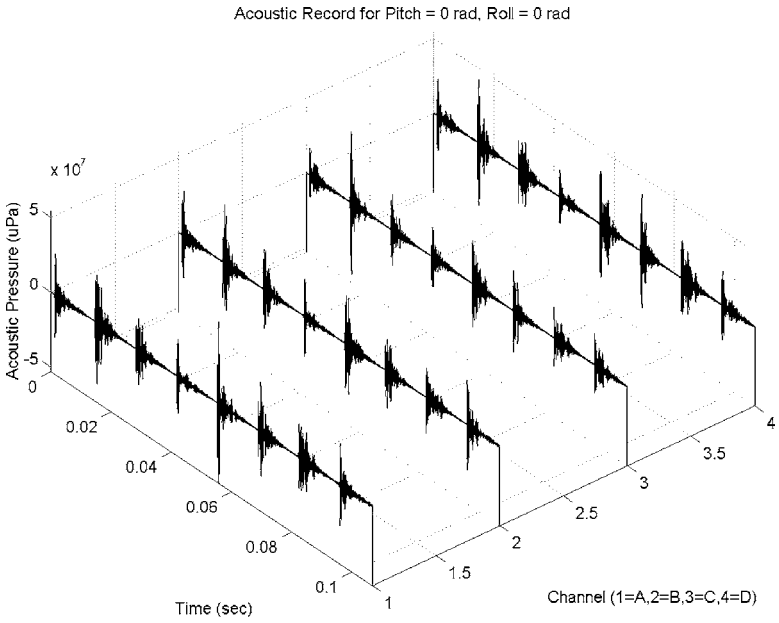


FIG. 14. Example of recorded signals.

$$x_0 = r_0 \cos(\phi_0)\cos(\psi_0),$$

$$y_0 = r_0 \sin(\phi_0)\cos(\psi_0),$$

$$z_0 = r_0 \sin(\psi_0). \tag{26}$$

Similarly, the Cartesian coordinates for the estimated source position given r_{est}, Φ_{est}, and Ψ_{est} are

$$x_{\text{est}} = r_{\text{est}} \cos(\phi_{\text{est}})\cos(\psi_{\text{est}}),$$

$$y_{\text{est}} = r_{\text{est}} \sin(\phi_{\text{est}})\cos(\psi_{\text{est}}),$$

$$z_{\text{est}} = r_{\text{est}} \sin(\psi_{\text{est}}). \tag{27}$$

Equation (28) shows the distance between the true and estimated source positions for the n_tth test and n_ath source location,

$$d_{\text{pos},n_a,n_t} = \sqrt{(x_{\text{est}} - x_0)^2 + (y_{\text{est}} - y_0)^2 + (z_{\text{est}} - z_0)^2}. \tag{28}$$

The position estimation error is the position error relative to the range,

$$E_{\text{pos},n_a,n_t} = 100\frac{\sqrt{(x_{\text{est}} - x_0)^2 + (y_{\text{est}} - y_0)^2 + (z_{\text{est}} - z_0)^2}}{r_0}. \tag{29}$$

E_{pos,n_a,n_f} represents the position estimation error for a single trial and for a single source location. The mean position error for a single source location over the entire set of N_t trials is given by

$$\bar{E}_{\text{pos},n_a} = \frac{1}{N_t}\sum_{n_t=1}^{N_t} E_{\text{pos},n_a,n_t}. \tag{30}$$

Finally, the position error averaged over every tested position ($n_a = 1, \ldots, N_a$) and trial is given by

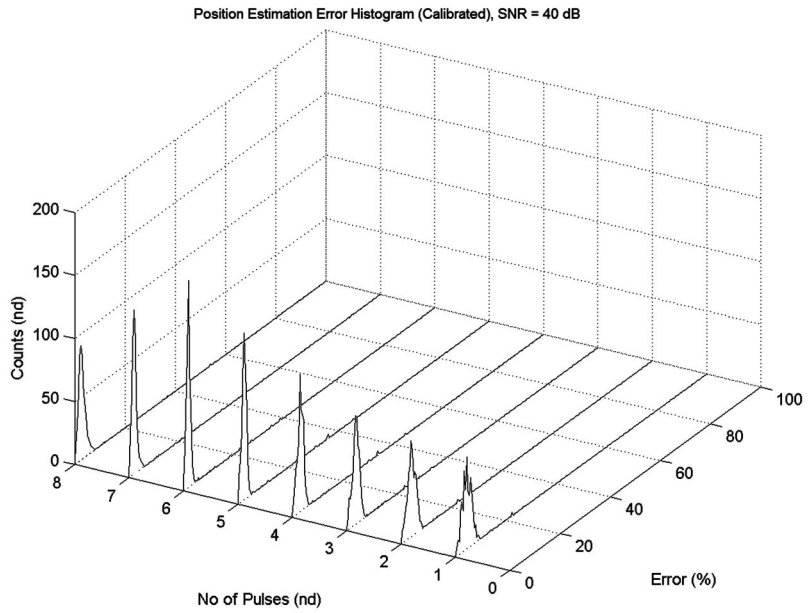


FIG. 15. Histogram of the position estimation error as a function of the number N_f of pulses processed, SNR =40 dB, using experimental data.

TABLE I. Simulation parameters.

c	1500 m/s
r_0	0.62 m
L_0 (n_b=1,2,3)	0.025 m
L_0 (n_b=4,5,6)	0.024 m
σ_L^2	$6.25e-6$ m^2
T	0.00062 s
W	1273 Hz
N_a	152 positions
N_t	50 trials per position

$$\bar{E}_{\text{pos}} = \frac{1}{N_a N_t} \sum_{\substack{n_a=1 \\ n_t=1}}^{N_a,N_t} E_{\text{pos},n_a,n_t}. \tag{31}$$

IV. EXPERIMENTAL PROCEDURE

A. Setup

A tank (Fig. 12) has been designed and built specifically for the calibration and testing of the USBL APS. The tank is 1.6 m by 1.6 m by 1.8 m tall. These dimensions allow for a reverberation-free time T_r of 0.62 ms if the USBL array is placed in the middle of the tank, and if the source is placed against the middle of a wall slightly below the USBL array. When the USBL array is vertical (at a pitch of 0 rad), the center-to-center distance from the source to the USBL array is 0.62 m. The time-frequency response of the tank, measured on hydrophone A of the array in vertical position, is shown in Fig. 13. The corresponding time signals recorded in this configuration are shown in Fig. 14.

The signals are acquired and processed using an embedded digital signal processing board, equipped with a TMS320C54-16 fixed point processor. The signal is transmitted using a second FAU-DPAM board. The control of the stepper motors and data collection is performed by a standard PC. The experimental process is fully automated. A TCM2-50 tilt sensor is also installed in a pressure housing just above the USBL array to provide redundancy in the array pitch estimation. The pressure vessel containing the sensor is oil-filled, and the articulated arms are made of plastic to minimize acoustic reflections. The hydrophones of the USBL array are cylindrical ceramics of 9.5 mm outer diameter, 9.7 mm height. The cylinder walls are 1.2 mm thick.

Change in the relative position of the source is achieved by changing the pitch and roll of the receiver array. The rotation of the array is performed by two stepper motors equipped with encoder feedback to control accurately the array orientation. Pitch and yaw are controlled with a resolution of 0.0061 and 0.0022 rad, respectively. Rotating the array rather than moving the source is advantageous as it simplifies the design of the calibration tank, but complicates the estimation of the source position, measured with respect to the geometrical center of the USBL array. The details of the operation are beyond the scope of this paper.[12]

B. Calibration and measurements

The calibration process is a necessary procedure before studying the fluctuations in positioning accuracy of the APS unit as a function of SNR and number of processed pulses. The relative position of the piezo-electric elements within the USBL array is not perfectly accurate. Calibrating an array consists in creating a look-up table containing correction factors to reduce the impact of the sensor location error.

The array calibration is performed in two steps. First, the Cartesian coordinates of the source with respect to the center of the array are estimated acoustically, for each array orientation and under excellent conditions (SNR of 40 dB). Each estimate is obtained using the maximum likelihood technique previously described, using N_f=8 pulses and N_t=5 trials. The final estimated location is obtained by averaging the coordinates found for each trial. For each array orientation, the Cartesian coordinates of the source are also measured using the stepper motor controllers and the tilt sensor. The difference between this measurement and the result obtained acoustically is recorded in a calibration look-up table. Once

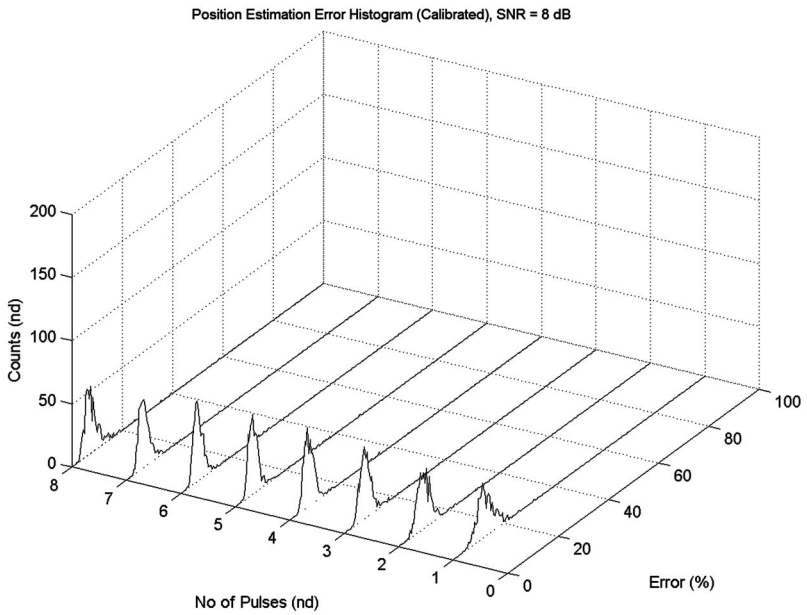


FIG. 16. Histogram of the position estimation error as a function of the number N_f of pulses processed, SNR =8 dB, using experimental data.

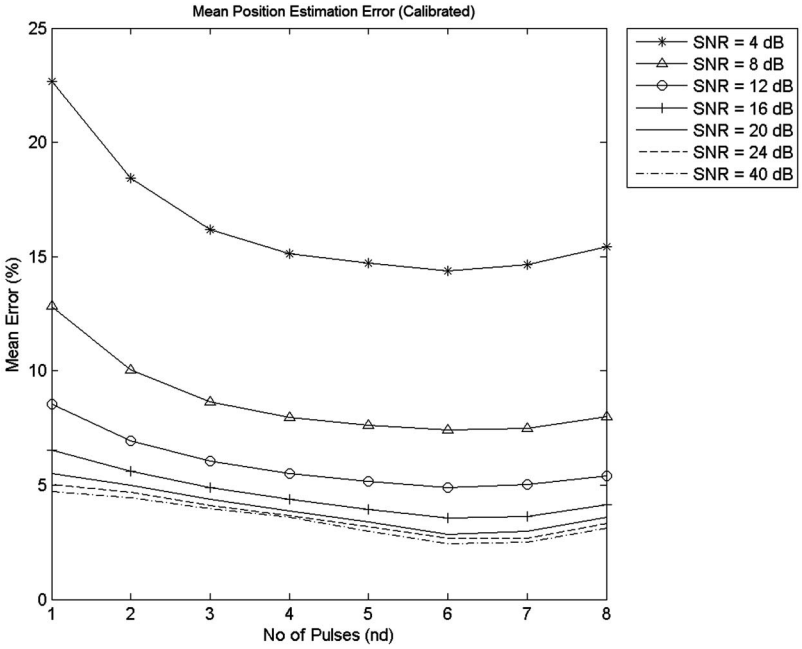


FIG. 17. Mean position estimation error as a function of the number N_f of pulses processed and SNR, using experimental data.

the array has been calibrated, measurements are performed as follows: (a) $N_t=5$ signal transmissions are recorded and processed for each orientation of the receiver array; (b) the pitch and roll as estimated by the stepper motors' controllers and by the tilt sensor are recorded; (c) a correction factor is applied to the maximum likelihood estimation using the calibration table; and (d) the position estimation error is determined from the measurements provided by the stepper motor controllers and tilt sensor.

The range of tested angles spans from 120° to 300° in azimuth and 7° to 45° in elevation. At first, the SNR is set to 40 dB by reducing the source level. Although the operation is automated, an entire set of measurements at a given SNR takes approximately 48 h. There is also a limitation to the minimum level of sound generated in the tank, and little guarantee that the measured noise is isotropic, as it has been previously assumed. Therefore, noise is artificially added to the signals measured on each channel to cover a wide range of SNR.

C. Simulation

An artificial set of data is generated to validate the experimental results presented in this document. A complete list of the simulation parameters is given in Table I. The simulation process uses the algorithm described in Fig. 3, with the exception that the bearing estimates are obtained directly from Eq. (16) using a Gaussian random noise generator. As in the experiment, the angles tested in the simulation vary from 120° to 300° in azimuth and 7° to 45° in elevation. The true azimuth and elevation are assumed to be perfectly known.

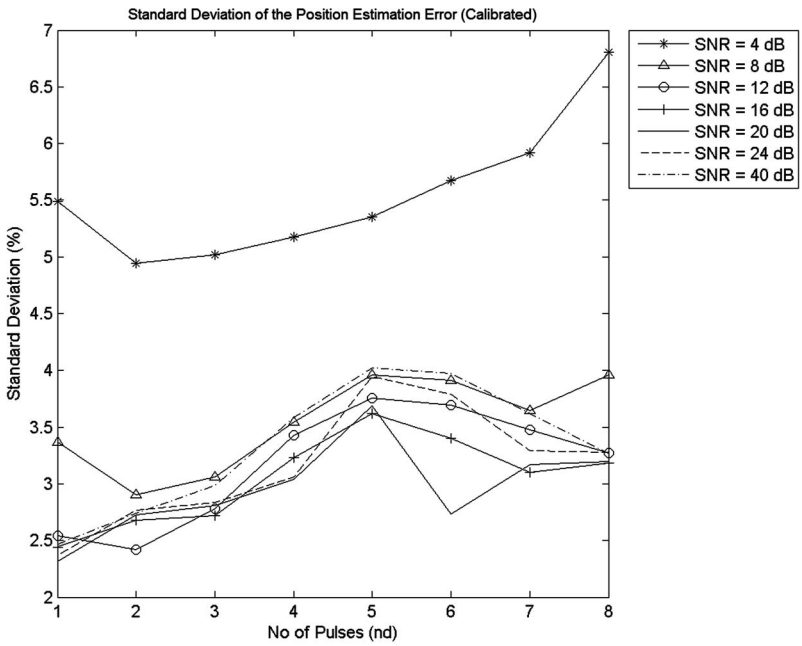


FIG. 18. Measured standard deviation of the position estimation error as a function of the number N_f of pulses processed and SNR.

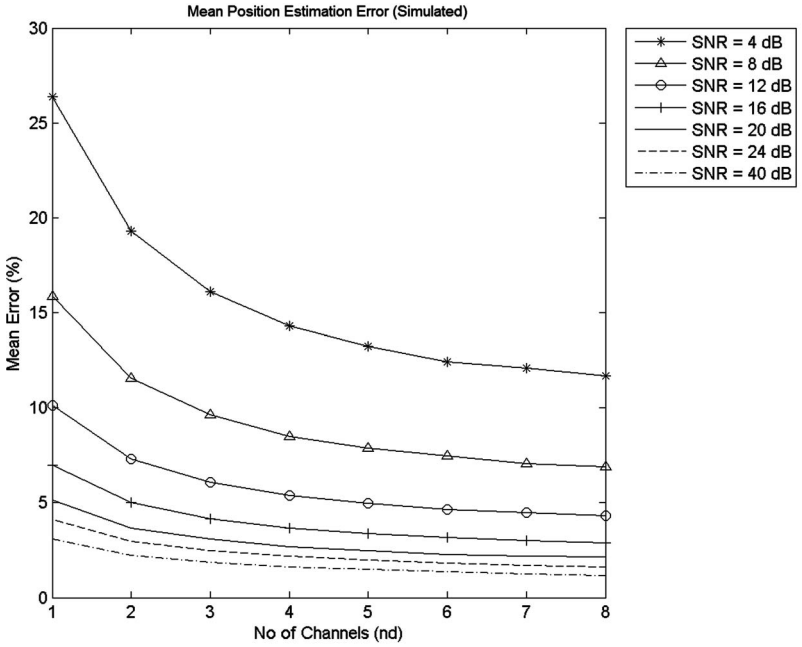


FIG. 19. Mean position estimation error as a function of the number N_f of pulses processed and SNR, using simulated data.

V. RESULTS

The main objective of this section is to show that, at a fixed signal-to-noise ratio, the positioning accuracy of the APS unit improves significantly as the number of processed pulses increases. This essential result indicates that an ACOMS source transmitting signals of limited output power can be located accurately with a small USBL array if a portion of the message is known. The positioning accuracy of the APS has been tested at SNR of 4, 8, 12, 16, 20, 24 and 40 dB, respectively. The simulation parameters given in Table I also apply to the experiment, with the exception of N_t. In the experiment, the number of trials per position is limited to $N_t=5$.

Figures 15 and 16 show the histogram of the position estimation error measured at $N_a=152$ locations, measured with a SNR of 40 and 8 dB, respectively. The total number of measurements is $N_tN_a=760$ per SNR. The total number of pulses used to estimate the source location varies from $N_f=1$ to $N_f=8$. Figures 4 and 5 already showed that the bearing estimate is more accurate as the central frequency of the pulse increases, due to the element spacing given in Eq. (2). As a result, the source position is computed as a function of an increasing number N_f of pulses, sorted by decreasing central frequency: 27, 25.8, 24.6, 23.4, 21, 19.8, 17.4, and 16.2 kHz. If $N_f=1$, the pulse received at 27 kHz is used. If $N_f=3$, the pulses received at 27, 25.8, and 24.6 kHz are used. If $N_f=8$, all the received pulses are used. The purpose of Figs. 15 and 16 is to show that the histograms are compact, the error remaining below 25% for a SNR of 8 dB, and 10% for a SNR of a 40 dB.

Figures 17 and 18 show the mean and standard deviation of the position estimation error measured across all $N_tN_a=760$ points per SNR, as a function of the number of pulses N_f used to estimate the source location. Figure 19 shows the mean position estimation error obtained from the simulation under identical conditions, with the exception of $N_t=50$. Given the standard deviation shown in Fig. 18, measured and simulated data are a close match, which validates the proposed method and the quality of the measurements. The most noticeable discrepancy between simulated and measured data occurs when $N_f>6$. This is easily explained by the fact that the half-wavelength of the pulses received at 17.4 and 16.2 kHz is significantly larger than the length of each baseline, and lead to more ambiguous measurements.

Focusing on $N_f\leqslant 6$, Figs. 17 and 19 clearly indicate that combining multiple estimates from the MFSK multitone sequence is highly beneficial: for a SNR of 8 dB and $N_f=1$, the mean position estimation error is 13%, but drops to 8% when $N_f=3$. The same error of 8% is observed for a SNR of 12 dB and $N_f=1$. Therefore, combining the three pulses of the multitone is equivalent to a gain in SNR of 4 dB. At a moderate SNR of 4 dB, the position estimation error drops by 65% when $N_f=6$ pulses are used instead of 1, which is almost equivalent to a gain in SNR of 4 dB. At higher SNR, the drop in position estimation error as N_f increases is not as pronounced, due to the error in sensor location which becomes a dominant factor. However, the equivalent gain in SNR is even more significant. Figures 17 and 19 show that the APS will perform with very similar accuracy when the SNR exceeds 24 dB. Overall, the APS can achieve a position estimation estimation error of 3% using six pulses given a SNR of 24 dB. Obviously, other environmental factors can severely impact the performance of the system and in particular sound speed variations. This specific issue is considered as future work.

VI. CONCLUSION

Experimental and simulated results indicate there is a significant advantage in using a known portion of a MFSK multitone sequence to estimate the location of a modem source, especially at moderate SNR. Therefore, a USBL APS can be used to accurately locate acoustic transmissions of a MFSK modem during the transmission of a message without impacting the modem operation. The accuracy improves with

the number of pulses and the SNR, and a positioning accuracy of 3% can be achieved for a SNR of 24 dB. Sensor positioning and proper array calibration are essential factors in the performance of the APS, especially at high SNR.

ACKNOWLEDGMENT

This work is sponsored in part by the Office of Naval Research, Code No. ONR 321OM, Dr. Tom Swean.

[1]P. H. Milne, *Underwater Acoustic Positioning Systems* (Spon, New York, 1983).
[2]R. J. Urick, *Principles of Underwater Sound*, 3rd ed. (McGraw-Hill, New York, 1983).
[3]K. Vickery, "Acoustic positioning systems, a practical overview of current systems," *Proceedings of the IEEE 1998 Workshop on Autonomous Underwater Vehicle*, August 1998, pp. 5–7, 20, 21.
[4]J. Opderbecke, "At-sea calibration of a USBL underwater vehicle positioning system," *Proceedings of the Oceans' 97 MTS/IEEE Conference*, October 1997, Vol. **1**, pp. 721–726.
[5]D. B. Kilfoyle and A. B. Baggeroer, "The state of the art in underwater acoustic telemetry," IEEE J. Ocean. Eng. **25**, 4–27 (2000).
[6]P. P. J. Beaujean and L. R. LeBlanc, "Adaptive array processing for high-speed communication in shallow water," IEEE J. Ocean. Eng. **29**, 807–823 (2004).
[7]L. Freitag, M. Johnson, M. Grund, S. Singh, and J. Preisig, "Integrated acoustic communication and navigation for multiple UUVs," *Proceedings of the MTS/IEEE Oceans' 2001*, Vol. **4**, pp. 2065–2070.
[8]P. P. J. Beaujean and E. P. Bernault, "A new multi-channel spatial diversity technique for long range acoustic communications in shallow water," *Proceedings of the MTS/IEEE Oceans'03*, San Diego, CA, 2003.
[9]E. A. Carlson, P.-P. Beaujean, and E. An, "An ad hoc wirless acoustic network simulator applied to multiple underwater vehicle operations in shallow waters using high-frequency acoustic modems," JUA(USN) **56**, in press (2006).
[10]L. Freitag, M. Stojanovic, M. Johnson, and S. Singh, "Analysis of channel effects on direct-sequence and frequency-hopped spread-spectrum acoustic communication," IEEE J. Ocean. Eng. **26**, 586–593 (2001).
[11]R. L. Peterson, R. E. Ziemer, and D. E. Borth, *Introduction to Spread Spectrum Communications* (Prentice-Hall, Englewood Cliffs, NJ, 1995).
[12]R. Warin, "Maximum likelihood estimates of azimuth and elevation for a frequency-hopped active source using a tetrahedral ultra short baseline," M.S. thesis publication, Department of Ocean Engineering, Florida Atlantic University, Boca Raton, FL, 2004.
[13]P.-P. J. Beaujean, A. I. Mohamed, and R. Warin, "Maximum likelihood estimates of a spread-spectrum source position using a tetrahedral ultra-short baseline array," *Proceedings of the IEEE Oceans' 2005 Europe*, June 2005, Vol. **1**, pp. 142–146.
[14]A. H. Quazi and D. T. Lerro, "Passive localization using time delay estimates with sensor positional errors," J. Acoust. Soc. Am. **78**, 1664–1670 (1985).
[15]P. Stoica, M. Viberg, K. M. Wong, and Q. Yu, "Maximum likelihood bearing estimation with partly calibrated arrays in spatially correlated noise fields," IEEE Trans. Signal Process. **44**, 888–899 (1996).
[16]R. O. Nielsen, "Azimuth and elevation angle estimation with a three-dimensional array," IEEE J. Ocean. Eng. **19**, 84–86 (1994).
[17]D. D. Lee and R. L. Kashyap, "Robust maximum likelihood bearing estimation in contaminated Gaussian noise," IEEE Trans. Signal Process. **40**, 1983–1986 (1992).
[18]P. K. Ray and A. Mahajan, "Optimal configuration of receivers in an ultrasonic 3D position estimation system by using genetic algorithms," *IEEE, Proceedings of the American Control Conference*, June 2000, Vol. **4**, pp. 2902–2906.
[19]J. S. Bendat and A. G. Piersol, *Random Data, Analysis and Measurement Procedures*, 2nd ed. (Wiley, New York, 1986).
[20]W. S. Burdic, *Underwater Acoustic System Analysis* (Peninsula, Los Altos, CA, 2002).
[21]A. Quazi, "An overview on the time delay estimate in active and passive systems for target localization," IEEE Trans. Acoust., Speech, Signal Process. **29**, 527–533 (1981).
[22]H. Cramer, *Mathematical Method of Statistics* (Princeton University Press, Princeton, NJ, 1968).
[23]W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes in C* (Cambridge University Press, Cambridge, 1988), pp. 290–328.

Measurement of the acoustic reflectivity of sirenia (Florida manatees) at 171 kHz

Jules S. Jaffe and Fernando Simonet
Marine Physical Lab, Scripps Institution of Oceanography, La Jolla, California 92093-0238

Paul L. D. Roberts
Department of Electrical and Computer Engineering, University of California, San Diego, California 92093

Ann E. Bowles
Hubbs-Sea World Research Institute, 2595 Ingraham Street, San Diego, California 92109

(Received 7 June 2006; revised 27 September 2006; accepted 6 October 2006)

The Florida manatee (*Trichechus manatus latirostris*) is an endangered sirenian. At present, its adult population (~2200) seems stable, but tenuous. Manatee-boat collisions are a significant proportion (~25%) of mortalities. Here, the potential use of active sonar for detecting manatees by quantifying sonic reflectivity is explored. In order to estimate reflectivity two methods were used. One method measured live reflections from captive animals using a carefully calibrated acoustic and co-registered optical system. The other method consisted of the analysis of animal tissue in order to obtain estimates of the sound speed and density and to predict reflectivity. The impedance measurement predicts that for a lateral view, the tissue reflectivity is close to 0.13, with a critical grazing angle of 28°. Data measured from live animals indicate that substantial reflections can be recorded, however in many instances observed "empirical target strengths" were less than an experimentally dependent −48-dB threshold. Conclusions favor the hypothesis that the animals reflect substantial amounts of sound; however, the reflections can often be specular, and therefore impractical for observation by a manatee detection sonar operating at 171 kHz. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2384845]

PACS number(s): 43.30.Vh, 43.58.Bh, 43.80.Ev [WWA] Pages: 158–165

I. INTRODUCTION

The Florida manatee (*Trichechus manatus latirostris*) is an endangered sirenian inhabiting the shallow coastal waters of Florida (USA). At present, its adult population (~2200) appears to be stable or increasing (Craig and Reynolds, 2004; Langtimm *et al.*, 2004; Florida Manatee Biological Review Panel, 2005), but natality and mortality have been nearly equal in recent years, suggesting that the population has limited potential for growth. Human-caused mortalities account for a significant proportion of the mortalities, approximately 25%. Collisions with watercraft are the major cause [on average 24% of losses over the last five years; in 2005 80 manatees died in collisions (Florida Fish and Wildlife Conservation Commission, unpublished; Marine Mammal Pathobiology Lab, 2005)]. Furthermore, this is the only cause of mortalities that can be managed directly. Therefore, Florida management agencies are actively supporting research to find ways of reducing collisions, particularly if they can be used to reduce mortalities while maintaining recreational opportunities for millions of registered boaters. Sonar techniques certainly have this potential.

One strategy for reducing manatee-boat collisions consists of alerting the boater to the presence of the animals. Given suitable time to respond, the boater can take action to minimize the chance of collision. Active sonar systems have the potential to fulfill the role of detecting animals at suitable ranges provided they can function in the manatee habitat and the animals are sufficiently reflective.

Although the intent of the article is not to summarize in detail the various strategies that might be used to detect the presence of manatees, we do remark that various researchers have been concerned with this problem over the years. Work (Gerstein, 2002) suggested that the animals cannot hear an oncoming craft because of the Lloyd's mirror effect. Certainly this makes the animal's job of evading boats more difficult. In order to ameliorate the situation, it has been suggested that motor craft in areas frequented by manatees carry an acoustic transmitter so that the animals will be alerted to the presence of boats. To our knowledge, the success of this strategy has not been tested under field conditions. In the laboratory, manatees respond to audible tone pips by approaching and manipulating sound sources (Bowles *et al.*, 2001; Bowles, 2002). They have also been shown to respond in unpredictable and potentially inappropriate ways to boat noise (Nowacek *et al.*, 2004). An alternative that has been tested is passive detection using vocalizations. Manatees produce short chirps in the range from 500 Hz to 6 kHz, with most energy at 2–3 kHz (Schevill and Watkins, 1965; Gerstein *et al.*, 1999; Niezrecki *et al.*, 2003). However, estimated source level for the vocalizations is very low (maximum 110 dB *re*:1 μPa @ 1 m) and the manatees only vocalize intermittently. In addition, the sounds are sometimes difficult

to distinguish from other short transients in the manatee habitat, especially at low signal-to-noise ratios. Therefore, the method is likely to be most effective when used concurrently with other techniques.

Although the active acoustic detection of manatees theoretically is an option, there has been little work in estimating sonar reflectivity of large marine animals. Probably the closest underwater acoustic models of the manatee that have been subjects of target strength measurements are cetaceans. Au (1996) measured target strengths of bottlenose dolphins of −11 to −23 dB as a function of frequency between 23 and 79 kHz. Measurements showed a strong dependence on dolphin orientation. However, the available evidence indicates that dolphin blubber and skin differ from those of the manatee (Kipps *et al.*, 2002), suggesting that the sonic properties of manatees are quite different as well. Such differences may explain why previous attempts to detect manatees using sonar have not been particularly successful (e.g., Dickerson *et al.*, 1996).

With respect to larger animals, Miller and Potter (2001) formulated a three-layer model in conjunction with density and sound speed estimates of right whale skin and blubber in order to predict target strengths. These were then compared with field data and good agreement was found. Based on skin measurements (density, sound speed) of 1700 m/s and 1200 kg/m^3 and blubber (density, sound speed) estimates of 1600 m/s and 900 kg/m^3, they predicted target strengths at 86 kHz to be between 0 and −5 dB depending on blubber thickness (0–30 cm). Measured target strengths of two adult right whales were consistent with these estimates.

Based on lack of data on manatee reflectivity and the increasing interest in estimating the possible role that active sonar might play in reducing animal-boat collisions, a program to measure animal sonic reflectivity was undertaken. In order to measure the sonic reflectivity of the Florida manatee, two measurement techniques were employed. The first consisted of measuring reflections from captive animals in an enclosed pool using a calibrated sonar system. Another method used frozen and subsequently thawed animal tissues to measure sound speed and density, from which acoustic impedance could be calculated. These measurements resulted in an estimate of reflectivity as a function of angle, including the critical angle at which all sound was reflected.

Since the thick manatee "skin" is mostly composed of densely woven collagen (Kipps *et al.*, 2002; Sokolov, 1982), the acoustic properties of this material are relevant. Collagen is the most abundant protein in mammals (White *et al.*, 1968), contributing 17% of the total body mass and 70% of the negative buoyant force (Sokolov, 1982). Density in its native state can vary from 1160 to 1330 kg/m^3 (Hulmes *et al.*, 1977; Dweltz, 1962). Such dense protein exhibits remarkably high ultrasonic attenuation, absorption, and velocity in tissues. Goss and Dunn (1980) studied the ultrasonic propagation properties of collagen and estimated a sound speed of 2094 m/s in a sample with a collagen concentration of 80%, at 20 °C at 8.97 MHz. More directly applicable to this study, Kipps *et al.* (2002) measured the skin density from 27 male and female manatee carcasses. The average observed value was 1121 kg/m^3.

Goss and Dunn (1980) established an acoustic absorption coefficient per unit concentration in collagen suspension at 20 °C as a function of frequency as $\alpha_C = 114 f^{-0.53}$, where f is the frequency in MHz and α_C is the absorption coefficient per unit concentration. Attenuation can then be calculated via the expression $\Delta\alpha = \alpha_C f^2 C_S$, where C_S is concentration, in dB as $\Delta\alpha_{dB} = \log_{10}(e)\Delta\alpha$. Agemura *et al.* (1990) obtained attenuation results of 80 dB/mm at 100 MHz from articular cartilage (composed of mostly collagen) of young steers. This result is slightly lower than that measured by Goss at 86 dB/mm. Extrapolating these results using Goss' formula yields an expected value of 2.3 dB/mm at 10 MHz and 0.6 dB/mm at our working frequency of 171 kHz.

II. METHODS

A. Sound speed and density measurements

Owing to the difficulty of working with live animals, both the density and sound speed measurements were obtained from a post-mortem subject. Samples were collected from an adult male that collided with a boat on 22 April 2004. Shortly after the collision, the manatee was euthanized, and samples were collected within 1 h of death and then frozen. Three 8×8 cm^2 samples were extracted from the dorsal, lateral, and ventral aspects of the manatee at the umbilicus that included skin and the full depth of blubber. Before the tests were performed, the samples were thawed, divided into smaller pieces, and then refrozen for storage.

1. Dissection

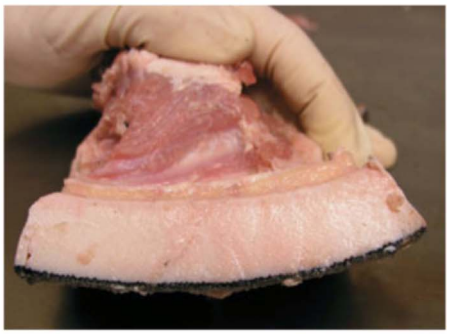
Figure 1 shows the dorsal, lateral, and ventral samples. All three skin samples were composed of a thin epidermis (2 mm) over a thick dermis, which was attached to blubber and muscle. The three samples had different skin thicknesses, 1 cm on the ventral, 1.7 cm on the dorsal, and 2.4 cm on the lateral aspect. The dorsal sample had ~1 cm of blubber followed by muscle, the lateral sample had a thicker layer of blubber (~2 cm) and then muscle. The ventral sample had skin attached to a ~1-cm layer of blubber followed by muscle (~2 cm), then more blubber (~1 cm) and another muscle layer.

2. Sound speed and attenuation

Sound speed was measured using a Krautkramer Branson USD10 Ultrasonic Digital Flaw Detector with two Krautkramer Branson transducers (Alpha series, 10 MHz, 0.25 mm) mounted onto a Mitutoyo digital caliper (Model CD-8"CS) (Fig. 2). The Krautkramer equipment was used to measure the delay time of the transmitted pulse from the top to bottom transducer. Prior to use, the instrument was calibrated by measuring the travel time for a 10-MHz pulse to propagate through a manufacturer-supplied copper strip (with known sound speed) and also distilled water at 19.5 °C whose sound speed is known to be 1490 m/s. Calipers were used to measure the sample thickness before the sound speed measurement. The sound speed was then computed as the ratio of the thickness to the time interval between transmit and receive pulse. Attenuation was computed as

(a)

(b)



(c)

FIG. 1. (Color online) (a) Ventral sample. (b) Lateral sample. (c) Dorsal sample (figure online).

$10 \log_{10}(I_{rec}/I_{inc})/\Delta x$, where I_{rec} is the intensity of the received waveform, I_{inc} is the intensity of the transmitted waveform, and Δx is the sample thickness. Results (in dB/mm) were observed at 10 MHz.

Since the sound speed and attenuation measurements were performed at 10 MHz (high frequencies are needed to keep samples down to a workable size), some extrapolation to the frequency employed at 171 kHz was necessary. Assuming that the medium is nondispersive implies that the sound speed will be the same at 171 kHz as it is at 10 MHz. Unfortunately, the attenuation results cannot be treated in a similar manner because absorption and scattering change significantly as a function of frequency.

FIG. 2. (Color online) The experimental configuration used for the sound speed measurements (figure online).

3. Density

In order to compute sample density, samples were submerged in a graduated cylinder in distilled water and the resultant increase in volume displacement and sample weight was measured using a Sartorius Handy H51 scale (0.0001 g accuracy). Density was calculated as mass/volume.

4. Acoustic impedance

Acoustic impedance determines how much sound is transmitted and reflected at any given boundary. Using the familiar Fresnel law of reflection, the reflection coefficient at the manatee-water interface can be calculated as

$$R_{W-M} = \frac{\rho_M c_M - \rho_W c_W}{\rho_M c_M + \rho_W c_W}, \tag{1}$$

where ρ_M and c_M are the density and the sound speed of the manatee skin and ρ_W and c_W are the density and the sound velocity in the fresh water pool. Reflectance (R) as a function of incidence angle α can then be estimated as

$$R = \frac{\cos\alpha - b\sqrt{\sin^2\alpha - \sin^2\alpha_C}}{\cos\alpha + b\sqrt{\sin^2\alpha - \sin^2\alpha_C}}, \tag{2}$$

where $b=\rho_1/\rho_2$ and the critical grazing angle, where grazing angle is defined as the angle between a plane parallel to the surface and an incident ray $\alpha_C=\cos^{-1}(c_1/c_2)$.

B. Measurement of the acoustic reflectivity of live manatees

An estimate of live manatee acoustic reflectivity will be described in this section. Working at the Manatee Rescue facility at SeaWorld San Diego, a series of experiments was performed in order to estimate animal target strength. The measurements were complicated by the fact that they were necessarily performed in the near field of the acoustic system. In addition, because of the large size of the animal, the sonar system was in the near field of the energy reflected from the manatee. As a candidate choice of frequency, based mainly upon the availability of hardware, a frequency of 171 kHz was used. This frequency was well above the upper limit of hearing of the manatee (Gerstein *et al.*, 1999) and bottlenose dolphin (Johnson, 1967), the two marine mammal species that could be encountered in areas where an avoidance sonar might be deployed. Additional complications that arose in the course of performing the experiments were due to the high level of ambient electrical noise. In addition, significant reverberation due to the geometry of the pool made it difficult to interpret data from ranges greater than

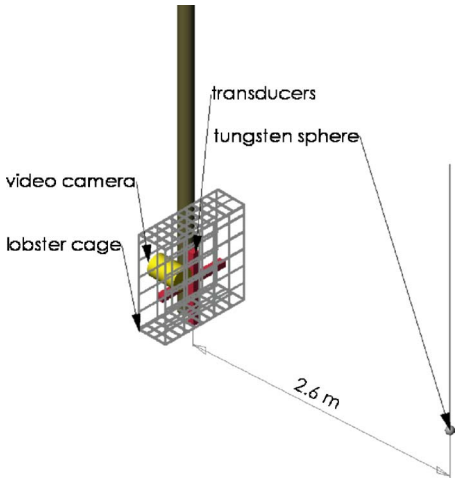


FIG. 3. (Color online) A diagram of the calibration setup consisting of (two) side scan transducers, a video camera, and a tungsten carbide sphere mounted inside of a cage.

4–5 m. Coupled with the seemingly low reflectivity of the manatees, the use of simple techniques for measuring target strengths was precluded.

1. Calibration of the sonar equipment

The permit for the experiments allowed a maximum source level of 180 dB *re*: 1 μPa @ 1 m. To calibrate source level, an E-8 Navy calibrated transducer was used as a hydrophone. The hydrophone was positioned at a known distance from the source and the output of the transducer was determined for a given input voltage amplitude using the sonar equation (Urick, 1967). Input voltages were adjusted so that the source level was always below 180 dB.

In order to measure animal reflectivity, two transducers of size 15 mm by 498 mm were employed. The approximate beam pattern in the far field of the transducers at 171 kHz (center frequency) was 20° by 1° as measured in the Transducer Evaluation Center (TRANSDEC), Naval Ocean Systems Center, San Diego. The sonars were mounted at right angles to each other as shown in Fig. 3, with one used in transmit mode while the other was used to receive. The resultant composite beam did not spread very much as the far field pattern is approximately $1° \times 1°$. This helped to reject the reverberation from the pools walls and air-water interface, providing a practical solution to the reverberation problem at modest ranges (1–4 m).

In fisheries acoustics, a standard technique for calibrating a sonar system involves the use of calibrated spheres (Vagle *et al.*, 1996). In a typical application, a sphere of known reflectivity and size and therefore measurable target strength is positioned at various three-dimensional locations in the field of view of the sonar transducer(s). The system response (output voltage of the system from the echo relative to input voltage) is then measured for each location. These values are then used to calibrate the system by noting the relationship between output levels as a function of target position. Given a measurement of target position, input, and output level, the target strength from a target at a measured position can then be estimated via extrapolation of these results when combined with an understanding of the three-dimensional energy density incident on targets with spreading loss via use of the sonar equation (Urick, 1967).

The procedure for the manatee experiments was based on this protocol. However, the exact reflectivity of a manatee in the sonar's field of view was more difficult to judge because, even at the furthest range of 5 m, the animal was in the near field of the sonar. So, for example, given a transducer of dimensions 15 mm by 498 mm, and using the formula for the start of the far field as D^2/λ (Kino, 1987), the start of the far field for the larger dimension was estimated to be 35 m. Although smaller transducers could be used and, hence, decrease the distance to the far field, a larger incident sound field pattern would result, perhaps complicating the measurements even more. In consideration of this, it was decided to use our existing transducers.

Since the goal of the experiment was to explore the potential use of an active sonar system to detect acoustic reflections from manatees, it was decided that some quantitatively determined proxy of animal reflectivity would suffice. Although the results are not entirely general in that they cannot be expressed as being the true target strength of the animal, a quantity that is independent of the exact geometry of the measuring apparatus, our system was calibrated and the results can be communicated to other researchers using a protocol that can be readily duplicated.

The experiments proceeded by characterizing the reflectivity of the manatee at a known set of three-dimensional locations relative to a small, exactly calibrated, and precisely positioned target. Given a voltage from the test target, V_{target} (−39 dB), and the voltage output from the manatee, V_{manatee} (at the same position), the estimated, empirically observed, "relative target strength" was defined as

$$\text{TS}_{\text{rel}} = 20 \log_{10}\left(\frac{V_{\text{manatee}}}{V_{\text{target}}}\right) - 39. \tag{3}$$

In order to judge both animal and test target position in three dimensions during the experiments a video camera was placed in a waterproof housing and mounted so that its field of view was oriented almost exactly along the axis of the combined transducer transmitting and receiving beam. The video allowed judgment of the bearing angle, while the reflected acoustic pulse permitted an estimate of range. Hence, the three-dimensional position of the animal could be measured. Past experiences with combining optics and acoustics resulted in enhanced characterization of targets (Jaffe *et al.*, 1998) and their precise three-dimensional locations (Schell and Jaffe, 2004).

The video camera and sonar transducers were calibrated together by translating a 38-mm tungsten carbide sphere (Foote, 1990), of target strength [TS]=−39 dB @ 171 kHz, in the test tank. The sphere was translated in two dimensions at several ranges while both video images and sonar reflections were recorded. Thus, effective beam patterns of the sonar were recorded for this omni-directional reflector, as measured at approximately 100 points in the field of view of the system over several ranges (1, 2, and 2.6 m).

Results indicated that at the very shortest ranges, 1 and 2 m, the near-field sound pattern was quite complicated. However, at 2.6 m, the field became a more interpretable sound pattern. At this and greater ranges, a "system response" could be used to estimate animal reflectivity from the calibration measurements.

At the completion of the calibration, the voltage obtained via reflection from the sphere could be accurately predicted as a function of absolute three-dimensional position. In addition, the calibration curves from data collected with the sphere at 2.6 m were extrapolated to a slightly greater range (4 m) for use. Since a range of 4 m was still in the near field of the system, spherical spreading could not be used to estimate reflectivity. A computer simulation was written to model the two-way spreading that the sound underwent for an object at 4-m range relative to the farthest calibrated sphere measurement at 2.6 m. At this range, simple 1/range $(1/R)$ spreading appeared to model the loss best. A $1/R$ spreading loss was therefore applied to estimate the system's output voltage from a target positioned at 4 m, extrapolated from the 2.6 m data.

2. Measurement of "relative" target strengths

The sonar setup was mounted in a protected cage to ensure that the manatees could not harm themselves or the equipment. The signals were 1-ms frequency swept "chirps" from 150 to 190 kHz. Output levels were kept within a safe range for manatee hearing in compliance with a permit that allowed performance of the experiments. Pressure levels 1 m from the sonar transducers were smaller than 180 dB *re*:1 μPa. Data from approximately 120 reflectivity measurements were used in the following analysis.

The processing method consisted of applying a matched filter detector to the received signals that was exactly the same as the transmitted wave. The resultant signal was displayed in "real time" so that the observer could monitor the results of the experiments. As an added calibration check, the spherical test target was also deployed in the field of view of the coincident optical-acoustic system during the experiments. This added measure insured that the reported values were truly related to the calibrated target.

The analysis took place in several stages. First, the capability of the system to detect targets in a reliable manner was tested by establishing a threshold for the sonar reflections. Based on analysis of the background noise level in the pool a value of −48 dB *re*:1 μPa was established. All reflections at least 10 dB over this threshold were considered to be potential returns from manatees. Upon detection of a sonar reflection that exceeded this value, the concurrent video data were inspected to see if manatees were present. Results indicated that a manatee was present in the field of view of the sonar system at the appropriate range in every case. Therefore, although the noise level in the pool was high, reflectivity from the manatees at ranges to 4 m were adequate to obtain unambiguous identification.

After acoustic detection, the video images were inspected to determine the section of the field of view of the sonar that the manatee was subtending. The chosen dataset included returns ranging from partial to total coverage of the field of view. Calibration data were then used to identify the most sonically reflective region within the field. The output voltage of the system was then used to compute the relative target strength via Eq. (3). This value constituted an estimate of target strength in dB. Taking the point inside the field of view with the largest V_{target} resulted in a conservative estimate of the animal's target strength. Other, less reflective areas would have yielded a higher ratio and therefore a larger inferred target strength for the manatee. However, the latter procedure was found to be numerically unstable in that the smaller values had more variability and were noisier. Moreover, the approach taken yielded an estimate of what might be expected from such a system in a practical sense.

III. RESULTS

A. Tissue measurements

In order to obtain an estimate of inherent variability in the tissue estimates of acoustic impedance the measurements were performed on many subsamples (connective tissue, blubber and muscle) that were extracted from ventral, dorsal, and lateral samples from approximately the midpoint of the body (at the umbilicus). Measurements were done at 21 °C (see Table I for results).

B. Acoustic reflectivity measurements

Figure 4 is a histogram of the observed relative target strengths for the more than 120 cases where the reflected signal was adequately above the noise level in the pool to be clearly detected. We note that in almost all cases, the reflected energy was substantially greater than that of the calibrated sphere. Although we know of no comparable values for other marine mammals at this frequency, the values seem reasonable especially when compared to the work of Au (1996) on bottlenose dolphins.

IV. DISCUSSION AND CONCLUSIONS

A. Tissue samples

Results for the connective tissue samples indicated an average sound velocity, density, and attenuation of 1708 m/s, 1149 kg/m^3, and 3.4 dB/mm (at 10 MHz), respectively. Table II contains a summary of the estimated reflection coefficients for the different regions. Reflection coefficient calculation under the Connective Tissue (CT) column is the reflection from the water-CT interface, under the Outer Blubber (OB) column is from the CT-OB interface, under the Outer Muscle (OM) column is from the OB-OM interface, under the IB column is from the OM-IB interface and, lastly, under the IM column is from the IB-IM interface.

To show the relationship of reflectivity to the angle between the incident sound wave and a plane parallel to the surface, a graph of reflectivity versus angle was then plotted using the data from the lateral umbilicus (1708 m/s and 1149 kg/m^3, from Table I) and the sound speed and density measurements in the manatee pool (1510 m/s and 1000 kg/m^3, respectively). Figure 5 shows the graph.

TABLE I. Sound velocity, attenuation, and density measurements of manatee tissue samples. CT=connective tissue, OB=outer blubber, OM=outer muscle, IB=inner blubber, IM=inner muscle, and n=number of samples.

Lateral	CT (n=7)	OB (n=7)	OM	IB	IM
Sound velocity (m/s)					
Mean	1707.82	1529.30	n/a	n/a	n/a
Standard deviation	5.75	8.34	n/a	n/a	n/a
10 MHz attenuation (db/mm)					
Mean	3.43	5.57	n/a	n/a	n/a
Standard deviation	0.20	1.66	n/a	n/a	n/a
Density (kg/m^3)					
Mean	1148.80	965.12	n/a	n/a	n/a
Standard deviation	70.53	63.78	n/a	n/a	n/a
Dorsal	(n=9)		(n=2)		
Sound velocity (m/s)					
Mean	1704.20	n/a	1601.77	n/a	n/a
Standard deviation	6.23	n/a	19.47	n/a	n/a
10 MHz attenuation (db/mm)					
Mean	4.18	n/a	3.39	n/a	n/a
Standard deviation	0.43	n/a	1.29	n/a	n/a
Density (kg/m^3)					
Mean	1113.60	n/a	1018.61	n/a	n/a
Standard deviation	40.58	n/a	2.94	n/a	n/a
Ventral	(n=6)	(n=6)	(n=6)	(n=6)	(n=6)
Sound velocity (m/s)					
Mean	1679.81	1525.42	1621.10	1519.01	1626.87
Standard deviation	13.16	17.58	10.05	21.91	10.65
10 MHz attenuation (db/mm)					
Mean	5.64	5.02	3.71	7.80	4.87
Standard deviation	0.88	0.64	0.81	1.48	1.06
Density (kg/m^3)					
Mean	1029.92	1064.42	1067.85	1023.21	1072.60
Standard deviation	63.78	106.66	42.64	108.53	82.01

B. Acoustic reflectivity measurements

As one aspect of this work, it is of interest to reconcile the "relative target strength" measurements with the reflectivity estimates in order to predict observed target strengths from the reflectivity data. To this end, an approximate target strength can be estimated as follows: Assuming the area subtended by the sonar at the target due to the composite beam pattern of $1^\circ \times 1^\circ$ is equal to A_{inc}, the power integrated over this area can be computed as $I_{\text{inc}}A_{\text{inc}}$, where I_{inc} is the intensity of the incident field. Next, the reflected power can be computed via multiplication by the reflectivity of the target ρ to yield $\rho I_{\text{inc}}A_{\text{inc}}$. The acoustic intensity measured at a 1-m range in the backscatter direction can them be computed by dividing this quantity by the area subtended by the backscattered hemisphere: 2π. Taking the logarithm of the ratio of this quantity to the incident acoustic intensity yields an estimate of target strength:

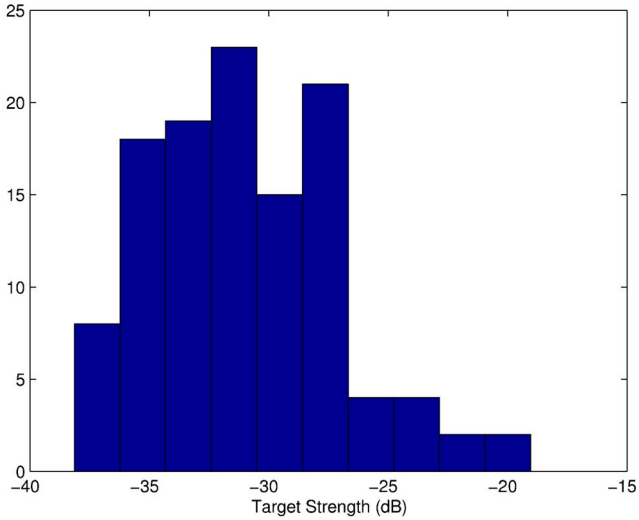


FIG. 4. (Color online) A histogram of "relative target strengths" for the approximate 120 reflections recorded from the animals.

$$\text{TS}_{\text{rel}} = 10\log_{10}\frac{I_{\text{ref}}}{I_{\text{inc}}} = \left(10\log_{10}\frac{I_{\text{inc}}A_{\text{inc}}\rho}{2\pi}\right) \Big/ I_{\text{inc}}$$
$$= 10\log_{10}\frac{A_{\text{inc}}\rho}{2\pi}. \quad (4)$$

This equation yields a TS_{rel} of −41 dB for the lateral aspect, the most common orientation for our studies. However, since this expression assumes that the reflected energy is uniformly scattered into the backward hemisphere, the estimate is probably low. The other extreme, that is total specular reflection, neglects the factor of 2π and yields a TS_{rel} of −30 dB, a value more consistent with those observed.

One surprising result of the experiments was that in a large number of cases the reflected energy was unobservable, even though a manatee was clearly in the sonar beam. Since the large air-breathing animals were assumed to reflect a large fraction of the sound, initial impressions of this lack of reflected energy were puzzling. In addition, based on the absorption estimates, as described above, it seemed unlikely that the animal was acting as a sound sink. In lieu of the above tissue measurements, the most likely reason for the

TABLE II. A summary of the reflection coefficients for the various regions.

Reflection coefficient Lateral	CT	OB	OM	IB	IM
Mean	0.129	−0.141	n/a	n/a	n/a
Standard deviation	0.029	0.054	n/a	n/a	n/a
No. of samples (n)	7	7	n/a	n/a	n/a
Dorsal					
Mean	0.113	n/a	−0.072	n/a	n/a
Standard deviation	0.019	n/a	0.006	n/a	n/a
No. of samples (n)	9	n/a	2	n/a	n/a
Ventral					
Mean	0.067	−0.005	0.024	−0.007	0.059
Standard deviation	0.028	0.078	0.053	0.094	0.076
No. of samples (n)	6	6	6	6	6

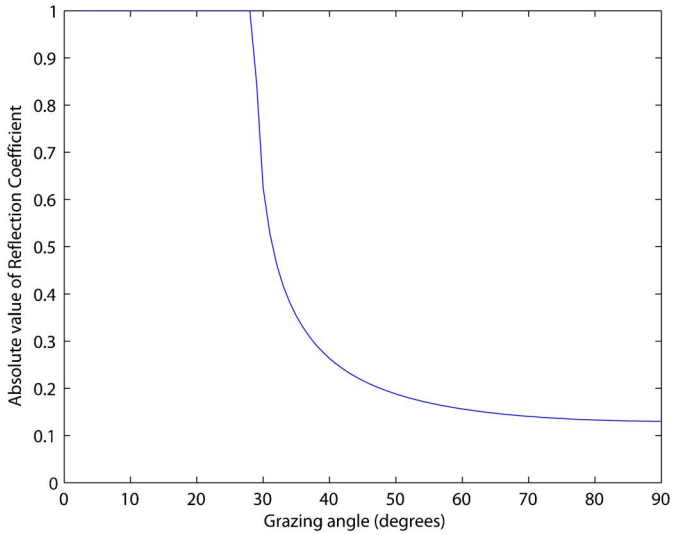


FIG. 5. (Color online) A plot of reflectivity versus grazing angle using Eq. (2) and the data from the lateral umbilicus [1708 m/s and 1149 kg/m^3 (from Table I)] relative to the fresh water sound speed and density measurements in the Sea World manatee pool (1510 m/s and 1000 kg/m^3).

discrepancy is that the animal acts as a specular reflector at the frequency tested. This implies that the sound was reflected, but at an unobservable angle between source and receiver. Future work to confirm this hypothesis would require three-dimensional characterization of the animal's reflectivity. In the context of the observations reported here it is also interesting to note that a report by Dickerson *et al.* (1996) describes the difficulty of obtaining adequate acoustic reflections from the animals. That study was initiated in order to create a manatee detection system in order to prevent deaths in canal locks. The innovative solution considered there was to devise a system that detected the interruption of the acoustic signal, much the same as an electro-optic door opener, in order to detect the presence of the animal and, hence, prevent the locks from closing.

Based on results discussed in this article, a detection frequency of 171 kHz does not appear to be reliable enough to use for detection. Reconciliation of the tissue samples with the observations strongly suggests that the animals' skin acts as a specular reflector, resulting in the reflection of a large part of the signal, but not necessarily back at the observer. Clearly what is needed is a more omni-directional scatter pattern. It is possible that other frequencies might offer a substantial advantage in this regard as the surface of the animal will appear to have different roughness depending on undulations relative to the frequency of incident sound.

Lastly, any true program to detect manatees acoustically should take into consideration the propagation of sound in their native habitat. Shallow water channels with either muddy or smooth rock bottoms present challenging environments in which to work. In this regard, higher frequency systems with narrow beam patterns can be used to minimize interactions with the animals' environment. However, only a truly systematic study of animal reflectivity versus frequency can determine whether active acoustic systems can help to preserve these dwindling marine animals.

ACKNOWLEDGMENTS

The authors would like to thank the Florida Fish and Wildlife Conservation Commission, especially Katherine Frisch and Elsa Haubold, for support and advice throughout this project. We thank Dr. Dave Demer (Southwest Fisheries Science Center) for the loan of a set of calibrated tungsten carbide spheres and Jerry Smith (UCSD/SIO) for the loan of the transducers. Dr. Graham Worthy, University of Central Florida and Hubbs-SeaWorld Research Institute (HSWRI), kindly collected the manatee tissue samples, which were provided by the Marine Mammal Pathobiology Laboratory in St. Petersburg, FL. Dr. Robert Shadwick (UCSD/SIO) let us use his lab for the tissue measurements and Jeremy Goldbogen (UCSD/SIO) helped with the dissections. D. V. Holliday and C. Greenlaw (BAE Systems) kindly provided loan of the calibrated E-8 transducer. We thank the Animal Care Staff at SeaWorld's Manatee Rescue Facility for providing access to the manatees and helping us during experiments. Tina Yack, Chris Martindell, and Jennifer Keating assisted with the setup and collection of data at SeaWorld. Experiments with live manatees were conducted under permit from the U.S. Fish and Wildlife Service, Division of Management Authority (MA054026-0, issued to Dr. Bowles).

Agemura, D. H., O'Brien, Jr., W. D., Olerud, J. E., Chun, L. E., and Eyre, D. E. (**1990**). "Ultrasonic propagation properties of articular cartilage at 100 MHz," J. Acoust. Soc. Am. **87**, 1786–1791.

Au, W. L. (**1996**). "Acoustic reflectivity of a dolphin," J. Acoust. Soc. Am. **99**, 3844–3848.

Bowles, A. E. (**2002**). "Design for a Manatee Finder: Sonar Techniques to Prevent Manatee-Vessel Collisions," report by Hubbs-SeaWorld Research Institute for Florida Marine Research Institute, Manatee Avoidance Technology Program. Purchase Order S-7701–617591, 25 November.

Bowles, A. E., Alves, C. D., and Anderson, R. C. (**2001**). "Manatee behaviors in the presence of fishing gear: response to novelty and the potential for reducing gear interactions," report by Hubbs-SeaWorld Research Institute, San Diego, CA, for U. S. Fish and Wildlife Service, Jacksonville, FL on Purchase Order 401819M390. Hubbs-SeaWorld Research Institute Technical Report No. 2001-317.

Craig, B. A., and Reynolds III, J. E. (**2004**). "Determination of manatee population trends along the Atlantic coast of Florida using a Bayesian approach with temperature-adjusted aerial survey data," Marine Mammal Sci. **20**(3), 386–400.

Dickerson, D. D., Reine, K. J., Newton, J. C., and Dickerson, Jr., C. E. (**1996**). "Evaluation of hydroacoustic techniques for detecting manatees at water control structures," report by U.S. Army Corps of Engineers, Waterways Experiment Station, Vickburg, MS, February.

Dweltz, N. E. (**1962**). *Collagen*, edited by N. Ramanathan (Interscience,

New York), p. 179.
Florida Manatee Biological Review Panel (**2005**). "Preliminary Status Review of the Florida Manatee, *Trichechus manatus latirostris*," report by Florida Fish and Wildlife Conservation Commission, Fish and Wildlife Research Institute, December.
Foote, K. G. (**1990**). "Spheres for calibrating an eleven-frequency acoustic measurement system," J. Cons., Cons. Int. Explor. Mer **46**, 284–286.
Gerstein, E. R. (**2002**). "Manatees, bioacoustics and boats," Am. Sci. **90**(2), 154–163.
Gerstein, E. R., Gerstein, L., Forsythe, S. E., and Blue, J. E. (**1999**). "The underwater audiogram of the West Indian manatee (*Trichechus manatus*)," J. Acoust. Soc. Am. **105**, 3575–3583.
Goss, S. A., and Dunn, F. (**1980**). "Ultrasonic propagation properties of collagen," Phys. Med. Biol. **25**(5), 827–837.
Hulmes, D. J. S., Miller, A., White, S., and Doyle, B. B. J. (**1977**). "Interpretation of the meridional X-ray diffraction pattern from collagen fibers in terms of the known amino acid sequence," J. Mol. Biol. **1**(10), 643–666.
Jaffe, J. S., Ohman, M. D., and De Robertis, A. (**1998**). "OASIS in the sea: measurement of the acoustic reflectivity of zooplankton with concurrent optical imaging," Deep-Sea Res., Part II **45**(7), 1838–1844.
Johnson, C. S. (**1967**). "Sound detection thresholds in marine mammals," in *Marine Bioacoustics, Vol. 2*, edited by W. N. Tavolga (Pergamon, Oxford, UK), pp. 247–260.
Kino, G. S. (**1987**). *Acoustic Waves: Devices, Imaging, and Analog Signal Processing* (Prentice-Hall, Englewood Cliffs, NJ).
Kipps, E. K., McLellan, W. A., Rommel, S. A., and Pabst, D. A. (**2002**). "Skin density and its influence on buoyancy in the manatee (*Trichechus manatus latirostris*), harbor porpoise (*Phocoena phocoena*) and bottlenose dolphin (*Tursiops truncatus*)," Marine Mammal Sci. **18**(3), 765–778.
Langtimm, C. A., Beck, C. A., Edwards, H. H., Fick-Child, K. J., Ackerman, B. B., Barton, S. L., and Hartley, W. C. (**2004**). "Survival estimates for Florida manatees from the photo-identification of individuals," Marine Mammal Sci. **20**(3), 438–463.
Marine Mammal Pathobiology Laboratory (**2005**) "Manatee Mortality Table," Florida Fish and Wildlife Conservation Commission, Fish and Wildlife Research Institute, http://www.floridamarine.org/ Last viewed 12/6/06.
Miller, J. H., and Potter, D. C. (**2001**). "Active high frequency phased-array sonar for whale shipstrike avoidance: target strength measurements," MTS/IEEE Oceans 2001 Conference Proceedings, pp. 2104–2107.
National Marine Manufacturer's Association (**2004**) "Recreational Boating Statistical Abstract," National Marine Manufacturer's Association, Chicago, IL (available at http://www.nmma.org/facts/boatingstats/2004/). Last viewed 12/6/06.
Niezrecki, C., Phillips, R., Meyer, M., and Beusse, D. O. (**2003**). "Acoustic detection of manatee vocalizations," J. Acoust. Soc. Am. **114**, 1640–1647.
Nowacek, S. M., Wells, R. S., Owen, E. C. G., Speakman, T. R., Flamm, R. O., and Nowacek, D. P. (**2004**). "Florida manatees, Trichechus manatus latirostris, respond to approaching vessels," Biol. Conserv. **119**, 517–523.
Schevill, W. E., and Watkins, W. A. (**1965**). "Underwater calls of *Trichechus* (manatee)," Nature (London) **205**(4969), 373–374.
Schell, C., and Jaffe, J. S. (**2004**). "Experimental verification of an interpolation algorithm for improved estimates of animal position," J. Acoust. Soc. Am. **116**, 254–261.
Sokolov, V. E. (**1982**). *Mammal Skin* (Univ. of California, Berkeley, CA).
Urick, R. J. (**1967**). *Principles of Underwater Sound for Engineers* (McGraw Hill, New York).
Vagle, S., Foote, K. G., Trevorrow, M. V., and Farmer, D. M. (**1996**). "A Technique for Calibration of Monostatic Echosounder Systems," IEEE J. Ocean. Eng. **21**(3), 298–305.
White, A., Handler, P., and Smith, E. L. (**1968**). *Principles of Biochemistry* (McGraw-Hill, New York).

Comparison of ocean-acoustic horizontal coherence predicted by path-integral approximations and parabolic-equation simulation results

Michael D. Vera[a)]
The University of Southern Mississippi, Hattiesburg, Mississippi 39406

(Received 18 May 2006; revised 2 October 2006; accepted 3 October 2006)

A line-integral approximation to the acoustic path integral has been used to generate predictions for the characteristic length scale of horizontal, cross-range coherence in long-range ocean-acoustic propagation. These estimates utilize a single range-independent sound-speed profile and the mean variance, as a function of depth, of fractional sound-speed perturbations due to internal waves. The length scales predicted by the integral approximation have been compared to the values generated by parabolic-equation simulations through multiple realizations of Garrett-Munk internal waves. One of the simulation environments approximates the Slice89 experiment; transmissions from a 250-Hz source were simulated in a deep-water transect to a maximum range of 1000 km. The second environment corresponds to one of the propagation paths in the North Pacific Acoustic Laboratory (NPAL) experiment. The source in this experiment was bottom-mounted near Kauai, Hawaii and the relevant receiver consisted of five vertical line arrays oriented transverse to the propagation path with cross-range separations ranging from approximately 500 to 3500 m. The receiver was at a range of 3889.8 km from the source. The predicted length scales are consistently shorter than the parabolic-equation results by 30%–80%, depending on the range and environment examined. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2382345]

PACS number(s): 43.30.Zk, 43.30.Pc, 43.30.Qd, 43.30.Re [AIT] Pages: 166–174

I. INTRODUCTION

A series of ocean-acoustic experiments with ranges of a megameter and beyond has demonstrated that acoustic arrivals can be detected, identified, and inverted for ocean properties at basin-scale ranges. These include the Slice89 experiment,[1] the Acoustic Thermometry of Ocean Climate (ATOC) project and its associated Acoustic Engineering Test (AET),[2] and the North Pacific Acoustic Laboratory (NPAL) experiment.[3] Internal waves can often be regarded as the dominant contributor to small-scale acoustic fluctuations in these experiments.[4–6] A model spectrum for internal waves, vertical oscillations of water about equilibrium density, was developed by Garrett and Munk.[7,8] This spectrum has been used to examine the effect of internal waves on acoustic signal characteristics such as travel-time variability and vertical coherence.[4,9]

Predictions from two different numerical approaches for the horizontal, cross-range coherence of acoustic arrivals are presented here: approximations to the acoustic path integral and simulations using a parabolic equation. Both of these methods yield estimates of horizontal coherence based on the effects of a stochastic field of internal waves conforming to the Garrett-Munk spectrum. Neither method accounts for the role played by other features, such as mesoscale, in the final coherence. Each numerical method is applied to two different acoustic propagation environments corresponding to experimental geometries. The first model environment is based upon the Slice89 experiment. This model corresponds to the one used in previous tests of the performance of path-integral approximations.[9] The second simulation environment corresponds to one of the transects in the North Pacific Acoustic Laboratory (NPAL) experiment. A two-dimensional receiver array was deployed near the coast of California as part of the NPAL project. This 2D array consists of five vertical line arrays (VLA) of receivers at a range of 3889.8 km oriented transverse to the propagation path from the source near Kauai, Hawaii. The design of this receiver was motivated by a desire to investigate the coherence of acoustic arrivals separated in the horizontal, cross-range direction. The components of this receiver allow for an examination of horizontal coherence at ten cross-range separations ranging from approximately 500 to 3500 m. Aspects of the experimental horizontal coherence behavior at this receiver have been investigated.[10,11] A detailed comparison between individual arrivals in the experimental data and the numerical models presented here remains an area for future research.

Propagation of sound in the ocean interior can be effectively modeled using a parabolic approximation to the full acoustic wave equation. Hardin and Tappert first applied a parabolic wave equation to the problems of ocean acoustics and presented an efficient split-step Fourier algorithm for its solution.[12,13] Simulation of acoustic propagation through ocean internal waves using a parabolic equation was first described by Flatté and Tappert; they examined a 100-Hz signal at a range of 250 km.[14] Important advances in the parabolic-equation method were made by Collins;[15] he developed an improved method based upon Padé expansions of an operator in the wave equation.

[a)]Electronic mail: michael.vera@usm.edu

 0001-4966/2007/121(1)/166/9/$23.00

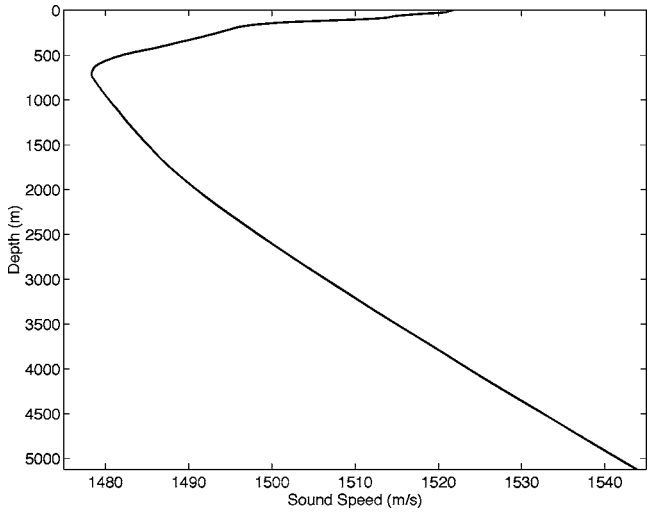


FIG. 1. The sound speed used in the Slice89 simulations is shown as a function of depth. The acoustic source in these simulations is located at 800-m depth, near the sound-speed minimum.

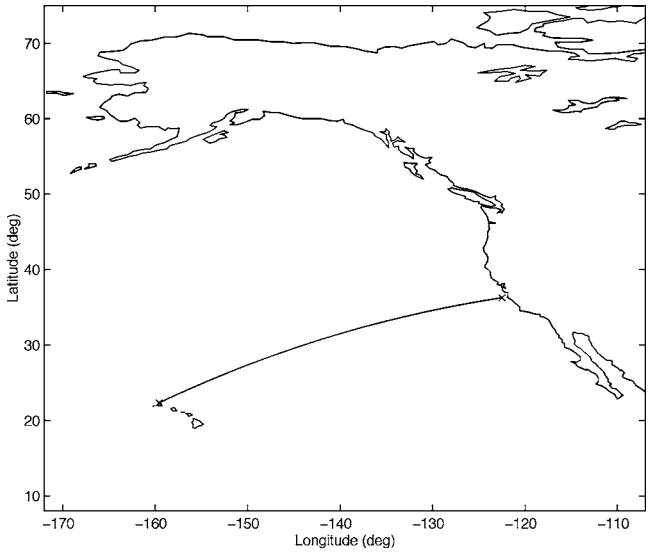


FIG. 2. The propagation path from the source near Kauai, Hawaii to the 2D receiver array near California is shown as a solid line. The total range is 3889.8 km. This transect is included in the NPAL network. The model environment discussed in the text is based on environmental measurements along this path.

The parabolic equation, in the "standard" form presented by Hardin and Tappert, is mathematically related to the Schrödinger equation,[16] so it can be formally solved using the method of path integrals. Approximations to the path integral based upon line integrals along deterministic ray paths were developed for the problems of ocean acoustics.[17–19] The accuracy of these estimates for acoustic fluctuations can be determined by a comparison to parabolic-equation simulations through multiple realizations of internal waves conforming to the Garrett-Munk spectrum. A prior investigation examined the predictions for travel-time statistics and vertical coherence.[9] The characteristic length scale of the mutual coherence function for horizontal separations predicted by the integral approximation is compared here to simulations of acoustic propagation using a parabolic equation. Numerical internal-wave fields were generated in three space dimensions; two-dimensional slices through this field were used in simulations of acoustic propagation along five different horizontal azimuths, with specified cross-range separations at the final range.

II. SLICE89 PROPAGATION MODEL

One of the modeled propagation environments used in the simulations below corresponds approximately to the Slice89 experiment performed in the North Pacific. The source for this experiment had a center frequency of 250 Hz, and a bandwidth of 100 Hz. The source depth in the simulations was 800 m. A vertical line array (VLA) of receivers recorded the arrivals at a range of 1000 km. The propagation occurred entirely in deep water; acoustic interaction with the bottom did not significantly impact the received field.[1,4,20]

The environment used in the following calculations is an idealized version of the Slice89 case. The same parameters were used as in a prior investigation into the accuracy of certain path-integral approximations.[9] The simulations occurred in a model environment with a constant bottom depth of 5118.75 m. The single profile of sound speed shown in Fig. 1 was used. Though only a single receiving array was involved in the experiment, the field in the simulations was examined at two different ranges and five discrete cross-range separations. At the maximum range of 1000 km, the five simulated transects had cross-range separations between adjacent receivers ranging from 300 to 1200 m with a maximum separation between receivers of 3250 m.

III. NORTH PACIFIC ACOUSTIC LABORATORY EXPERIMENT

A broadband ($Q=2$) source with a center frequency of 75 Hz located to the north of Kauai, Hawaii was employed in both the Acoustic Thermometry of Ocean Climate (ATOC) and NPAL experiments and continues to operate. The source is bottom-mounted; its acoustic center is approximately 2 m above the seafloor.[21] The water depth at this location is about 800 m. A two-dimensional receiver array located near the coast of California recorded acoustic arrivals on five vertical line arrays oriented transverse to the propagation geodesic. This path is depicted in Fig. 2. The range from the source to the center VLA was 3889.8 km. Horizontal separations between adjacent VLAs ranged from approximately 500 m to nearly 1200 m. The distance between the most widely separated VLAs was approximately 3.5 km.[3]

A series of environmental measurements was performed in support of the NPAL experiment during the summers of 1998 and 1999. Objective maps of salinity and temperature were derived from a high-resolution CTD/XBT survey. The buoyancy frequency profile used to generate internal-wave fields in the simulations was generated from the 1999 measurements. The sound-speed field used in the calculations discussed below was derived from the 1999 environmental measurements using the Del Grosso equation for sound speed.[22] Adjacent profiles of temperature and salinity from an objective map with a horizontal spacing of 39 km were

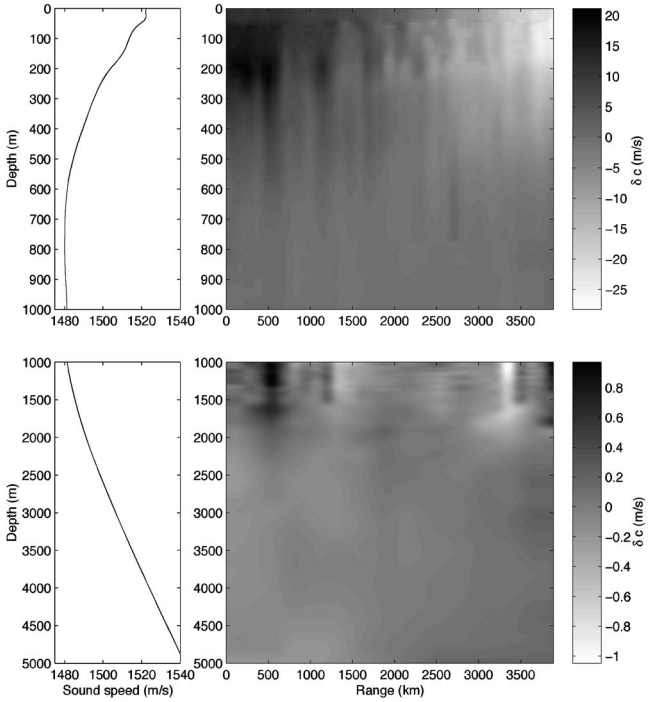


FIG. 3. Sound speeds derived from environmental measurements taken during the summer of 1999. The left panels contain the range-averaged sound-speed profile for different intervals of depth. The right panels display the difference from this average as a function of both range and depth. This sound-speed field is perturbed in the model by Garrett-Munk internal waves.

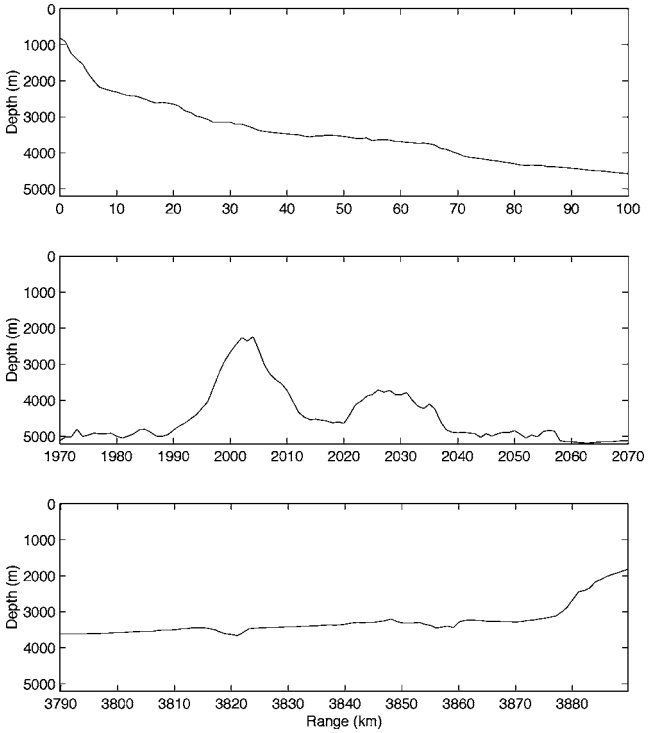


FIG. 4. Bathymetry measurements are shown for three regions of range along the propagation transect. The top panel is the source region. The source is bottom-mounted at approximately 800-m depth. The middle panel displays a seamount at roughly half of the total range. The bottom panel shows the upslope as the receiver is approached; the final depth is about 1800 m. Acoustic interaction with the bathymetry in these regions complicates the simulated arrival structure.

averaged in order to suppress small-scale features such as internal waves. The resulting sound speed is displayed in Fig. 3.

Along the transect from the Kauai source to the 2D NPAL receiver array, SEABEAM bathymetry measurements were performed with a horizontal resolution of approximately 50 m.[3] These bathymetry measurements proved critical to successfully modeling this portion of the NPAL experiment. The geometry of the seafloor plays a prominent role in the structure of the far-field acoustic arrival pattern.[23] The most important regions of acoustic interaction with the bathymetry are depicted in Fig. 4. Reflection from the downslope near the source generates acoustic arrivals that evolve differently in range than the refracted portion of the field. A seamount was detected at approximately half the range reaching to a depth of approximately 2500 m; this feature leads to additional complications in the arrival structure. As the signal approaches the receiver, the water depth again decreases from abyssal values to about 1800 m. Propagation along this transect leads to an acoustic arrival structure that is a complicated mixture of refracted and reflected components of the field.

IV. SIMULATION TECHNIQUES

Simulated three-dimensional stochastic internal-wave fields that conform to the Garrett-Munk spectral model were generated for each environment using a method presented by Colosi and Brown.[24] Two-dimensional slices through these fields provided the displacements in the range-depth plane connecting the source to each of the receivers in the simulations. These realizations of the internal-wave spectrum were used to perturb the "background" sound-speed structure displayed in Figs. 1 and 3.

The Garrett-Munk (GM) spectrum of internal-wave vertical displacement as a function of vertical mode number j and horizontal wave number k is

$$S_\zeta(j,k) = \frac{2B^2E}{\pi M}\frac{N_0}{N}\frac{1}{(j^2+j_*^2)}\frac{k_j k^2}{(k^2+k_j)^2}, \tag{1}$$

where $B=1000$ m is the thermocline depth scale, $E=6.3\times10^{-5}$ (for the reference internal-wave energy referred to as 1 GM), $N_0=3$ cph and $j_*=3$ are empirical constants, $k_j \equiv \pi\omega_i j/N_0 B$, ω_i is the inertial frequency from the Earth's rotation at the latitude of the internal waves, and M is the normalization constant for the sum over mode number, $M=\Sigma_{j=1}^{\infty}(j^2+j_*^2)^{-1}$.[7,8,25] The latitude (used to define the inertial frequency) for the Slice89 model was 30 deg. For the NPAL experiment, the latitude was taten as 30.2 deg, the average value along the path from the Kauai source to the 2D receiver array. The lowest 50 vertical modes were used. A single, range-independent buoyancy frequency profile, $N(z)$, was used in constructing the internal-wave fields for each case. A field of 1-GM internal waves has a root-mean-square vertical displacement of $\sqrt{\langle\zeta^2\rangle}=7.3$ m at the depth where $N(z)=N_0$; this strength was chosen for the simulations based upon mooring measurements of the perturbations in the actual NPAL experiment.[26]

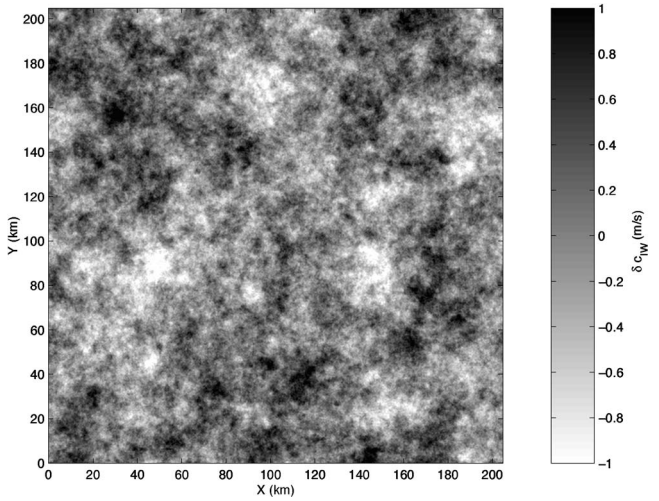


FIG. 5. An overhead plan view of the simulated sound-speed perturbations due to internal waves at the reference depth where the buoyancy frequency is equal to 3 cph. This numerical field, with an extent of 204.8 km in both horizontal directions, is converted into range-depth slices for the path from the source to each VLA. The conversion preserves correlations among the slices appropriate to the cross-range separation at each range. The method is described in the text.

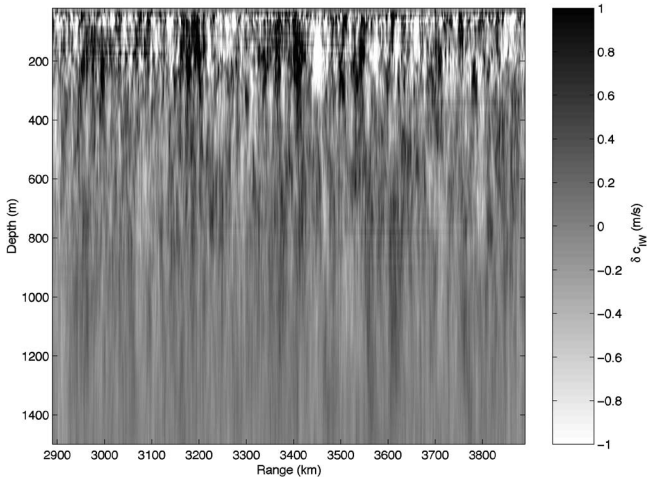


FIG. 6. The sound-speed perturbations caused by a sample range-depth slice of internal-wave displacements are shown. The final 1000 km of range and the shallowest 1500 m of the water column are displayed. The addition of these perturbations to the sound speed depicted in Fig. 3 gives the sound speed as used in the PE simulation.

Numerical internal-wave fields were constructed with an extent of 204.8 km in each of the horizontal directions using wavelengths ranging from 1 to 204.8 km in a Fourier expansion.[24] This interval of horizontal scales contains the dominant contributions in Eq. (1) as well as the effective internal wave correlation length of $O(10\ \text{km})$. The periodicity of the Fourier technique was exploited in order to use a $(204.8\text{-km})^2$ horizontal section (as shown in Fig. 5) as a building block in constructing range-depth slices for the entire range in each of the simulation environments. In order to avoid the introduction of numerical artifacts, the range direction was chosen to be as far as possible from the X, Y coordinate axes of the modeled field. As range increases from 0 to about 205 km and then from 205 to 410 km, (X, Y) goes from (0,0) to (204.8 km, 10 km) and then from (0 km, 10 km) to (204.8 km,20 km). This mitigates the artificial periodicity of the numerical internal-wave field; a given transect between the source and a specific receiver never samples the numerical field at effective separations of less than 10 km and this occurs only when 200 km of intervening range has been traversed.

The total sound speed can be written as

$$c(r,z) = c_{\text{background}} + \delta c_{\text{IW}}(r,z), \tag{2}$$

where $c_{\text{background}}$ is either a single, depth-dependent profile (Slice89 environment) or the smoothed field derived from the 1999 measurements (NPAL) and δc_{IW} is generated from the simulated internal waves. The sound-speed perturbation is calculated from the vertical displacement according to

$$\delta c = \left(\frac{\partial c}{\partial z}\right)_{\text{pot}} \zeta \approx \frac{1}{2} c_0 \left(\frac{\mu}{g}\right) N^2 \zeta, \tag{3}$$

where $c_0 = 1500$ m/s is a reference sound speed, g is gravitational acceleration, and $\mu = 24.5$.[24,26] A portion of a sample range-depth slice of internal-wave sound-speed perturbations is displayed in Fig. 6.

A. Path-integral approximations

The integral approximation to the length scale of horizontal coherence requires the mean-square value of the relative sound-speed perturbations, $\langle \mu^2(z) \rangle = \langle [\delta c_{\text{IW}}/c_0]^2 \rangle$, and an expression for the correlation length of the internal-wave disturbances along a ray with inclination θ and depth z, $L_p(\theta, z)$. The average fractional variance, $\langle \mu^2(z) \rangle$, was determined either by range-averaging a simulated internal-wave field (NPAL) or by a WKBJ approximation to internal-wave vertical structure[19](Slice89). The correlation length, $L_p(\theta, z)$, is determined by the empirical expression,[19,27]

$$L_p(\theta,z) = L_{p0} \frac{[1 - \exp[-(\sigma_c/\sigma_{\text{curv}})^p]]^q}{1 + (\rho_s/\rho_c)^{q_a} (\sigma_c/\sigma_{\text{curv}})^{p_a}}, \tag{4}$$

where $\rho_c = 3.5$, $\sigma_c = 0.0204$, $p = 0.385$, $p_a = 0.5$, $q = 1.3$, and $q_a = 2.0$ are dimensionless constants, and $L_{p0} = 12.7$ km. The quantity ρ_s is a measure of the anisotropy of the internal waves given by $\rho_s = [N(z)\tan\theta]/\omega_i$. The quantity σ_{curv} a dimensionless measure of the ray curvature, obtained by dividing the real ray radius of curvature by about 1000 km. Typical internal-wave correlation lengths are a few kilometers in the horizontal direction and $O(100\ \text{m})$ in the vertical.

These internal-wave parameters were used in an integral approximation for the horizontal length scale of acoustic coherence. The mutual coherence function of the acoustic field at two horizontally separated points can be expressed as[19]

$$C_{\text{integral}}(\delta y) \equiv \langle \psi^*(\delta y)\psi(0) \rangle \approx \exp\left[-\frac{1}{2}\left(\frac{\delta y}{y_0}\right)^{3/2}\right], \tag{5}$$

where ψ is the complex demodulated acoustic pressure, δy is the horizontal separation between the points of interest, and y_0 is the length scale of horizontal coherence.

The value of y_0 was determined by the following line integral along a ray path:[19]

$$y_0^{-3/2} = 2q_0^2 \int_0^R dx \langle \mu^2(z) \rangle L_p \left[\frac{x}{R} + \frac{\delta y_s}{\delta y_r} \left(1 - \frac{x}{R} \right) \right]^{3/2} y_h^{-3/2}, \tag{6}$$

where q_0 is the acoustic wave number, R is the total range, δy_s is the separation of ray paths at the source range (set to 0), δy_r is the separation at the receiver, and y_h characterizes the correlation of internal-wave perturbations in the horizontal. Acoustic horizontal coherence can be equivalently expressed as the root-mean-square fluctuations in the horizontal arrival angle given by $\theta_h = (\delta y_r q_0^4 y_0^3)^{-1/4}$.

The length scale of horizontal coherence was calculated using a computer code package referred to as "CAFI" (Computation of Acoustic Fluctuations due to Internal waves) prepared by Flatté and Rovner.[19] A single range-independent profile of sound speed was used, $c(z) = \langle c_{\text{background}}(r,z) \rangle$. For the Slice89 environment, 7201 ray paths were calculated. Only rays determined to be entirely refracted (no reflection from the surface or the flat bottom) were used. Approximately 1600 ray paths were calculated for the NPAL case, with launch angles ranging from −5.5 to 10.4 deg ($\delta\theta_{\text{launch}} = 0.01$ deg). This region of launch angle was found to yield fully refracted rays in a ray simulation that allowed for specular reflection from the seafloor.[23] Ray paths that corresponded to the arrival depths and timefront segments of interest were included. The reasons for selecting specific timefront segments are discussed in Sec. V. Depth intervals were chosen to avoid caustics. The predictions for horizontal coherence scale from all suitable paths were incorporated into an average for each range and propagation environment. Typical values for the calculated length scale of horizontal coherence, y_0, are approximately 300 m for the Slice89 case at 1000 km, and 600 m for the NPAL case at 3889.8 km.

B. Parabolic-equation simulations

Simulations of acoustic propagation were performed for each environment using the RAM code developed by Collins.[15] The geoacoustic parameters for the measured NPAL bathymetry were estimated and incorporated into the simulation using a complex-density equivalent fluid.[23,28] The modeled parameters for the homogeneous bottom in the simulations were a compressional speed of 2200 m/s, a shear speed of 1100 m/s, a compressional attenuation of 0.1 dB/λ, and a shear attenuation of 0.2 dB/λ. These same bottom parameters were used for the flat bottom in the Slice89 simulations, but acoustic interaction with the bottom was insignificant in this case. The computational grid had a vertical spacing of 0.25 m and horizontal range steps of 25 m. Four Padé terms were included in the approximation for the parabolic-equation operator.[15,29] This computational grid was based upon acceptable convergence of the transmission loss at the central frequency. A set of 401 (Slice89) or 451 (NPAL) discrete frequencies was used to span the frequency range from $f_c - \frac{1}{4} f_{bw}$ to $f_c + \frac{1}{4} f_{bw}$; this interval is from 225 to 275 Hz for Slice89 transmissions and 56.25 to 93.75 Hz for NPAL transmissions. This accounts for the most energetic, central half of the total bandwidth. The outer regions of the full bandwidth were neglected. The sound speed displayed in Fig. 3 was interpolated onto 300 equally spaced ranges in the NPAL simulations.

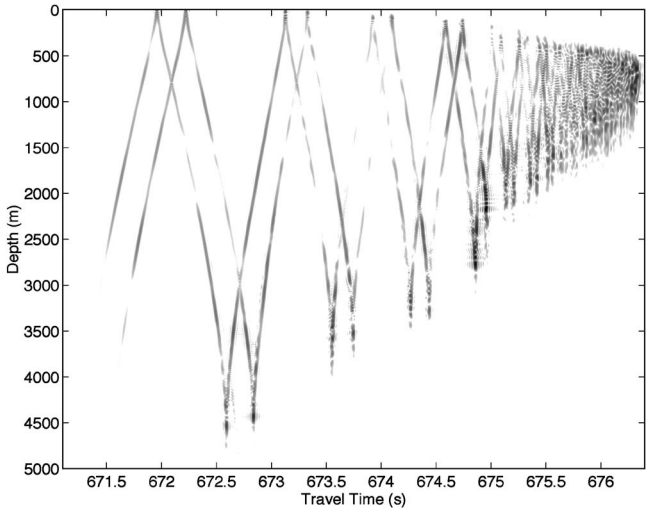


FIG. 7. A sample timefront from a Slice89 simulation at a range of 1000 km. The front shows the expected effects of 1-GM internal waves. A dynamic range of 30 dB is used.

Multiple realizations of the stochastic internal-wave field were used in simulations to each of the five vertical line arrays for each environment. Eight realizations were used in NPAL simulations. The 1-GM internal waves induced sound-speed perturbations with a root-mean-square value of 0.9 m/s at about 70-m depth (maximum), and about 0.3 m/s at 500-m depth. Four realizations of 1-GM internal waves were used in the Slice89 case. The effect of the sound-speed perturbations on the timefronts is displayed in Figs. 7 and 8.

V. COMPARISON

The process of comparing the parabolic-equation (PE) simulations and the integral prediction begins with the selec-

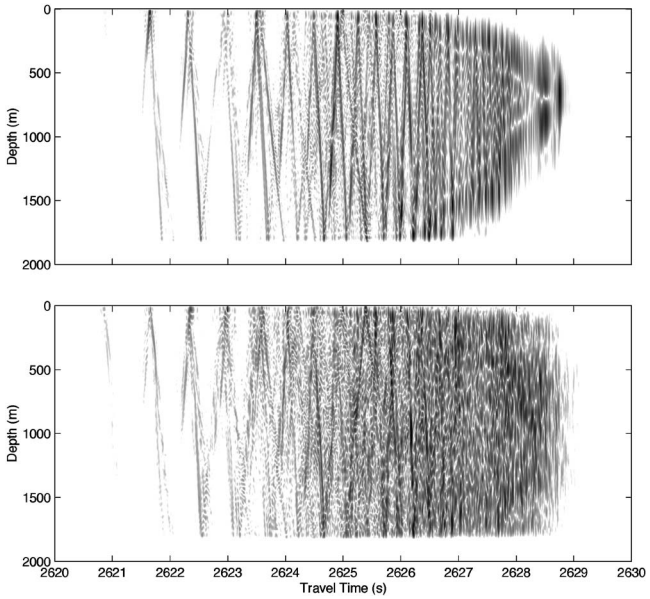


FIG. 8. Sample timefronts for propagation from the Kauai source to the 2D NPAL receiver array. A simulation without internal waves yielded the front in the top panel. The bottom panel shows the effect of 1-GM internal waves. A dynamic range of 30 dB is used in both cases.

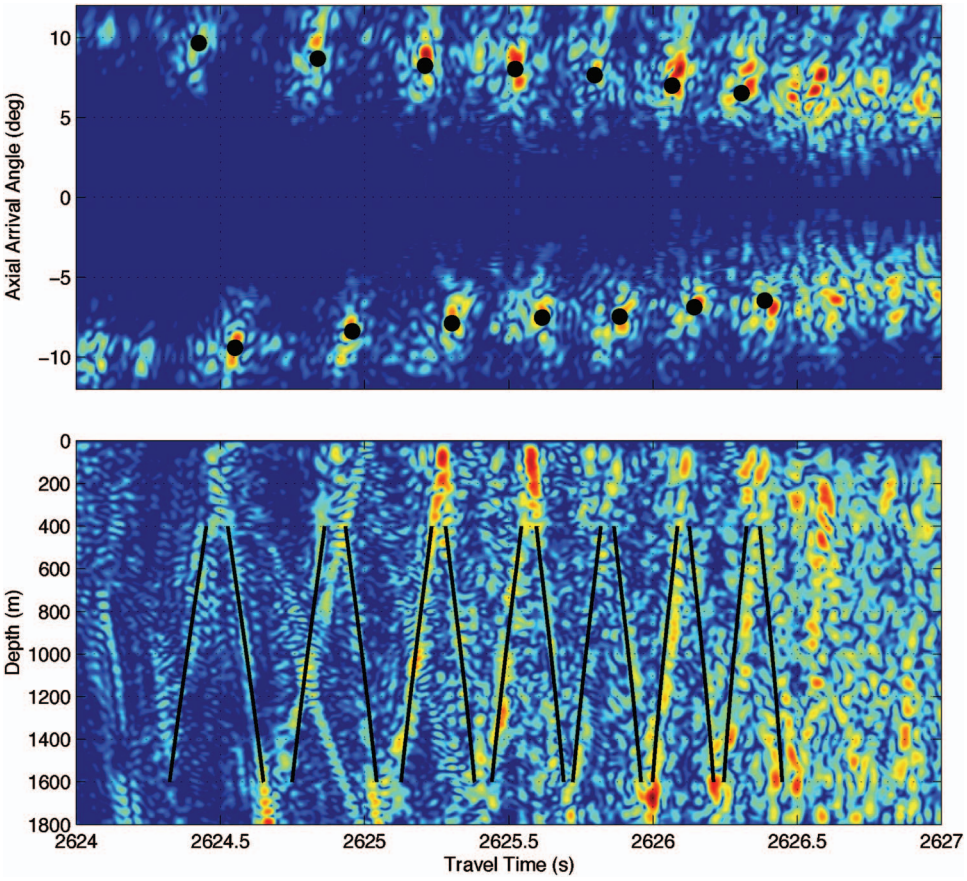


FIG. 9. The method for selecting arrivals from a sample 1-GM simulation is shown. The top panel shows the turning-point filter representation of the simulation; the selected peaks are marked with black circles. The bottom panel displays the corresponding timefront; the black curves result from applying the inverse of the turning-point filter to the points from the top panel. The color scale uses a dynamic range of 30 dB.

tion of corresponding arrivals from the simulations. The results of ray simulations were used as a template in the selection. Only fully refracted rays were used. This is because bottom interaction was not treated in the integral calculation. The simulation fields for the NPAL case also contain acoustic energy that has reflected from the seafloor; the comparison requires that the relevant refracted energy be extracted from the field at the range of interest. Turning-point filter representations[30] of the simulations were compared to the refracted-ray template. Application of the turning-point filter transforms the acoustic field into a plane with coordinates of travel time and axial arrival angle, the inclination of the acoustic energy at the sound-speed minimum. For both sets of simulations, the maximum in the quality function,

$$Q = |T^2(\theta_{\mathrm{ax}}, t)| \cos\left(\frac{\pi \delta\theta}{\theta_W}\right) \cos\left(\frac{\pi \delta t}{t_W}\right), \tag{7}$$

was located for each ray of interest where $T(\theta_{ax}, t)$ is the complex turning point filter field, $\delta\theta$ and δt are the differences between the point of the field under consideration and the ray point, $t_W = 0.2$ s is the chosen scale for time separations, and $\theta_W = 2°$ is the scale for separations in axial arrival angle. Finally, the turning-point filter transformation was inverted to generate curves that approximate the arrivals in the (z,t) plane from each of the points in the $(\theta_{\mathrm{ax}}, t)$ plane for depths from 500 to 2500 m (Slice89) or 400 to 1600 m (NPAL).

The presence of bottom-interacting acoustic energy complicated the selection of suitable NPAL arrivals in the simulations. At both of the examined ranges, an interval of travel time was examined that was late enough that the acoustic field contains refracted arrivals, but early enough that individual timefront segments could be extracted. This travel-time interval is from 2624 to 2626.5 s for a range of 3889.8 km. Earlier arrivals have reflected from the seafloor near the source.[23] Each of the chosen segments can be assigned a timefront ID number[2] based upon its identification with a ray path. The 14 refracted ID segments considered in the NPAL analysis at the maximum range were +158 through +171. An example of this selection procedure is depicted in Fig. 9.

The selection of suitable arrivals from the Slice89 results was straightforward. The ten ID numbers used in the Slice89 case included both positive and negative values: −39 through −35 and +36 through +40. An alternate selection procedure was also investigated in the Slice89 case at a range of 1000 km. Using the curves in the (z,t) plane from turning-point filter inversion as a template, local maxima in intensity were selected directly from the timefront according to the highest value of the quality function,

$$Q = I_{\mathrm{cand}} \left(1 - \sqrt{2 \frac{|t_{\mathrm{cand}} - t_{\mathrm{tmplt}}|}{t_Q}}\right), \tag{8}$$

where I_{cand} and t_{cand} are the intensity and travel time of the candidate peak, t_{tmplt} is the template time for the relevant ID number and depth value, and $t_Q = 160$ ms is a time scale for the quality function.[9] Coherence calculations from this alternate procedure were within statistical uncertainty of those using the original method, and calculations of the

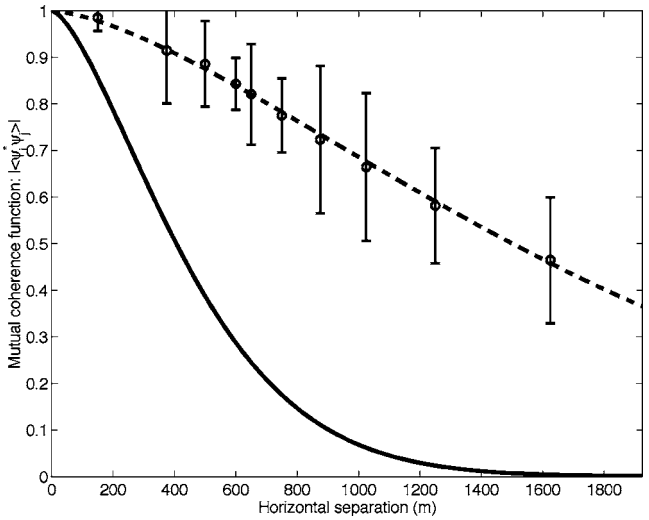


FIG. 10. The mutual coherence function with the length scale predicted by the integral method [see Eqs. (5) and (6)] is shown as a solid line for Slice89 propagation to a range of 500 km. The points with error bars were derived from PE simulations. The dashed curve is the best fit to the form of Eq. (5).

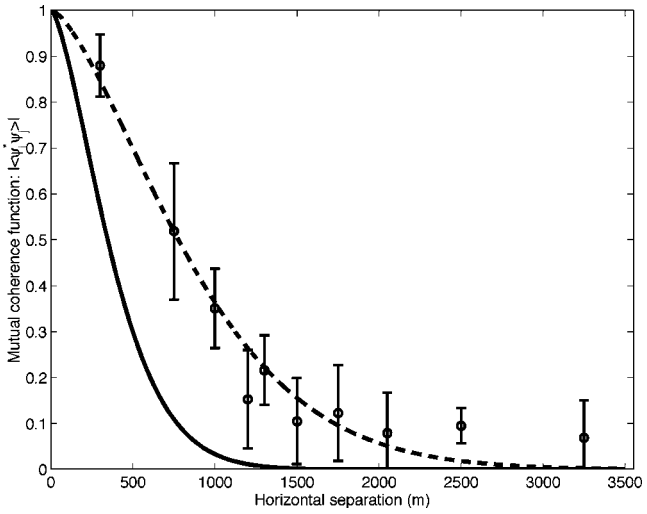


FIG. 11. The mutual coherence function with the length scale predicted by the integral method [see Eqs. (5) and (6)] is shown as a solid line for Slice89 propagation to a range of 1000 km. The points with error bars were derived from PE simulations. The dashed curve is the best fit to the form of Eq. (5).

final length scale of horizontal coherence differed by only about 5%. This agreement supports the validity of the more flexible turning-point-filter inversion technique.

After selection of the arrival segments at each receiver from the simulations, the mutual coherence function for horizontal separations was calculated. The value of the complex acoustic field was determined along the nearly linear segments generated by the selection procedure for each receiver and internal-wave realization. These complex numbers were used to calculate the coherence function,

$$C_{\mathrm{PE}}(\delta y) = \frac{\langle \psi^*(y;z,ID)\psi(y+\delta y;z,ID)\rangle}{\sqrt{\langle|\psi^2(y;z,ID)|\rangle\langle|\psi^2(y+\delta y;z,ID)|\rangle}}, \tag{9}$$

where y is the position of the first VLA under consideration, δy is the horizontal separation of the second set of receivers from the first, and ψ is the value of the complex field along the curve generated by the selection procedure. The mutual coherence function from the simulations is an average over internal-wave realization, the selected depth region, and the relevant timefront segments. The values for different depths on a given segment are correlated. The coherence scales from the different ID segments examined were also similar. The statistical uncertainty depicted in Figs. 10–13 is due to the variance across different internal-wave realizations.

The scale of acoustic horizontal coherence from the integral calculations was averaged similarly. Those ray paths that arrive at the receiver range at an appropriate depth with a relevant timefront ID number were incorporated into an average of $y_0^{-3/2}$. The integral prediction for the behavior of the mutual coherence [see Eq. (5)] as a function of cross-range separation δy can be stated as $\exp\left[-\frac{1}{2}\delta y^{3/2}\langle y_0^{-3/2}\rangle\right]$. The mutual coherence functions from the PE simulations and from the integral technique are displayed for the Slice89 environment in Figs. 10 (500-km range) and 11 (1000-km range). The NPAL results are also shown for two ranges in Figs. 12 (1000-km range) and 13 (3889.8-km range). The length scale derived from the path-integral approximations is consistently smaller than the length scale determined from the PE simulations. A fit of the PE results to the form of Eq. (5) yields longer characteristic length scales than the predictions; the predictions for y_0 underestimate the simulation values by 30%–80%. The y_0 values from the integral predictions and from the simulations are collected in Table I.

VI. CONCLUSION

A previous comparison of the predictions of path-integral approximations to parabolic-equation simulations through simulated internal waves resulted in reasonable agreement for the variance of travel time, and significant discrepancies for the travel-time bias, the pulse spread, and the scale of vertical coherence. It was noted that these problematic predictions all involved geometric expressions for the vertical separation of nearby rays.[9] The necessary approximations for the statistical properties of the perturbation

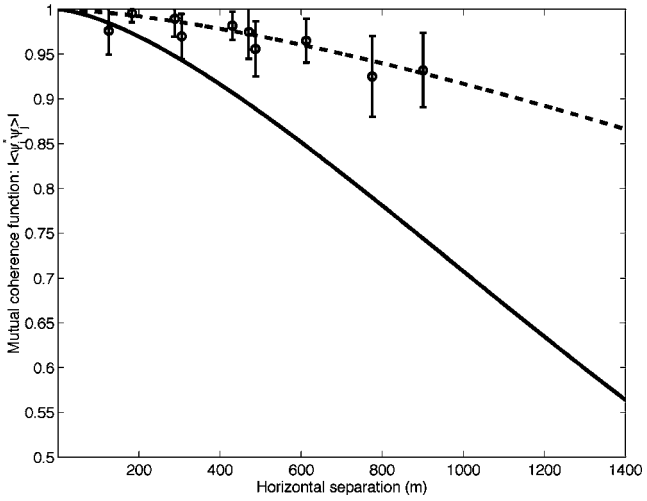


FIG. 12. The mutual coherence function with the length scale predicted by the integral method [see Eqs. (5) and (6)] is shown as a solid line for NPAL propagation to a range of 1000 km. The points with error bars were derived from PE simulations. The dashed curve is the best fit to the form of Eq. (5).

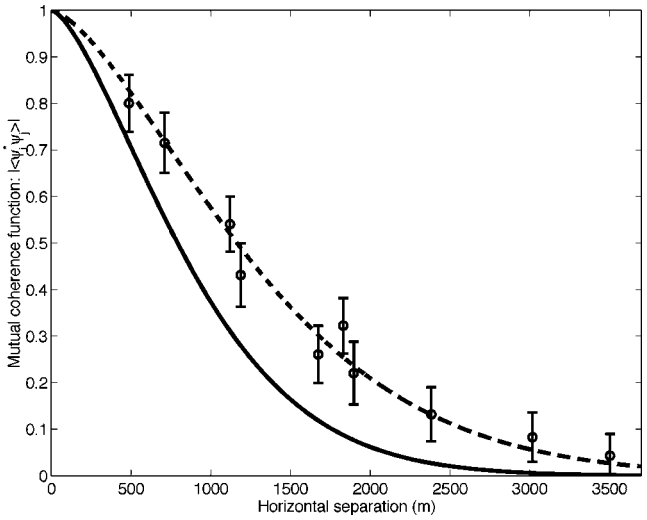


FIG. 13. The mutual coherence function with the length scale predicted by the integral method [see Eqs. (5) and (6)] is shown as a solid line for NPAL propagation to the maximum range of 3889.8 km. The points with error bars were derived from PE simulations. The dashed curve is the best fit to the form of Eq. (5).

field had apparently broken down over the depth intervals spanned by rays representing sound that could interfere constructively—the Fresnel zone scale.

The prediction for horizontal coherence examined here also involves an expression for the separation of nearby rays [see Eq. (6)]. This separation, however, is in the horizontal, cross-range direction. This expression is simpler (and apparently more tractable mathematically) than one for a vertical separation since the vertical refraction of the ocean sound channel does not play a role. The approximation for internal-wave properties over the separation of interfering rays is also less restrictive in this case. The internal-wave field is markedly anisotropic, with much shorter vertical than horizontal scales. Nevertheless, it may be the case that relaxing the assumption that ray separations are small [used in deriving Eq. (6)] could improve the accuracy of the technique at the expense of additional complexity and computational requirements.

The agreement between the integral prediction for the horizontal coherence scale and the simulation results is better than was found for those expressions involving vertical separation functions, even in a much more complicated model environment (NPAL). Nevertheless, there are substantial discrepancies in the coherence length estimates generated by the two approaches. Whether the method of path-integral approximation for horizontal coherence is a useful approach depends largely on accuracy requirements and the availability of detailed environmental data. For those applications in which environmental information is largely unknown and only approximate quantities are desired, the technique could be a useful initial guide. Such cases could include, perhaps, the initial stages of experimental design. The computational demands are far less than those of simulation and the analysis is much more straightforward. If, however, sufficient environmental information is available to permit the consideration of more stringent accuracy requirements, then computationally demanding simulations may be the preferred choice.

TABLE I. The y_0 values calculated for the different combinations of range and propagation environment are collected here. "PE" values are derived from the parabolic-equation simulations; "PI" values are the predictions generated by the integral approximations.

	PE y_0 (m)	PI y_0 (m)
Slice89, 250 Hz, 500 km	1200	330
Slice89, 250 Hz, 1000 km	630	280
NPAL, 75 Hz, 1000 km	3200	1300
NPAL, 75 Hz, 3889.8 km	940	640

ACKNOWLEDGMENTS

Portions of this research were performed at the Scripps Institution of Oceanography. The author benefited greatly from the assistance of Bruce Cornuelle, Matthew Dzieciuch, Walter Munk and Peter Worcester. Members of the NPAL Group provided valuable comments and environmental data. This work was supported by the University of Southern Mississippi, by the Strategic Environmental Research and Development Program through Defense Advanced Research Projects Agency (DARPA) Grant No. MDA972-93-1-0003, and by the Office of Naval Research (ONR) through Grant Nos. N0014-1-97-1-0258 and N00014-03-1-0182.

[1]T. Duda, S. M. Flatté, J. Colosi, B. Cornuelle J. Hildebrand, W. Hodgkiss,Jr., P. Worcester, B. Howe, J. Mercer, and R. Spindel, "Measured wavefront fluctuations in 1000-km pulse propagation in the Pacific Ocean," J. Acoust. Soc. Am. **92**, 939–955 (1992).

[2]P. F. Worcester, B. D. Cornuelle, M. A. Dzieciuch, W. H. Munk, B. M. Howe, J. A. Mercer, R. C. Spindel, J. A. Colosi, K. Metzger, T. G. Birdsall, and A. B. Baggeroer, "A test of basin-scale acoustic thermometry using a large-aperture vertical array at 3250-km range in the eastern North Pacific Ocean," J. Acoust. Soc. Am. **105**, 3185–3201 (1999).

[3]P. Worcester and R. Spindel, "North Pacific Acoustic Laboratory," J. Acoust. Soc. Am. **117**, 1499–1510 (2005).

[4]J. Colosi, S. M. Flatté, and C. Bracher, "Internal-wave effects on 1000-km oceanic acoustic pulse propagation: Simulation and comparison with experiment," J. Acoust. Soc. Am. **96**, 452–468 (1994).

[5]J. A. Colosi, E. K. Scheer, S. M. Flatté, B. D. Cornuelle, M. A. Dzieciuch, W. H. Munk, P. F. Worcester, B. M. Howe, J. A. Mercer, R. C. Spindel, K. Metzger, T. G. Birdsall, and A. B. Baggeroer, "Comparisons of measured and predicted acoustic fluctuations for a 3250-km propagation experiment in the eastern North Pacific Ocean," J. Acoust. Soc. Am. **105**, 3202–3218 (1999).

[6]J. Colosi, F. Tappert, and M. Dzieciuch, "Further analysis of intensity fluctuations from a 3252-km acoustic propagation experiment in the eastern North Pacific," J. Acoust. Soc. Am. **110**, 163–169 (2001).

[7]C. Garrett and W. Munk, "Space-time scales of ocean internal waves," Geophys. Fluid Dyn. **2**, 225–264 (1972).

[8]C. Garrett and W. Munk, "Space-time scales of internal waves: A progress report," J. Geophys. Res. **80**, 291–297 (1975).

[9]S. Flatté and M. Vera, "Comparison between ocean-acoustic fluctuations in parabolic-equation simulations and estimates from integral approximations," J. Acoust. Soc. Am. **114**, 697–706 (2003).

[10]R. Andrew, B. Howe, J. Mercer, and the NPAL Group, "Transverse horizontal spatial coherence of deep arrivals at megameter ranges," J. Acoust. Soc. Am. **117**, 1511–1526 (2005).

[11]A. Voronovich, V. Ostashev, and the NPAL Group, "Horizontal refraction of acoustic signals retrieved from North Pacific Acoustic Laboratory billboard array data," J. Acoust. Soc. Am. **117**, 1527–1537 (2005).

[12]R. Hardin and F. Tappert, "Applications of the split-step Fourier method to the numerical solution of nonlinear and variable coefficient wave equations," SIAM Rev. **15**, 423 (1973).

[13]F. Tappert, "The parabolic approximation method," in *Wave Propagation and Underwater Acoustics*, edited by J. Keller and J. Papadakis (Springer,

Berlin, 1977), pp. 224–287.
[14]S. Flatté and F. Tappert, "Calculation of the effect of internal waves on oceanic sound transmission," J. Acoust. Soc. Am. **58**, 1151–1159 (1975).
[15]M. Collins, "A split-step Padé solution for the parabolic equation method," J. Acoust. Soc. Am. **93**, 1736–1742 (1993).
[16]S. Flatté, "The Schrödinger equation in classical physics," Am. J. Phys. **54**, 1088–1092 (1986).
[17]R. Dashen, S. Flatté, and S. Reynolds, "Path-integral treatment of acoustic mutual coherence functions for rays in a sound channel," J. Acoust. Soc. Am. **77**, 1716–1722 (1985).
[18]S. Flatté, R. Dashen, W. Munk, K. Watson, and F. Zachariasen, *Sound Transmission Through a Fluctuating Ocean, Mechanics and Applied Mathematics* (Cambridge University Press, Cambridge, 1979).
[19]S. Flatté and G. Rovner, "Calculations of internal-wave-induced fluctuations in ocean-acoustic propagation," J. Acoust. Soc. Am. **108**, 526–534 (2000).
[20]P. Worcester, B. Cornuelle, J. Hildebrand, W. Hodgkiss, T. Duda, J. Boyd, B. Howe, J. Mercer, and R. Spindel, "A comparison of measured and predicted broadband acoustic arrival patterns in travel time-depth coordinates at 1000 km range," J. Acoust. Soc. Am. **95**, 3118–3128 (1994).
[21]B. Howe, "Acoustic Thermometry of Ocean Climate (ATOC): Pioneer Seamount Source Installation," Technical Report No. APL TM3-96, Applied Physics Laboratory, University of Washington (1996).
[22]V. D. Grosso, "New equation for the speed of sound in natural waters (with comparisons to other equations)," J. Acoust. Soc. Am. **56**, 1084–1091 (1974).
[23]M. Vera, K. Heaney, and the NPAL Group, "The effect of bottom interaction on transmissions from the North Pacific Acoustic Laboratory Kauai source," J. Acoust. Soc. Am. **117**, 1624–1634 (2005).
[24]J. Colosi and M. Brown, "Efficient numerical simulation of stochastic internal-wave-induced sound-speed perturbation fields," J. Acoust. Soc. Am. **103**, 2232–2235 (1998).
[25]R. Esswein and S. Flatté, "Calculation of the phase-structure function density from oceanic internal waves," J. Acoust. Soc. Am. **70**, 1387–1396 (1981).
[26]F. Beron-Vera, M. Brown, J. Colosi, S. Tomsovic, A. Virovlyansky, M. Wolfson, and G. Zaslavsky, "Ray dynamics in a long-range acoustic propagation experiment," J. Acoust. Soc. Am. **114**, 1226–1242 (2003).
[27]S. Flatté and G. Rovner, "Path-integral expressions for fluctuations in acoustic transmission in the ocean waveguide," in *Methods of Theoretical Physics Applied to Oceanography*, edited by P. Müller, Proceedings of the Ninth 'Aha Huliko'a Hawaiian Winter Workshop, 1997, pp. 167–174.
[28]Z. Zhang and C. Tindle, "Improved equivalent fluid approximations for a low shear speed ocean bottom," J. Acoust. Soc. Am. **98**, 3391–3396 (1995).
[29]F. Jensen, W. Kuperman, M. Porter, and H. Schmidt, *Computational Ocean Acoustics* (American Institute of Physics, New York, 1994).
[30]M. Dzieciuch, P. Worcester, and W. Munk, "Turning point filters: Analysis of sound propagation on a gyre-scale," J. Acoust. Soc. Am. **110**, 135–149 (2001).

The response of rectangular piezoelectric sensors to Rayleigh and Lamb ultrasonic waves

Francesco Lanza di Scalea,[a] Howard Matt, and Ivan Bartoli
NDE & Structural Health Monitoring Laboratory, Department of Structural Engineering, University of California, San Diego, 9500 Gilman Drive, M.C. 0085, La Jolla, California 92093-0085

(Received 16 May 2006; revised 25 October 2006; accepted 25 October 2006)

A fundamental understanding of the response of piezoelectric transducer patches to ultrasonic waves is of increasing interest to the field of structural health monitoring. While analytical solutions exist on the interaction of a piezoelectric actuator with the generated Lamb waves, the behavior of a piezoelectric sensor has only been examined for the limited case of a piezo-actuated Lamb wave in a pitch-catch configuration. This paper focuses on the fundamental response of surface-bonded piezoelectric sensors to ultrasonic waves. The response to both Rayleigh waves and Lamb waves is examined, starting with harmonic excitation fields and moving to broadband and narrowband excitation fields. General oblique incidence of the wave on rectangular sensors is treated first; parallel incidence is then derived as a particular case. The solutions are developed analytically for the harmonic and the narrowband excitations, and semianalytically for the broadband excitation. The results obtained can be used to design ultrasonic sensors that are either particularly sensitive to a given mode or possess unique directivity behavior which, in turn, can greatly simplify current algorithms for damage detection and localization. © *2007 Acoustical Society of America.*
[DOI: 10.1121/1.2400668]

PACS number(s): 43.35.Zc, 43.35.Yb, 43.40.Le, 43.38.Fx [TDM] Pages: 175–187

I. INTRODUCTION

With the advent of on-board structural health monitoring systems based on ultrasonic waves,[1–4] piezoelectric transducer patches are being increasingly used due to their low-cost, low-profile, and ease of integration within the structure.[5–18]

The fundamental electromechanical behavior of piezoelectric transducers is well understood.[18–21] The transducer-structure interaction has been previously addressed for applications involving Lamb waves. One of the earliest works was conducted by Monkhouse *et al.*[1] on piezoelectric polymer films for Lamb wave transduction. The authors extended to the Lamb wave case the concept of interdigitated electrodes previously exploited for generating narrowband surface acoustic waves. Moulin *et al.*[22] utilized a hybrid method (finite element and normal mode expansion) to solve for the Lamb wave response resulting from surface mounted and embedded piezoelectric actuators in CFRP laminates. Lin and Yuan[23] analytically derived the voltage response of a piezoceramic sensor attached to a plate and subjected to narrowband (tone burst) excitation originating from a piezoceramic actuator. In this study Mindlin plate theory was used to examine the fundamental antisymmetric A_0 mode below the cut-off frequencies of higher-order modes. Giurgiutiu[18] analytically derived and experimentally validated the Lamb wave tuning capability of attached piezoceramic actuators for both the fundamental symmetric, S_0, and antisymmetric, A_0, mode. He found that the preferential excitement of specific mode-frequency combinations occurs according to the relation between the Lamb wavelength (or wave number) and the dimension of the actuator (wavelength tuning). Giurgiutiu's closed form solution for the actuator-generated waves assumes harmonic excitation and was obtained by coupling a pin-force shear stress at the plate's surface (two concentrated in-plane forces in opposite directions at either end of the transducer) with the theoretical Rayleigh-Lamb displacement and strain fields.

Building on the above-mentioned work, Raghavan and Cesnik[24] used three-dimensional linear elasticity to obtain the general solution for S_0 and A_0 circularly crested waves excited by a circular piezoceramic actuator. The pin-force shear stress excitation was, again, assumed to predict displacement and strain fields generated in the plate. The solution was derived for both harmonic excitation and narrowband (tone burst) excitation. This study also examined the voltage response of piezoceramic *sensors* subjected to harmonic Lamb wave fields generated by either rectangular piezo-actuators (plane wave fields) or circular piezo-actuators (circularly crested wave fields). One important conclusion was that wavelength tuning conditions do not necessarily govern the response of piezo-sensors. Rather, the general trend is a monotonic increase in the output voltage with decreasing sensor dimension as a result of averaging the wave strain field over the sensor gauge length.

It should be noted that the difference in response between a piezo-actuator and a piezo-sensor, indeed, does not contradict elastodynamic reciprocity conditions. The pin-force applied by a finite-size actuator will generally excite multiple modes of the structure. Based on reciprocity, the response of the same finite-size sensor to an excitation consisting of the same, multiple modes would be equivalent to the response of the actuator. This equivalence has recently

[a]Author to whom correspondence should be addressed; electronic mail: flanza@ucsd.edu

been used in baseline-free damage detection based on time-reversal acoustics using pitch-catch actuator-sensor pairs. However, the response of a sensor to a single incoming mode cannot be immediately recovered from the pin-force actuation solution on the basis of reciprocity.

Greve *et al.*[25] and Nieuwenhuis *et al.*[26] compared Giurgiutiu's theoretical models[18] for S_0 and A_0 harmonic excitation by a piezo-actuator to finite element simulations of narrowband (tone burst) excitation. Their model showed slight shifts in the frequencies at which wavelength tuning occurs compared to the theoretical harmonic case. This work also examined the response of piezo-sensors to the piezo-actuated fields under narrowband excitation. It was confirmed that the response of the piezo-sensor to Lamb waves cannot be predicted by the pin-force piezo-actuator model.

The majority of the above-reviewed works has covered the response of piezoelectric sensors to piezo-generated wave fields. It is of interest to examine the fundamental response of a piezo-sensor to an incoming wave field. This knowledge is relevant, for example, to passive-only structural monitoring (e.g., acoustic emission testing) and, also, to active monitoring when the piezo-actuated wave changes dispersive properties prior to reaching the piezo-sensor. The latter case, for example, can occur through mode conversions caused by scattering at small discontinuities or simply by changes in the waveguide thickness.

The present paper examines the fundamental response of surface-mounted, rectangular piezoelectric sensors subjected to harmonic, broadband, and narrowband wave fields under plane wave hypotheses. The study first treats the case of Rayleigh surface waves and subsequently that of Lamb waves. The sensor voltage response is derived by coupling the direct piezoelectric effect to the wave strain field that is averaged over the sensor area. Analytical expressions are obtained for the general case of waves propagating at oblique incidence relative to the rectangular sensor. The predictions are compared to experimental data for the broadband excitation cases. The results can help in the design of sensors which respond favorably to a given mode-frequency combination while suppressing other mode-frequency combinations. In addition, the directivity in the response of the rectangular sensors can be exploited to triangulate the location of an ultrasound source without prior knowledge of the wave velocity in the medium.

II. STRAIN SENSITIVITY OF PIEZOELECTRIC SENSOR PATCHES

The piezoelectric constitutive equations can be written as[27]

$$\mathbf{D} = \mathbf{e}^{\sigma}\mathbf{E} + \mathbf{d}\boldsymbol{\sigma} \quad \text{(direct effect)}, \tag{1a}$$

$$\boldsymbol{\sigma} = -(\mathbf{d}\mathbf{C}^{E})^{T}\mathbf{E} + \mathbf{C}^{E}\boldsymbol{\varepsilon} \quad \text{(converse effect)}, \tag{1b}$$

where in Eq. (1a), $\mathbf{D}$ is the charge density vector (3×1), $\mathbf{e}^{\sigma}$ is the dielectric permittivity matrix (3×3) measured at zero stress, $\mathbf{E}$ is the electric field vector (3×1), $\mathbf{d}$ is the piezoelectric coefficient matrix (3×6), and $\boldsymbol{\sigma}$ is the stress vector (6×1). In Eq. (1b), $\mathbf{C}^{E}$ is the elastic stiffness matrix (6×6) measured at zero electric field, $\boldsymbol{\varepsilon}$ is the strain vector (6×1), and the superscript T indicates transposed matrix.

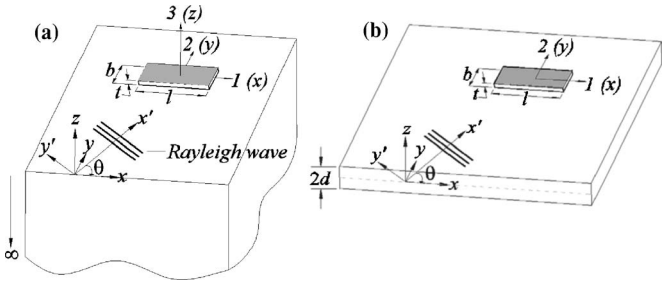


FIG. 1. A rectangular piezoelectric sensor and its interaction with (a) Rayleigh waves and (b) Lamb waves in general oblique incidence.

Consider a thin, rectangular piezoelectric sensor with length l, width b, and thickness t along directions 1, 2, and 3, respectively, Fig. 1(a). The piezoelectric element can be considered as transversely isotropic where the plane of isotropy corresponds to the (1, 2) plane. The electrical poling is assumed to be along the transducer's thickness direction, 3. Under plane stress conditions ($\sigma_{33}=\sigma_{23}=\sigma_{13}=0$) and after substitution of Eq. (1b) into Eq. (1a), the direct piezoelectric effect reduces to[28]

$$D_3 = [d_{31}\ d_{32}\ 0]\begin{bmatrix} Q_{11} & Q_{12} & 0 \\ Q_{12} & Q_{11} & 0 \\ 0 & 0 & Q_{66} \end{bmatrix}\left(\begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{22} \\ \gamma_{12} \end{bmatrix} - E_3\begin{bmatrix} d_{31} \\ d_{32} \\ 0 \end{bmatrix}\right) + e_{33}^{\sigma}E_3, \tag{2}$$

where d_{31} and d_{32} are the piezoelectric constants with units (m/V), and ε_{11}, ε_{22}, γ_{12} are the strain components in the sensor. In the case of monolithic piezoelectric elements ($d_{31}=d_{32}$), the reduced stiffness matrix, $[Q]$, can be written as

$$[Q] = \begin{bmatrix} \dfrac{Y^E}{1-\nu^2} & \dfrac{\nu Y^E}{1-\nu^2} & 0 \\ \dfrac{\nu Y^E}{1-\nu^2} & \dfrac{Y^E}{1-\nu^2} & 0 \\ 0 & 0 & \dfrac{Y^E}{2(1+\nu)} \end{bmatrix}, \tag{3}$$

where Y^E and ν are the Young's modulus and Poisson's ratio of the piezoelectric transducer in the (1, 2) plane of isotropy. Simplifying Eq. (2), the expression describing the charge density becomes

$$D_3 = d_{31}(Q_{11}+Q_{12})(\varepsilon_{11}+\varepsilon_{22}) - [2d_{31}^2(Q_{11}+Q_{12}) - e_{33}^{\sigma}]E_3. \tag{4}$$

Considering the electrical boundary conditions of the piezoelectric patch to be an open circuit, the total charge over the electrode area is zero, thus $\int\int D_3 dxdy = 0$.[29] Under such conditions, the transducer voltage can be calculated as

$$V = -\frac{\int\int\int E_3 dxdydz}{lb}. \tag{5}$$

Substituting Eq. (4) into Eq. (5):

$$V = \frac{d_{31}(Q_{11}+Q_{12})t \int\int (\varepsilon_{11}+\varepsilon_{22})dxdy}{lb[e_{33}^{\sigma} - 2d_{31}^2(Q_{11}+Q_{12})]}, \quad (6)$$

where the in-plane normal strains are assumed constant through the transducer's thickness.

Substituting the reduced stiffness components from Eq. (3) into Eq. (6), the following expression is finally obtained for the voltage generated by the piezoelectric sensor patch as a function of the in-plane strains:

$$V = \frac{d_{31}Y^E t \int\int (\varepsilon_{11}+\varepsilon_{22})dxdy}{lb[e_{33}^{\sigma}(1-\nu) - 2d_{31}^2 Y^E]}. \quad (7)$$

The sensor response to Rayleigh and Lamb ultrasonic waves will be examined in the following sections by considering the appropriate strain fields in Eq. (7). It is presumed in this analysis that the presence of the sensor does not significantly alter the strain field of the incident wave.

III. RESPONSE TO RAYLEIGH WAVES

A. Harmonic Rayleigh excitation

Consider a rectangular piezoelectric sensor bonded to the surface of an isotropic elastic medium and subjected to a harmonic surface (Rayleigh) plane wave. In general, the wave propagates along a direction x' inclined at an angle θ from the sensor's lengthwise direction x, Fig. 1(a). In order to calculate the sensor voltage response from Eq. (7), the surface strain components along the wave propagation axes (x',y') must be rotated to the sensor geometrical axes (x,y). For plane waves $(\varepsilon_{y'y'}=0)$, the following strain invariant can be written:

$$\varepsilon_{xx}+\varepsilon_{yy} = \varepsilon_{x'x'}+\varepsilon_{y'y'} = \varepsilon_{x'x'}, \quad (8)$$

where the Rayleigh surface strain $\varepsilon_{x'x'}$ can be written as[30]

$$\varepsilon_{x'x'}|_{z=0} = -\frac{ik^2 B}{2q}(r-2sq)e^{i(kx'-\omega t)}. \quad (9)$$

In Eq. (9), k is the wave number, B is the arbitrary amplitude of the wave, and $r=2-(c/c_T)^2, q=\sqrt{1-(c/c_L)^2}, s=\sqrt{1-(c/c_T)^2}$. Ignoring shear lag, and thus assuming a very thin and stiff adhesive layer,[31] the sensor experiences the same strain as the structure, thus $(\varepsilon_{11}+\varepsilon_{22})=(\varepsilon_{xx}+\varepsilon_{yy})|_{z=0}$. Substituting Eq. (9) into Eq. (7) yields the following voltage response:

$$V = -\frac{d_{31}Y^E tik^2 B(r-2sq)\int\int e^{i(kx'-\omega t)}dxdy}{2qlb[e_{33}^{\sigma}(1-\nu) - 2d_{31}^2 Y^E]}. \quad (10)$$

Using the variable transformation $x'=x\cos\theta+y\sin\theta$, evaluating the integral in y over the sensor width $[-b/2, b/2]$, and considering the integration limits of x over the sensor length $[-l/2, l/2]$, the previous expression can be written as

$$V = -\frac{SkBR\int_{-l/2}^{l/2} e^{i(kx\cos\theta-\omega t)}\left(\exp\left(i\frac{kb\sin\theta}{2}\right)-\exp\left(-i\frac{kb\sin\theta}{2}\right)\right)dx}{lb\sin\theta}, \quad (11)$$

where S and R are frequency-independent constants defined as

$$S = \frac{d_{31}Y^E t}{e_{33}^{\sigma}(1-\nu) - 2d_{31}^2 Y^E},$$

$$R = \frac{(r-2sq)}{2q}. \quad (12)$$

Using Euler's and prosthaphaeresis formulas in Eq. (11), the sensor voltage response to a harmonic Rayleigh wave can be simplified to

$$V = -i\bar{V}e^{-i\omega t}, \quad (13)$$

where the amplitude is

$$\bar{V} = \frac{SA_{\text{Rayleigh}}}{lb\sin\theta\cos\theta}\sin\left(\frac{kb\sin\theta}{2}\right)\sin\left(\frac{kl\cos\theta}{2}\right)$$

(oblique incidence). (14)

The term $A_{\text{Rayleigh}}=4BR$ is related to the Rayleigh wave amplitude.

When the wave propagation direction, x', is parallel to the sensor's lengthwise direction, x, the incidence angle, θ, is equal to zero. In this case, since $\lim_{\alpha\to 0}(\sin\alpha)/\alpha=1$, Eq. (14) simplifies to the following result, expressed in terms of either the wave number, k, or the wavelength, λ:

$$\bar{V} = \bar{V}|_{\theta\to 0} = \frac{SA_{\text{Rayleigh}}k}{2l}\sin\left(\frac{kl}{2}\right)$$

$$= \frac{SA_{\text{Rayleigh}}\pi}{\lambda l}\sin\left(\frac{\pi l}{\lambda}\right) \quad \text{(parallel incidence)}. \quad (15)$$

Eqution (15) shows that the response to a given wave number (or wavelength) will be influenced by the dimension of

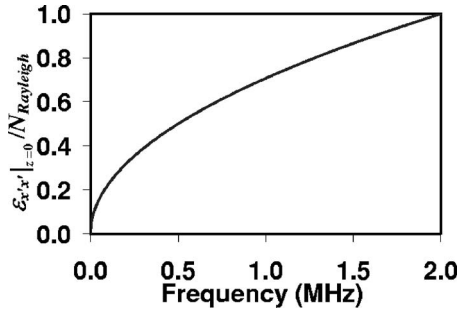


FIG. 2. Surface strain of a Rayleigh wave in steel normalized by the wave power flow through the solid at various frequencies.

the sensor parallel to the wave propagation direction (wavelength tuning). From the sinusoidal term, the response will be large at $\lambda=2l/(2n-1)$ for $n=1,2,3,\ldots$ corresponding to sensor lengths equal to an odd multiple of half the wavelength. Contrary to what was concluded for piezoelectric actuators,[18] these conditions do not necessarily correspond to local maxima of the response due to the presence of the factor $(SA_{\text{Rayleigh}}\pi/\lambda l)$. The response will be zero at $\lambda=l/n$ for $n=1,2,3,\ldots$ corresponding to sensor lengths equal to an integer multiple of the wavelength. Furthermore, from the term l in the denominator of the multiplication factor, the response will increase with decreasing sensor length.

Through the factor A_{Rayleigh}, Eqs. (14) and (15) contain the arbitrary amplitude of the wave, B. Therefore, a normalization is required in order to compare the sensor response to different wavelengths/frequencies of the incoming field. One approach is normalizing the response by the power flow carried by the wave. The following normalization factor was considered for the Rayleigh response:

$$N_{\text{Rayleigh}} = \sqrt{\int_{-\infty}^{0} P_{x'}\,dz}, \tag{16}$$

where $P_{x'}$ is the x' component of the wave Poynting vector which, in turn, is calculated from the relevant stress components and particle velocities as

$$P_{x'} = -\frac{1}{2}\text{Re}\left[\sigma_{x'x'}\left(\frac{\partial u_{x'}}{\partial t}\right)^{*} + \sigma_{x'z}\left(\frac{\partial u_{z}}{\partial t}\right)^{*}\right], \tag{17}$$

where Re indicates real part, and the asterisk indicates complex conjugate.

For Rayleigh waves propagating in steel ($c_L=5.9$ mm/μs, $c_T=3.2$ mm/μs, $c=2.99$ mm/μ), the normalized surface strain ($\varepsilon_{x'x'}|_{z=0}/N_{\text{Rayleigh}}$) is shown in Fig. 2 as a function of frequency. The resulting nonlinear trend indicates that, with increasing frequency, the square root of the power flow increases faster than the surface strain.

The normalized voltage response to harmonic Rayleigh waves propagating along the lengthwise direction of the sensor, $\bar{V}/N_{\text{Rayleigh}}$ calculated from Eqs. (15) and (16), is plotted in Fig. 3(a) for the following sensor dimensions: $l=12.7$ mm (0.5 in.), $l=6.35$ mm (0.25 in.), and $l=3.17$ mm (0.125 in.). The limit case of $l=0$ (point sensor) is also shown. As discussed earlier, the response is large at frequencies where the sensor length is equal to an odd multiple of

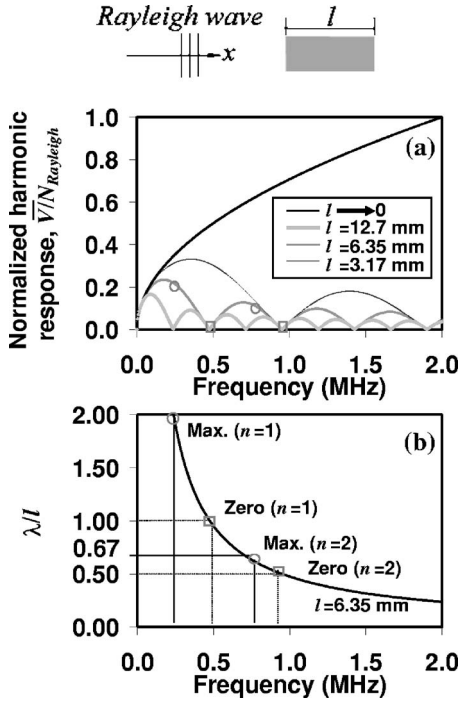


FIG. 3. (a) Normalized voltage response of sensor subjected to parallel incident harmonic Rayleigh wave for varying sensor lengths; and (b) ratio of Rayleigh wavelength to sensor length of 6.35 mm.

half the wavelength; the response is, instead, zero at frequencies where the sensor length is equal to an integer multiple of the wavelength. This result is further emphasized in Fig. 3(b) where the nondimensional ratio of the Rayleigh wavelength to sensor length is plotted for the 6.35-mm-long sensor. Notice that the first two local maxima of the response at $f=0.25$ and 0.75 MHz, circled in Fig. 3, are slightly shifted from the conditions $\lambda=2l/(2n-1)$. The shift is due to the factor $(SA_{\text{Rayleigh}}\pi/\lambda l)$ in Eq. (15) and to the frequency dependence of N_{Rayleigh}. Figure 3(a) shows that the magnitude of the local response maxima for a given sensor length decreases with increasing frequency. Figure 3 also confirms that the response to a given wave frequency increases with decreasing sensor length, as a result of the term l in the denominator of Eq. (15).

The limit case of the point sensor ($l=0$) results in a larger response than any of the finite-length sensors. The point-sensor response also identically matches the normalized surface strain shown in Fig. 2, confirming the "ideal strain sensor" behavior where smearing effects due to strain averaging over the sensor length are not present.

B. Arbitrary excitation

The harmonic Rayleigh wave excitation assumed in Eqs. (14) and (15) is of theoretical interest. However, harmonic excitation will seldom occur in a practical structural health monitoring application where the wave fields will likely be either broadband (e.g., an acoustic emission detection) or narrowband (e.g., a pitch-catch testing configuration).

The amplitude $\bar{U}=\bar{U}(\omega)$ of the frequency-domain response of a linear system subjected to an arbitrary excitation can be calculated from the product of the frequency-domain

response of the system to harmonic excitation, $\bar{V}=\bar{V}(\omega)$, and the amplitude of the Fourier transform of the excitation, $\bar{P}=\bar{P}(\omega)$:[32]

$$\bar{U}=\bar{V}\times\bar{P}. \tag{18}$$

In turn, the sensor excitation can be expressed in terms of the amplitude spectrum of the surface strain induced by the arbitrary source at the point of entry into the sensor, $\varepsilon^a_{x'x'}|_{z=0}(\omega)$, in the following manner:

$$\bar{P}=\frac{\varepsilon^a_{x'x'}|_{z=0}}{\varepsilon_{x'x'}|_{z=0}}, \tag{19}$$

where $\varepsilon_{x'x'}|_{z=0}(\omega)$ is the amplitude spectrum of the surface strain excited under harmonic conditions, and the dependence of all terms on ω is implicit. The sensor response to arbitrary excitation can be thus calculated as

$$\bar{U}=\bar{V}\times\frac{\varepsilon^a_{x'x'}|_{z=0}}{\varepsilon_{x'x'}|_{z=0}}. \tag{20}$$

The following sections address the response to broadband and narrowband Rayleigh wave fields, respectively.

C. Broadband Rayleigh excitation

1. Semianalytical analysis

The harmonic surface strain field at a point is equivalent to the sensor response to harmonic excitation from Eq. (14) when the sensor length l and width b are taken to be 0 (point sensor). In order to transform voltage into strain, the pointwise response must be further divided by the term S, which contains the piezoelectric constant of the sensor. Thus the arbitrary excitation spectrum can be written as

$$\bar{P}=\frac{\varepsilon^a_{x'x'}|_{z=0}}{\varepsilon_{x'x'}|_{z=0}}=\frac{\varepsilon^a_{x'x'}|_{z=0}}{\lim\limits_{l,b\to 0}\bar{V}/S}=\varepsilon^a_{x'x'}|_{z=0}\frac{\lambda^2}{A_{\text{Rayleigh}}\pi^2}. \tag{21}$$

Substituting Eqs. (14) and (21) into Eq. (18), the response of the sensor to broadband Rayleigh wave excitation is

$$\bar{U}=\frac{S\lambda^2\,\varepsilon^a_{x'x'}|_{z=0}}{\pi^2 lb\sin\theta\cos\theta}\sin\left(\frac{\pi b\sin\theta}{\lambda}\right)\sin\left(\frac{\pi l\cos\theta}{\lambda}\right). \tag{22}$$

This expression shows that the response to broadband waves follows a similar wavelength tuning behavior as the response to harmonic waves, with appropriate scaling factors. For example, the response to parallel incidence will be large for sensor dimensions equal to an odd multiple of half the wavelength, and it will be zero for sensor dimensions equal to an integer multiple of the wavelength.

In this study, the surface strain spectrum $\varepsilon^a_{x'x'}|_{z=0}(\omega)$ in Eq. (22) was obtained through a conventional, two-dimensional finite element analysis (FEA) of a broadband Rayleigh wave generation. The total sensor response $\bar{U}$ was thus predicted by a semianalytical analysis.

The FEA used ABAQUS EXPLICIT[33] to discretize a steel block (c_L=5.9 mm/μs, c_T=3.2 mm/μs, density ρ=7700 kg/m^3), 250 mm (9.8 in.) in length and 50 mm (1.98 in.) in thickness. The block was assumed infinitely wide to reflect plain strain conditions. The excitation was a half-sinusoid forcing function (0.24 μs in duration), with a nearly uniform energy spectrum in the DC-2 MHz frequency range of interest, applied to a surface node and directed along the thickness direction to simulate an ablative generation regime. Material damping effects were neglected. The block's material, dimensions, and excitation-detection distance were equivalent to what was used in the experimental tests to be discussed in the next section. The discretization used four-node, bilinear plane strain quadrilateral elements with two degrees of freedom per node. Element sizes are typically conditioned by the minimum wavelength, $\lambda_{\min}$, as determined by the highest frequency of interest.[34,35] The largest element dimension was chosen equal to $\lambda_{\min}/10$. The integration time step was set to Δt=0.01 μs.

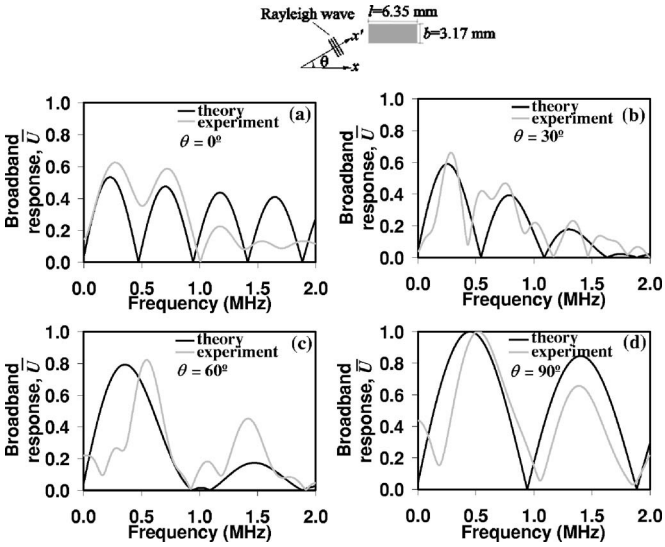


FIG. 4. Experimental and theoretical frequency response of 6.35 mm $\times$3.17 mm rectangular piezoelectric sensor to broadband Rayleigh waves propagating at (a) 0°, (b) 30°, (c) 60°, and (d) 90° from the lengthwise sensor dimension.

2. Experimental analysis

Experimental tests were conducted on a 250 mm $\times$250 mm$\times$50 mm (9.8 in.$\times$9.8 in.$\times$1.98 in.) steel block upon which a rectangular Pb(Zr-Ti)O_3-PZT patch of dimensions 6.35 mm$\times$3.17 mm$\times$0.5 mm (0.25 in.$\times$0.125 in. $\times$0.02 in.) was bonded using a thin layer of instant adhesive. A 12 ns pulse from an Nd:YAG Q-switched laser was focused to a line on the block's surface to generate broadband Rayleigh waves in the ablative regime. The block's orientation was varied such that the waves propagated at various angles relative to the sensor's geometrical axes. Signal acquisition was performed by a National Instruments PXI-1010 unit running under LABVIEW software which recorded the first Rayleigh wave arrivals and subsequently performed the fast-Fourier transform in the DC-2 MHz range.

The results for wave propagation directions inclined at θ=0°, 30°, 60°, and 90° from the lengthwise direction of the sensor are shown in Fig. 4. Each response spectrum was

normalized by the peak response of a commercial acoustic emission transducer bonded next to the PZT so as to eliminate the effects of the laser shot-to-shot variations. The corresponding responses calculated theoretically from Eq. (22) are also shown for comparison. The amplitude of the response increases with decreasing sensor dimension along the wave propagation direction, and it is thus smallest for the 0° incidence (6.35 mm length) in Fig. 4(a), and largest for the 90° incidence (3.17 mm length) in Fig. 4(d). There is a reasonably good agreement between experiment and theory regarding the position of the local maxima and minima of the response according to wavelength tuning. The relative amplitude of the response also compares relatively well between theory and experiment. The strongest discrepancies are seen above 1 MHz, where the theory overestimates the experimental response for the 0° and the 90° incidence, while it underestimates the experimental response for the 60° incidence. These differences could result from either wave attenuation losses or shear lag effects, both of which were neglected in the theory. Another potential source of discrepancy is the finite thickness of the transducer which may modify the incoming strain field due to local changes in mass and stiffness of the waveguide.

D. Narrowband Rayleigh excitation

The specific case of narrowband excitation examined is that of a tone burst wave field that may be excited by an actuator in a pitch-catch testing configuration. Only parallel incidence ($\theta=0°$) is considered in this analysis. The extension to oblique incidence is a relatively straightforward matter. An unmodulated tone burst in the time domain, $f(t)$, can be expressed as

$$f(t)=\begin{cases} D\sin\omega_0 t & \text{for}|t|<L \\ 0 & \text{for}|t|>L, \end{cases} \tag{23}$$

where ω_0 is the central frequency and $2L$ is the time duration of the wave train.

The Fourier transform amplitude of the fixed time window toneburst of central frequency ω_0 can be expressed as[36]

$$\bar{F}=D\left[\frac{\sin n\pi\left(\frac{\omega}{\omega_0}+1\right)}{(\omega+\omega_0)}-\frac{\sin n\pi\left(\frac{\omega}{\omega_0}-1\right)}{(\omega-\omega_0)}\right], \tag{24}$$

where n is number of cycles within the time window. Assuming that this quantity corresponds directly to the surface strain $\varepsilon^a_{x'x'}|_{z=0}$, the amplitude spectrum of the sensor excitation from Eq. (19) becomes

$$\bar{P}=\frac{\varepsilon^a_{x'x'}|_{z=0}}{\varepsilon_{x'x'}|_{z=0}}=\frac{\bar{F}}{\lim\limits_{l,b\to 0}\bar{V}/S}=\bar{F}\frac{\lambda^2}{A_{\text{Rayleigh}}\pi^2}. \tag{25}$$

Substituting Eqs. (15) and (25) into Eq. (18), the response amplitude of the piezoelectric sensor to a tone burst Rayleigh wave excitation of amplitude D and center frequency ω_0 is

$$\bar{U}=D\left[\frac{\sin n\pi\left(\frac{2\pi c}{\lambda\omega_0}+1\right)}{\left(\frac{2\pi c}{\lambda}+\omega_0\right)}-\frac{\sin n\pi\left(\frac{2\pi c}{\lambda\omega_0}-1\right)}{\left(\frac{2\pi c}{\lambda}-\omega_0\right)}\right]\frac{S\lambda}{\pi l}\times\sin\left(\frac{\pi l}{\lambda}\right), \tag{26}$$

where the wavelength λ has been used instead of the frequency ω in the square brackets for consistency with the other terms in the expression.

It is clear from this equation that the response will be dependent on both the tone burst center frequency, ω_0, and the harmonic wavelength tuning term, $\sin(\pi l/\lambda)$. Results from Eq. (26) are shown in Figs. 5(a)–5(c) for a sensor length $l=6.35$ mm, and assuming steel ($c=2.99$ mm/μs) as the test material. Three different center frequencies for the excitation tone burst were chosen for these plots. Two of them coincided with a large value ($f_{\max}$) and the first zero value ($f_{\min}$) of the harmonic sensor response (where $f_{\max}$ corresponds to $\lambda=2l$, and $f_{\min}$ corresponds to $\lambda=l$ from wavelength tuning). The third center frequency was chosen to coincide with an intermediate value (f_{int}) of the harmonic sensor response. From Fig. 3, these frequency values were: $f_{\max}=0.24$ MHz, $f_{\text{int}}=0.35$ MHz and $f_{\min}=0.47$ MHz for $l=6.35$ mm. The expected conclusion from Fig. 5 is that the response increases when the excitation tone burst frequency approaches the favorable sensor length-to-wavelength ratio. Notice also that a non-negligible response is obtained in Fig. 5(c) for tone burst frequencies corresponding to zero harmonic response as a result of the finite bandwidth and the sidelobes of the tone burst spectrum. The same results are plotted for a smaller sensor ($l=3.17$ mm) in Figs. 5(d)–5(f), considering the corresponding $f_{\max}=0.48$ MHz, $f_{\text{int}}=0.72$ MHz, and $f_{\min}=0.95$ MHz as tone burst center frequencies.

IV. RESPONSE TO LAMB WAVES

A. Harmonic Lamb excitation

Consider a rectangular piezoelectric sensor bonded to the upper surface of an isotropic plate of thickness $2d$ and subjected to a harmonic strain field associated to Lamb waves propagating in the plane (x',z) along direction x', Fig. 1(b). As for the Rayleigh wave case, the wave propagation direction x' forms an angle θ with the lengthwise direction of the sensor. The origin of the thickness coordinate, $z=0$, is at the midplane of the plate. For plane waves ($\varepsilon_{y'y'}=0$) the only strain relevant to the surface-mounted sensor is the in-plane component in the wave propagation direction given by[37]

$$\varepsilon_{x'x'}|_{z=d}=ik^2E\left(\tanh rd-\frac{2rs}{k^2+s^2}\cdot\tanh sd\right)\times\exp\left[i\left(kx'-\omega t-\frac{\pi}{2}\right)\right]$$

(antisymmetric modes) (27)

and

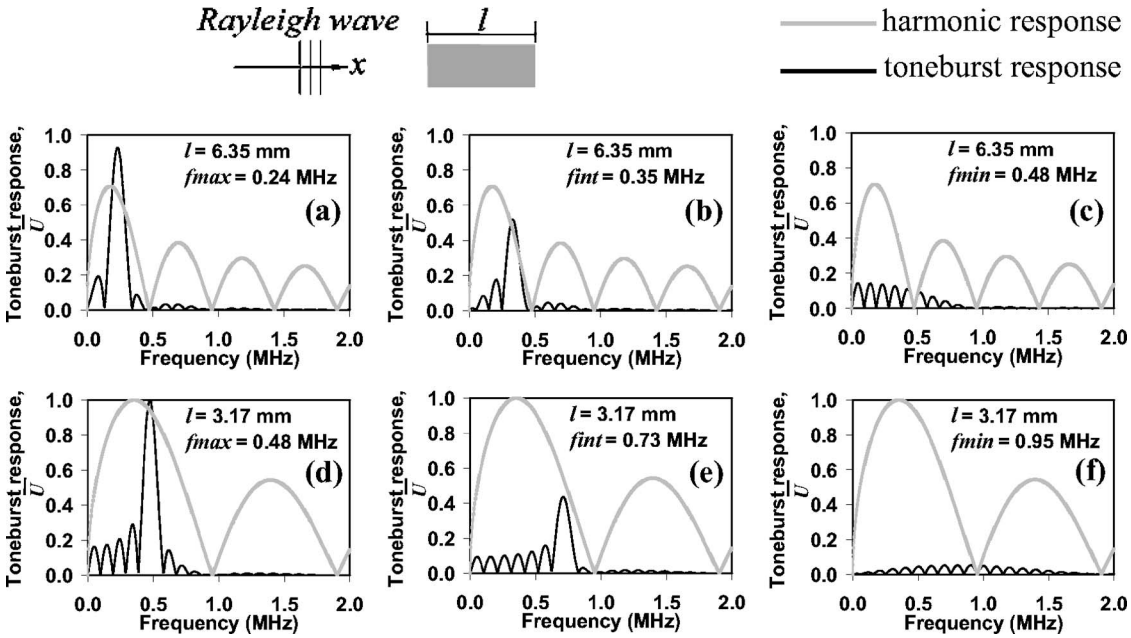


FIG. 5. Frequency response of piezoelectric sensor to a narrowband (tone burst) Rayleigh wave centered at frequencies corresponding to a large (a), an intermediate (b), and a zero (c) harmonic response for sensor length $l=6.35$ mm. (d)–(f) The same for sensor length $l=3.17$ mm. Harmonic responses are in gray lines.

$$\varepsilon_{x'x'}|_{z=d} = ik^2F\left(\coth rd - \frac{2rs}{k^2+s^2}\cdot\coth sd\right) \times\exp\left[i\left(kx' - \omega t - \frac{\pi}{2}\right)\right] \quad \text{(symmetric modes)}, \tag{28}$$

where E and F are the two arbitrary wave amplitudes for the antisymmetric and the symmetric modes, respectively, and $r=\sqrt{k^2-(\omega/c_L)^2}$, $s=\sqrt{k^2-(\omega/c_T)^2}$.

The voltage response to harmonic Lamb waves can be obtained by substituting Eq. (27) or (28) in Eq. (7). For the antisymmetric modes, this substitution results in

$$V = \frac{d_{31}Y^E tik^2E}{lb[e_{33}^{\sigma}(1-\nu)-2d_{31}^2Y^E]}\left(\tanh rd - \frac{2rs}{k^2+s^2}\cdot\tanh sd\right) \times\int\int \exp\left[i\left(kx' - \omega t - \frac{\pi}{2}\right)\right]dxdy. \tag{29}$$

Following the same steps of the Rayleigh wave case, Eq. (29) can be simplified to

$$V = i\bar{V}\exp\left[-i\left(\omega t + \frac{\pi}{2}\right)\right], \tag{30}$$

where the amplitude $\bar{V}$ of the sensor response to harmonic, antisymmetric, or symmetric Lamb waves is

$$\bar{V}_{\text{antisymm,symm}} = \frac{SA_{\text{antisymm,symm}}}{lb\sin\theta\cos\theta} \times\sin\left(\frac{kb\sin\theta}{2}\right)\sin\left(\frac{kl\cos\theta}{2}\right) \quad \text{(oblique incidence)}, \tag{31}$$

where the terms

$$A_{\text{antisymm}} = 4E\left(\tanh rd - \frac{2rs}{k^2+s^2}\tanh sd\right)$$

and

$$A_{\text{symm}} = 4F\left(\coth rd - \frac{2rs}{k^2+s^2}\coth sd\right)$$

are related to the Lamb wave fields. Comparing Eq. (31) with Eq. (14), it can be seen that the Lamb and Rayleigh wave responses are formally equivalent, with the difference being contained in the terms A_{Rayleigh}, A_{antisymm}, and A_{symm} which define the specific wave fields.

For parallel incidence ($\theta=0$), Eq. (31) simplifies to

$$\bar{V}_{\text{antisymm,symm}} = \bar{V}|_{\theta\to 0} = \frac{SA_{\text{antisymm,symm}}k}{2l}\sin\left(\frac{kl}{2}\right) = \frac{SA_{\text{antisymm,symm}}\pi}{\lambda l}\sin\left(\frac{\pi l}{\lambda}\right) \quad \text{(parallel incidence)}, \tag{32}$$

where the results are expressed in terms of either wave numbers or wavelengths.

The sinusoidal terms in Eq. (32) follow the same wavelength tuning behavior as for the Rayleigh wave case. The response will also generally increase with decreasing sensor length due to the presence of the term l in the denominator of the expressions.

In order to eliminate the arbitrary terms E and F implicitly present in Eqs. (31) and (32), the response can be normalized by the Lamb wave power flow

$$N_{\text{Lamb}} = \sqrt{\int_{-d}^{+d} P_{x'}dz}. \tag{33}$$

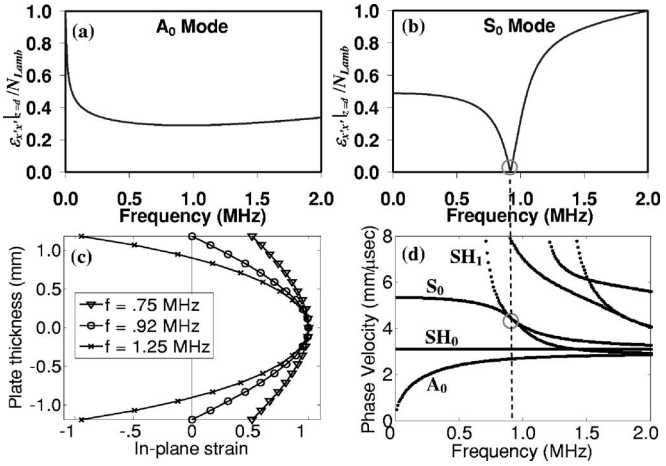


FIG. 6. In-plane surface strain of (a) A_0 Lamb wave and (b) S_0 Lamb wave in 2.38-mm-thick aluminum plate normalized by the wave power flow through the plate at various frequencies; (c) progression of cross-sectional mode shapes of in-plane strains for S_0 at various frequencies; and (d) phase velocity dispersion curves for the subject plate.

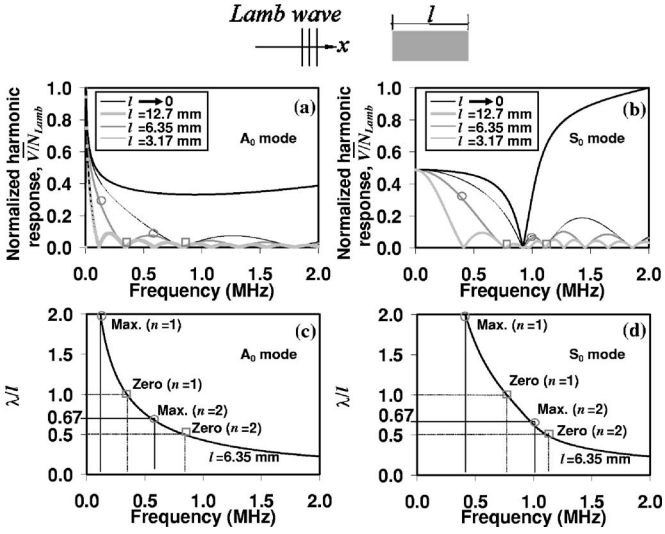


FIG. 7. Normalized voltage response of sensor subjected to parallel incident harmonic A_0 Lamb wave (a) and S_0 Lamb wave (b) for varying sensor lengths. Ratio of wavelength to sensor length of 6.35 mm for A_0 (c) and for S_0 (d).

Plots of the normalized strain at the top surface of the plate, $\varepsilon_{x'x'}|_{z=d}/N_{\text{Lamb}}$, are shown in Figs. 6(a) and 6(b) for the fundamental A_0 and S_0 modes, respectively. The plate is assumed 2.38 mm (0.09 in.) thick and made of aluminum (c_L=6.3 mm/μs, c_T=3.1 mm/μs, ρ=2700 kg/m^3).

In the case of A_0, Fig. 6(a), the normalized surface strain rapidly decreases from an asymptotically large value at extremely low frequencies to a relatively constant value over the remaining frequency range.

In the case of S_0, Fig. 6(b), two distinct trends exist: normalized strain values decrease with frequency for the first half of the frequency range, and increase with frequency for the second half of the frequency range. Separating these regions, a specific frequency (0.92 MHz) exists at which the normalized surface strain is zero. The 0.92 MHz value is the transition point between positive and negative strains at the plate's surface. The behavior can be seen in Fig. 6(c) showing the $\varepsilon_{x'x'}$ cross-sectional mode shapes of S_0 as the frequency changes from 0.75 to 0.92 MHz and then 1.25 MHz. In addition, this pivotal frequency occurs at the point of intersection between S_0 and the first-order shear horizontal mode, SH_1, as shown in the phase velocity dispersion curves of Fig. 6(d). It can be seen that the pivotal frequency is also a point of inflection of the S_0 dispersion curve. These conditions apply to isotropic plates of any thickness. In practice, the zero-strain pivotal frequency can be useful, for example, to exclusively detect (or excite) the A_0 mode by eliminating the influence of the S_0 mode.

The normalized voltage response $\bar{V}/N_{\text{Lamb}}$ calculated from Eq. (32) is plotted in Figs. 7(a) and 7(b) for harmonic A_0 and S_0 waves, respectively, propagating in the 2.38-mm-thick aluminum plate. The sensor dimensions in the wave propagation direction are the same as those considered for the Rayleigh response in Fig. 3: l=12.7, 6.35, and 3.17 mm, in addition to l=0 (point sensor). It is clear from Fig. 7 that the sensor response to Lamb waves is considerably different from that to Rayleigh waves.

For the A_0 mode, Fig. 7(a), the response asymptotically increases at extremely low frequencies. The response is generally smaller at higher frequencies, where it is primarily controlled by the wavelength tuning effect depending upon the ratio of Lamb wavelength to sensor dimension, λ/l. This ratio is plotted in Fig. 7(c) for A_0 and l=6.35 mm. As for the Rayleigh wave, the local maxima are slightly shifted from the $\lambda=2l/(2n-1)$ conditions due to the terms multiplying the sinusoidal function in Eq. (32). Notice that the first local maximum point ($\lambda=l$) is overshadowed by the large response at the low frequencies. Zero response is obtained at sensor lengths equal to an integer multiple of the wavelength ($\lambda=l/n$). The amplitude of the A_0 response generally decreases with increasing frequency (for a given sensor length) and, instead, increases with decreasing sensor length (for a given frequency). The largest response is always obtained from the ideal point sensor (l=0), which also matches identically the normalized, in-plane strain at the plate's surface in Fig. 6(a). Finally, the responses of the finite-size sensors converge to that of the point sensor at extremely low frequencies, where $\lambda \gg l$ and thus the exact strain solution at a point is retrieved.

The response to the S_0 mode is shown in Fig. 7(b). The asymptotic behavior at the low frequencies of the A_0 response is not seen in the S_0 response, which is consistent with the surface strain plot in Fig. 6(b). The wavelength tuning effect is also apparent in the S_0 response as confirmed by the λ/l plot in Fig. 7(d) for the 6.35-mm-long sensor. The zero response seen at f=0.92 MHz, however, is not determined by wavelength tuning; it is rather the result of the vanishing surface in-plane strain as discussed in Fig. 6. Consequently, the zero response at the pivotal frequency applies regardless of the dimensions of the sensor. This behavior will always occur for any isotropic plate at the point of inflection of the S_0 dispersion curve, as discussed earlier. The response to frequencies around the pivotal value is also affected by the phenomenon. For example, notice that above 0.92 MHz the response does not follow the usual trend of decreasing amplitude with increasing frequency (for a given sensor length).

As in the previous cases, the S_0 response to a given frequency increases with decreasing sensor length, with the point sensor ($l=0$) always yielding the largest response.

B. Arbitrary excitation

Equation (18) can be used to find the sensor response to an arbitrary, single mode Lamb wave excitation based on the harmonic response and the frequency content of the excitation. The same two cases examined for the Rayleigh waves, broadband and narrowband (tone burst) excitation, are considered for the Lamb waves in the following sections.

C. Broadband Lamb excitation

1. Semianalytical analysis

The term $\bar{P}$ in Eq. (18) should represent the amplitude spectrum of the Lamb wave entering the sensor, as generated by a broadband excitation of the plate. As for the Rayleigh wave case, the surface strain spectrum $\varepsilon^a_{x'x'}|_{z=d}$ in the expression for $\bar{P}$ of Eq. (19) was obtained through a FEA analysis of a plate subjected to broadband excitation. The harmonic response term $\bar{V}$ in Eq. (18) was, instead, derived analytically from Eq. (31). Thus the total sensor response $\bar{U}$ to broadband Lamb wave excitation was predicted semianalytically.

The plate studied by the FEA ABAQUS EXPLICIT simulations was aluminum ($c_L=6.3$ mm/μs, $c_T=3.1$ mm/μs, $\rho=2700$ kg/m^3), 2.38 mm (0.09 in.) in thickness and 400 mm (15.7 in.) in length. The discretization used four-node, bilinear plane strain quadrilateral elements with two degrees of freedom per node. Material damping effects were neglected. The spatial resolution criterion was $\lambda_{\min}/10$, and the integration time step was $\Delta t=0.01$ μs.

The simulated excitation used the half-sinusoid forcing function of 0.24 μs in duration. Two distinct analyses were performed in order to preferably excite the fundamental antisymmetric (A_0) or the symmetric (S_0) mode. For the A_0 excitation, an out-of-plane force was applied to the surface node in one of the plate's free ends. For the S_0 excitation, an in-plane force was applied to the mid-thickness node of the plate's free end. These conditions replicated the excitation conditions adopted in the experimental tests that follow. Consistent with the positions of the PZT sensors used in the experiment, the detection points were located at 40 mm (1.57 in.) from the antisymmetric excitation, and at 109 mm (4.3 in.) from the symmetric excitation. The calculated in-plane strains at the surface of the plate, $\varepsilon^a_{x'x'}|_{z=d}$, are shown in Figs. 8(a) and 8(b) for the antisymmetric and the symmetric excitation, respectively. The strain amplitude is normalized to the relative maximum in the plots.

One complication is that multiple Lamb modes are generated by practical plate excitations including the ones considered in this study. Single-mode excitation can only be achieved if pure cross-sectional mode shapes are induced. This is generally difficult to do, particularly over a wide frequency range such as the DC-2 MHz range considered here. Since the harmonic response of the sensor to a single mode is known, it is important to identify and separate its

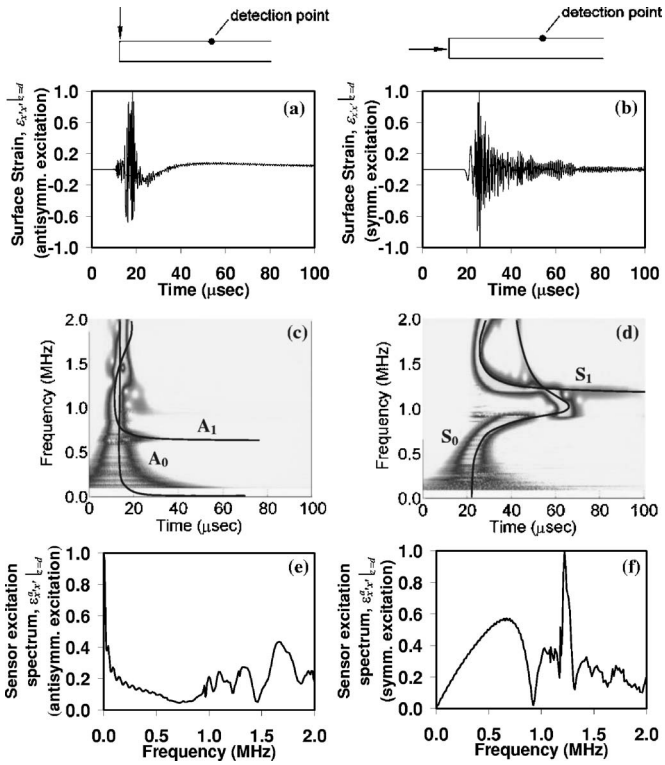


FIG. 8. FEA time histories of the in-plane surface strain in a plate generated by broadband antisymmetric (a) and symmetric (b) excitation. Wavelet transform spectrograms of antisymmetric (c) and symmetric (d) responses. Sensor excitation spectra for symmetric (e) and antisymmetric (f) excitations.

contribution from that of the other modes when possible. For this purpose, continuous wavelet transform (CWT) spectrograms, shown in Figs. 8(c) and 8(d), were calculated from the strain time histories. A complex Morlet mother wavelet (center frequency equal to 5, bandwidth parameter equal to 2) was used. The theoretical arrival times of the relevant modes from Rayleigh-Lamb theory are indicated as solid lines. It can be seen from the spectrograms that A_0 and S_0 are exclusively excited only below the cut-off frequencies of A_1($\sim$0.7 MHz) and S_1($\sim$1.2 MHz), respectively.

In order to compute the sensor excitation spectrum, $\bar{P}$, the fast Fourier transform (FFT) amplitude of the strain signals was calculated and is shown in Figs. 8(e) and 8(f) for the antisymmetric and the symmetric cases, respectively. Based on the CWT scalograms, prior to FFT processing appropriate time gates were applied to the strain time histories to isolate the contribution of the zero-order modes. For the antisymmetric excitation, Fig. 8(c), a time gate of 5–90 μs was employed to capture the A_0 contribution in the frequency ranges 0–0.65 MHz and 0.75–2 MHz. Between these frequency ranges, from 0.65 to 0.75 MHz, a separate time gate of 0–20 μs was applied in order to separate A_0 from A_1. For the symmetric excitation, Fig. 8(d), the S_0 contribution was calculated using a time gate of 0–80 μs in the frequency range 0–1.17 MHz, and a separate gate of 40–70 μs in the range 1.17–2 MHz. These gates, however, could not fully isolate the contribution of S_0 from that of S_1 at around 1.25 MHz due to their equivalent group velocities. As a result, the spike in the symmetric spectrum of Fig. 8(f) at

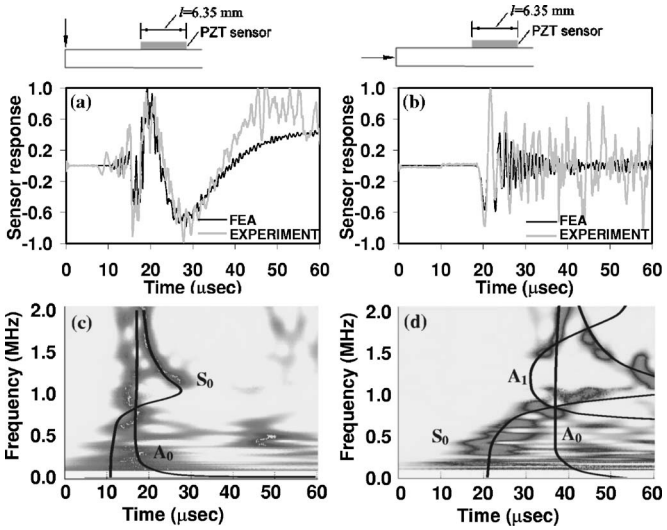


FIG. 9. Comparison between experiment and FEA of sensor response to broadband antisymmetric (a) and symmetric (b) excitation of a plate. Wavelet transform spectrograms of antisymmetric (c) and symmetric (d) responses.

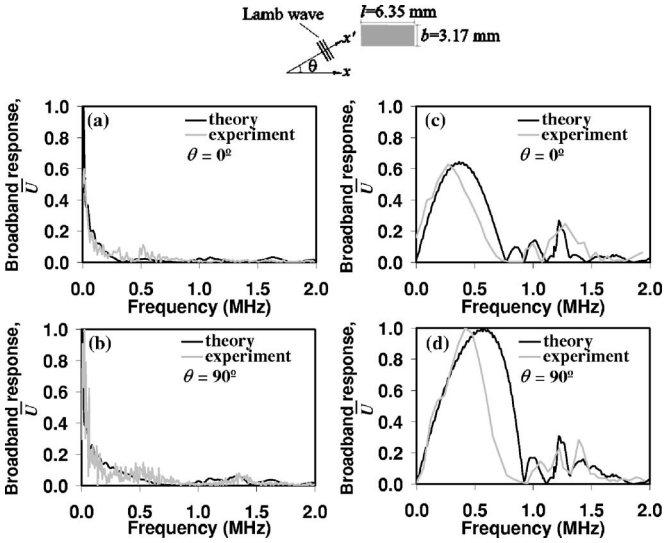


FIG. 10. Experimental and theoretical frequency response of 6.35 mm $\times$3.17 mm rectangular piezoelectric sensor to broadband antisymmetric [(a), (b)] and symmetric [(c), (d)] Lamb waves propagating at 0° [(a), (c)] and 90° [(b), (d)] from the lengthwise sensor dimension.

around 1.25 MHz reflects the contribution of S_1. Despite this interference, the sensor response $\bar{U}$ was evaluated by only considering the harmonic response to S_0. Although a more accurate representation should consider the superimposition of each modal contribution, this simplification is deemed sufficient given the dominant contribution of the zero order modes and the time gates applied to limit the role of higher modes.

2. Experimental analysis

Experimental tests were conducted on a 216 mm $\times$172 mm$\times$2.38 mm (8.5 in.$\times$6.7 in.$\times$0.09 in.) aluminum plate upon which a PZT sensor of dimensions 6.35 mm $\times$3.17 mm$\times$0.5 mm (0.25 in.$\times$0.125 in.$\times$0.02 in.) was bonded using a thin layer of instant adhesive. The same Nd:YAG laser used for the Rayleigh wave tests was employed to generate the broadband Lamb waves in the plate in the slightly ablative regime. For the antisymmetric excitation, the laser pulse was applied normal to the plate surface at 40 mm from the sensor. For the symmetric excitation, the laser pulse was applied at the plate's free edge at 109 mm from the sensor. Results were collected for waves propagating at angles θ=0° and θ=90° from the lengthwise direction of the sensor. The sensor readings were processed through the CWT, gated in time to isolate the fundamental Lamb modes when possible, and then processed by the FFT.

Figure 9 compares the voltage time histories measured by the sensor to the $\varepsilon^a_{x'x'}|_{z=d}$ surface strain computed by the FEA simulation and averaged over the sensor area. The two results should coincide if the simulation represented exactly the experimental conditions. The wave incidence angle is θ =0°. The plots in Figs. 9(a) and 9(b) correspond to the antisymmetric and the symmetric excitations, respectively. In both cases, the earlier portion ($<$30 μs) of the measurements is in good agreement with the averaged strain from the FEA. Beyond 30 μs, the agreement is poorer primarily due to the reflections from the plate edges that are not present in the simulation, and also due to slightly different contributions of the higher-order modes.

The CWT spectrograms of the experimental results are shown in Figs. 9(c) and 9(d), again with the theoretical Rayleigh-Lamb solutions for the visible modes. It can be seen that a S_0 contribution appears in the antisymmetric excitation, Fig. 9(c). Similarly, A_0 and A_1 contributions appear in the symmetric excitation, Fig. 9(d). An edge reflection can also be seen at around 50 μs for the antisymmetric excitation.

The experimental voltage responses of the sensor to broadband excitation are shown as grey lines in Fig. 10 for the wave incidence angles θ=0° and θ=90°. Superimposed as dark lines are the sensor responses, $\bar{U}$, calculated from the semianalytical analysis through Eq. (18). The response to antisymmetric excitation, Figs. 10(a) and 10(b), is dominated by the asymptotically large amplitude at DC. The wavelength tuning condition of zero response, $\lambda=l$, can be observed in both the θ=0° incidence at f=0.37 MHz, and in the θ=90° incidence at f=0.8 MHz.

As for the response to symmetric excitation, Figs. 10(c) and 10(d), local maxima are found at around 0.5 MHz for both incidence angles. This large response is due to the substantial S_0 contribution seen in Fig. 8(f), rather than to any wavelength tuning phenomena. Notice also that the magnitude of the response is generally larger for the perpendicular incidence (where the sensor dimension along the wave propagation direction, 3.17 mm, is small), consistent with the harmonic response of Fig. 7(b). The zero response seen in both Figs. 10(c) and 10(d) at 0.9 MHz is the pivotal point at which the S_0 in-plane strain at the surface vanishes (Fig. 6). Most of the other frequency values with zero response correspond to the wavelength tuning points $\lambda=l/n$. Once more, the spike at 1.25 MHz is due to the contribution of the S_1 mode.

Because of the difficulty in generating a pure mode in

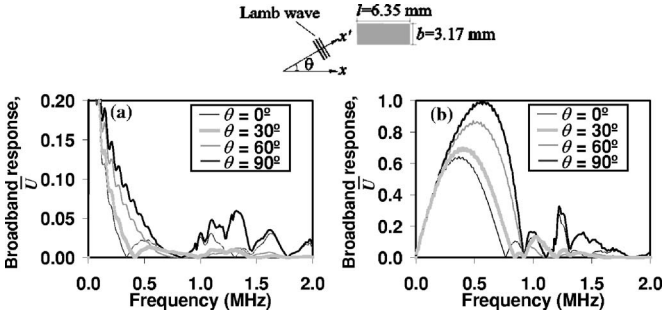


FIG. 11. Theoretical frequency response of rectangular sensor to broadband antisymmetric (a) and symmetric (b) Lamb wave excitation as a function of the wave incidence angle.

practical broadband Lamb wave tests, the exact sensor excitation will depend on the geometrical and the physical properties of the plate, as well as the distance between the excitation and the detection points. The good agreement between experiment and theory shown in Fig. 10 suggests that a semianalytical approach of the type used in this study could be used to reasonably predict the sensor response to broadband Lamb waves in other cases.

A more complete picture of the effect of the wave incidence angle is given in Fig. 11, where the response is calculated for $\theta=0°$, 30°, 60°, and 90° for broadband antisymmetric Lamb excitation, Fig. 11(a), and symmetric Lamb excitation, Fig. 11(b). These results represent $\bar{U}$ in Eq. (18) calculated from the semianalytical analysis. The plots for $\theta=0°$ and 90° coincide with those of Fig. 10. As expected, the results indicate an increase in response magnitude with decreasing sensor dimension along the wave propagation direction (i.e., increasing θ). Wavelength tuning effects are also visible. Notice again the pivotal, zero-response frequency of 0.9 MHz in the symmetric responses of Fig. 11(b) that occurs at all incidence angles.

D. Narrowband Lamb excitation

The same tone burst excitation examined for the Rayleigh wave is considered for the Lamb wave case. Equation (18) is, again, used to calculate the sensor response as $\bar{U}=\bar{V}\times\bar{P}$. For parallel incidence ($\theta=0°$), the harmonic response $\bar{V}$ is given by Eq. (32). The excitation spectrum $\bar{P}$ is given by Eq. (25) with the substitution of A_{Rayleigh} with A_{antisymm} or A_{symm} for the antisymmetric and symmetric cases, respectively. The total sensor response can thus be written as

$$\bar{U}=D\left[\frac{\sin n\pi\left(\frac{2\pi c_p}{\lambda\omega_0}+1\right)}{\left(\frac{2\pi c_p}{\lambda}+\omega_0\right)}-\frac{\sin n\pi\left(\frac{2\pi c_p}{\lambda\omega_0}-1\right)}{\left(\frac{2\pi c_p}{\lambda}-\omega_0\right)}\right]\frac{S\lambda}{\pi l}\times\sin\left(\frac{\pi l}{\lambda}\right), \tag{34}$$

where the wavelength λ is explicitly indicated, and c_p is the phase velocity of the given Lamb mode. Eqation (34) is formally identical to that obtained for the narrowband Rayleigh response in Eq. (26). Also, the expression applies to both antisymmetric and symmetric Lamb waves. The individual response spectra, however, will be different because the velocity and the wavelength at each frequency in Eq. (26) or (34) will depend on the particular wave (Rayleigh, antisymmetric Lamb, or symmetric Lamb mode).

Results from Eq. (34) are plotted in Fig. 12 for sensor

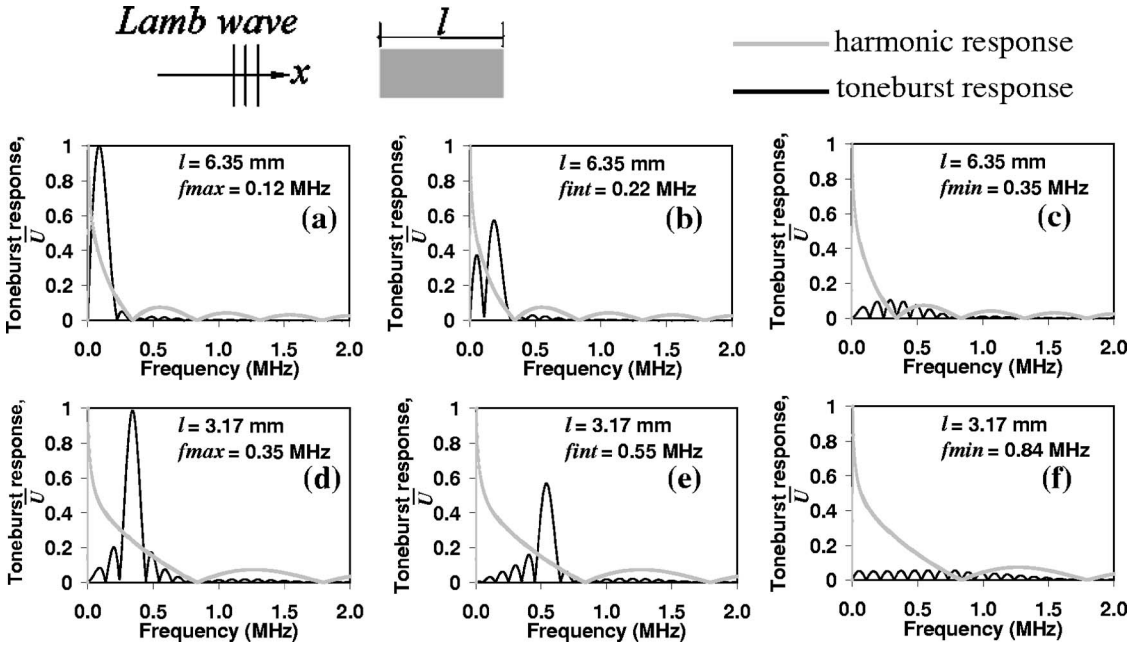


FIG. 12. Frequency response of piezoelectric sensor to narrowband (tone burst) antisymmetric A_0 Lamb wave centered at frequencies corresponding to a large (a), an intermediate (b), and a zero (c) harmonic response for sensor length $l=6.35$ mm. (d)–(f) The same sensor length $l=3.17$ mm. Harmonic responses in gray lines.

lengths $l=6.35$ and 3.17 mm subjected to a narrowband antisymmetric excitation in the 2.38-mm-thick aluminum plate. As for the Rayleigh wave case, the three different center frequencies for the tone bust correspond to a large value (f_{max} at $\lambda=2l$), an intermediate value (f_{int}), and the first zero value (f_{min} at $\lambda=l$) of the sensor's response to harmonic A_0. These frequencies are: $f_{max}=0.11$ MHz, $f_{int}=0.21$ MHz, $f_{min}=0.34$ MHz for the 6.35 mm sensor; $f_{max}=0.34$ MHz, $f_{int}=0.55$ MHz, $f_{min}=0.83$ MHz for the 3.17 mm sensor. Large responses are obtained when the tone burst center frequency corresponds to the favorable wavelength tuning condition. The non-negligible outputs at zero harmonic response, f_{min}, is a consequence of the finite bandwidth of the tone burst and its associated sidelobes. Similar results, not shown here, can be found for the symmetric excitation.

V. CONCLUSIONS

This paper addresses the fundamental response of ultrasonic piezoelectric sensor patches that are increasingly used in on-board structural health monitoring applications. The work complements previous investigations which studied the transduction efficiency of either piezoelectric actuators alone, or piezoelectric actuator-sensor pairs in pitch-catch testing configurations. The problem is studied by coupling the essential behavior of the sensor as a strain integrator to the strain field associated to a particular wave type.

Analytical expressions are first derived for the harmonic response to Rayleigh and Lamb waves impinging on the rectangular sensor at a generally oblique incidence. The parallel incidence (wave propagating along a sensor's geometrical axis) is treated as a particular case of the oblique incidence. The results are normalized by the power flow carried by the wave so as to compare different frequencies in a consistent manner. For parallel incidence, the harmonic solutions contain the wavelength tuning term, $\sin(\pi l/\lambda)$, previously found in the actuator response models. This term predicts local maxima in the response for sensor dimensions equal to an odd multiple of half the wavelength, and zeroes in the response for sensor dimensions equal to an integer multiple of the wavelength. However, the presence of other terms related to the wave strain field smears the wavelength tuning effect by shifting the position of the local maxima and controlling the relative magnitude of these maxima. The additional terms also produce interesting phenomena not related to wavelength tuning, such as an asymptotically large response to A_0 Lamb modes at low frequencies, and an identically zero response to S_0 Lamb modes at the point of inflection of their phase velocity dispersion curve. The responses generally increase with decreasing sensor dimensions. The point sensor (zero dimension) results in the largest response, identically matching the wave strain field spectrum and thus confirming the "ideal sensor" behavior.

Solutions are then calculated for the response of the sensor to both broadband and narrowband Rayleigh and Lamb wave fields, based on the harmonic responses and on the frequency spectrum of the particular excitation.

Regarding the broadband response, finite element analyses were performed to determine the wave field spectrum impinging onto the sensor as a consequence of a broadband excitation of the structure. The theoretical predictions match well with experimental tests conducted on instrumented specimens in which waves were excited by a Q-switched laser pulse. Although the harmonic wavelength tuning behavior is generally seen in the broadband response, some distinct differences exist. These include changes in the relative magnitude of the local response maxima in all cases examined, and a zero response to the S_0 excitation at DC.

One important issue when examining the response to broadband Lamb excitation is the presence of multiple modes. Although efforts were taken to isolate the contribution of the fundamental A_0 or S_0 modes, separating the effects of other modes was not possible in some limited frequency ranges. While the sensor excitation spectrum contained multiple modes, the corresponding harmonic responses were assumed to be those to the fundamental ones (A_0 or S_0). This approximation, which was strictly accurate below the cut-off frequencies of the higher-order modes, proved acceptable as the predicted sensor responses matched reasonably well with the measurements.

Regarding the narrowband response, the solutions were developed for unmodulated tone burst excitation. The results confirmed the intuitive conclusion that large responses are obtained when the tone burst center frequency coincides with a point of large harmonic response for the sensor.

The results presented demonstrate the critical role played by the dimension of the sensor. The solutions can be used to design specific sensors with maximum response to a given mode and minimum influence of other modes. It is also clear that the use of rectangular sensors with well-designed dimensions can achieve a very controllable directivity response behavior (for example, unidirectional response to a given ultrasonic mode in a certain frequency range). Such directivity behavior can be exploited to greatly simplify both active-passive damage detection systems based on diffraction measurements,[8] and passive-only, acoustic emission damage/impact detection systems. In both of these cases, in fact, the use of rosette-like detection nodes based on unidirectional sensors (as demonstrated recently with fiber-optic sensors[38]) would avoid the necessity for prior knowledge of the wave velocity in the medium, which is particularly appealing when anisotropic materials are involved.

The results presented assume straight-crested waves, which is accurate for sufficient distances between a point source and a receiver. A natural extension would be the study of circularly crested waves.

ACKNOWLEDGMENTS

This project was funded by the Los Alamos/UCSD Education Collaboration Task 2 "Structural Integrity Monitoring of UAV Composite Wings" and by the UCSD/Los Alamos Cooperative Agreement on Research and Education (CARE). Thanks are extended to Dr. Charles Farrar and Dr. Gyuhae Park of the Los Alamos National Laboratory, ESA Division, for useful technical discussions throughout this study.

[1]R. S. C. Monkhouse, P. Wilcox, and P. Cawley, "Flexible interdigitated PVDF transducers for the generation of Lamb Waves in structures," Ul-

trasonics **35**, 489–498 (1997).
[2]A. K. Mal, F. J. Shih, and W. Prosser, "Lamb waves from impact damage in composite plates," Instrumentation-Measure-Metrologie, Special Issue on Ultrasonic Methods for Material Characterization **3**, 11–37 (2003).
[3]A. K. Mal, "Structural health monitoring," Mechanics, American Academy of Mechanics **33** (2004).
[4]A. K. Mal, F. Ricci, S. Banerjee, and F. Shih, "A conceptual structural health monitoring system based on vibration and wave propagation," Struct. Health Monit. **4**, 283–293 (2005).
[5]S. H. Valdes and C. Soutis, "Health monitoring of composites using Lamb waves generated bypiezoelectric devices," Plast. Rubber Compos. **29**, 475–481 (2000).
[6]S. C. Wang and F.-K. Chang, "Diagnosis of impact damage in composite structures with built-in pieozelectrics network," Proceedings of the SPIE Smart Structures and Materials Conference, San Diego, 2000, Vol. **3990**, pp. 13–19.
[7]V. Giurgiutiu and A. N. Zagrai, "Characterization of piezoelectric wafer active sensors," J. Intell. Mater. Syst. Struct. **11**, 959–975 (2000).
[8]M. Lemistre and D. Balageas, "Structural health monitoring system based on diffracted Lamb wave analysis by multiresolution processing," Smart Mater. Struct. **10**, 504–511 (2001).
[9]R. P. Dalton, P. Cawley, and M. J. S. Lowe, "The potential of guided waves for monitoring large areas of metallic aircraft fuselage structures," J. Nondestruct. Eval. **20**, 29–46 (2001).
[10]E. Blaise and F.-K. Chang, "Built-in damage detection system for sandwich structures under cryogenic temperatures," Proceedings of the SPIE Smart Structures Conference, San Diego, 2002, Vol. **4701**, pp. 97–107.
[11]S. S. Kessler, S. M. Spearing, and C. Soutis, "Damage detection in composite materials using Lamb wave methods," Smart Mater. Struct. **11**, 269–278 (2002).
[12]C. A. Paget, S. Grondel, K. Levin, and C. Delabarre, "Damage assessment in composites by Lamb waves and wavelet coefficients," Smart Mater. Struct. **12**, 393–402 (2002).
[13]P. T. Coverly and W. J. Staszewski, "Impact damage location in composite structures usingoptimized sensor triangulation procedure," Smart Mater. Struct. **12**, 795–803 (2003).
[14]V. Giurgiutiu, "Embedded NDE with piezoelectric wafer active sensors in aerospace applications," Journal of Materials, http://www.tms.org/pubs/journals/JOM/0301/Giurgiutiu/Giurgiutiu-0301.html (2003). Last accessed 12/6/06.
[15]H. Sohn, G. Park, J. R. Wait, N. P. Limback, and C. R. Farrar, "Wavelet-based active sensing for delamination detection in composite structures," Smart Mater. Struct. **13**, 153–160 (2004).
[16]Z. Su and L. Ye, "Fundamental Lamb mode-based delamination detection for CF/EP composite laminates using distributed piezoelectrics," Struct. Health Monit. **3**, 43–68 (2004).
[17]H. M. Matt, I. Bartoli, and F. Lanza di Scalea, "Ultrasonic guided wave monitoring of composite wing skin-to-spar bonded joints in aerospace structures," J. Acoust. Soc. Am. **118**, 2240–2252 (2005).
[18]V. Giurgiutiu, "Tuned Lamb wave excitation and detection with piezoelectric wafer active sensors for structural health monitoring," J. Intell. Mater. Syst. Struct. **16**, 291–305 (2005).
[19]P. F. Lichtenwalner and D. Sofge, "A local area damage detection in composite structures using piezoelectric transducers," Proceedings of the SPIE Smart Structures Conference, San Diego, 1998, Vol. **3326**, pp. 509–515.
[20]G. L. Wojcik, A. N. Vaughan, and J. Mould Jr., "Electromechanical modeling using explicit time-domain finite elements," Proc.-IEEE Ultrason. Symp. 1107–1112 (1993).
[21]J. Sirohi and I. Chopra, "Fundamental understanding of piezoelectric strain sensors," J. Intell. Mater. Syst. Struct. **11**, 246–257 (2000).
[22]E. Moulin, J. Assaad, and C. Delebarre, "Modeling of Lamb waves generated by integrated transducers in composite plates using a coupled finite element-normal modes expansion method," J. Acoust. Soc. Am. **107**, 87–94 (2000).
[23]X. Lin and F. G. Yuan, "Diagnostic Lamb waves in an integrated piezoelectric sensor/actuator plate: Analytical and experimental studies," Smart Mater. Struct. **10**, 907–913 (2001).
[24]A. Raghavan and C. E. Cesnik, "Modeling of piezoelectric-based Lamb-wave generation and sensing for structural health monitoring," Proceedings of the SPIE Smart Structures Conference, San Diego, 2004, Vol. **5391**, pp. 419–430.
[25]D. W. Greve, J. J. Neumann, J. H. Nieuwenhuis, I. J. Oppenheim, and N. L. Tyson, "Use of Lamb waves to monitor plates: Experiments and simulations," Proceedings of the SPIE Smart Structures Conference, San Diego, 2005, Vol. **5765**, pp. 281–292.
[26]J. H. Nieuwenhuis, J. J. Newumann, D. W. Greve, and I. J. Oppenheim, "Generation and detection of guided waves using PZT wafer transducers," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **52**, 2103–2111 (2005).
[27]IEEE, *Piezoelectricity IEEE Standard 176* (IEEE, New York, 1978).
[28]M. S. Azzouz, C. Mei, J. S. Bevan, and J. J. Ro, "Finite element modeling of MFC/AFC actuators and performance of MFC," J. Intell. Mater. Syst. Struct. **12**, 601–612 (2001).
[29]H. S. Tzou, *Piezoelectric Shells: Distributed Sensing and Control of Continua* (Kluwer Academic, Dordrecht, 1993).
[30]J. L. Rose, *Ultrasonic Waves in Solid Media* (Cambridge University Press, Cambridge, 1999).
[31]E. F. Crawley and J. de Luis, "Use of piezoelectric actuators as elements of intelligent structures," AIAA J. **25**, 1373–1385 (1987).
[32]A. K. Chopra, *Dynamics of Structures* (Prentice-Hall, Upper Saddle River, NJ, 2001).
[33]*Abaqus User's Manual*, Version 6.5, Hibbit, Karlsson, and Sorensen, Inc.
[34]D. Alleyne and P. Cawley, "A two-dimensional Fourier transform method for the measurement of propagating multimode signals," J. Acoust. Soc. Am. **89**, 1159–1168 (1991).
[35]F. Moser, L. J. Jacobs, and J. Qu, "Modeling elastic wave propagation in waveguides with the finite element method," NDT & E Int. **32**, 225–234 (1999).
[36]A. D. Poularikas, *The Transforms and Applications Handbook* (CRC Press, Boca Raton, FL, 1996).
[37]I. A. Viktorov, *Rayleigh and Lamb Waves* (Plenum, New York, 1967).
[38]G. Thursby, B. Sorazu, D. Betz, and B. Culshaw, "Novel methods of Lamb wave detection for material damage detection and location," Proceedings of the SPIE Smart Structures and Materials Conference, San Diego, 2005, Vol. **5768**, pp. 313–322.

The stability of decentralized multichannel velocity feedback controllers using inertial actuators

Oliver Nicholas Baumann[a] and Stephen John Elliott[b]
Institute of Sound and Vibration Research, University of Southampton, Southampton, Hampshire SO17 1BJ, United Kingdom

(Received 29 June 2006; revised 24 October 2006; accepted 30 October 2006)

The application of direct velocity feedback control on vibrating structures is well known to provide additional damping and reduce vibration levels. A number of previously studied control systems use multiple feedback loops with ideal velocity sensors and force actuators. While accelerometer signals may be utilized to accurately estimate velocity, there is rarely a structure off which one may react an ideal force. This paper concentrates on the use of multiple electrodynamic inertial actuators as a means of applying a force. A time domain model of a plate structure with multiple velocity sensors and collocated inertial actuators is derived. This model is then used to optimize the decentralized controller in order to minimize the total kinetic energy of the plate. These results are compared with those obtained with a decentralized controller in which each local loop has the same gain. It is demonstrated that for low control efforts, and hence control gains, both controllers perform almost identically, however at large gains the equal gain controller becomes unstable. The cause of this instability is attributed to the resonance of the inertial actuator. The implications of using multiple inertial actuators is discussed and some experimental results are presented and compared with simulations. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2400674]

PACS number(s): 43.40.Vn, 43.40.At, 43.40.Dx [KAC] Pages: 188–196

I. INTRODUCTION

It is well known that direct feedback from velocity sensors to collocated ideal force actuators adds damping to a structure and is unconditionally stable, even for multiple channels.[1–3] While velocity signals may be derived relatively easily from accelerometers above a certain frequency, the notion of an ideal point force is rarely realizable since there is generally no other structure to react such a force off. Considerable attention has been paid to the use of piezoelectric materials for actuation. However in most cases the effect of the piezoelectric patch on the plate approximates a moment pair and not a point force, the result being that the sensor and actuator no longer form a collocated dual pair and stability cannot be guaranteed.[4] Here we concern ourselves with the use of electrodynamic inertial devices, such as that depicted schematically in Fig. 1, used in local feedback loops. In such devices the current in a coil suspended in a magnetic gap produces a force between the structure and a proof mass.[5–7]

This paper focuses on the use of inertial actuators in a multichannel decentralized velocity feedback controller for the reduction of the vibration on a plate. The plate and controller may be represented as in Fig. 2, in which both the primary disturbance **d** and the secondary control signals **u** act upon the plate which is represented as a generalized plant **G** and result in a set of measured velocities **y**. These velocities are fed back via a controller **H**, in this case consisting of a diagonal matrix of fixed gains, to produce the set of control signals. The error output **e** pertains to some measure of the system's performance, which may, for example, be the kinetic energy of the plate. In previous work[8] a plate was considered with collocated velocity sensors and ideal force actuators connected in individual loops. In this configuration each collocated sensor and actuator comprise a dual pair and therefore the system under control is passive and the decentralized controller is unconditionally stable. Decentralized controllers for vibration control have other attractive properties, namely the simplicity of implementation and scalability when compared to fully coupled, centralized controllers. It has also been demonstrated[8] that for dual, collocated sensors and actuators the performance of these controllers approaches that of a centralized controller. The majority of previous work,[1,3,9] however, has concentrated on the use of ideal forces or piezoelectric devices for actuation.

We first introduce a time domain model of a thin aluminum plate which includes the dynamics of multiple inertial actuators. The design of the controller is then discussed in terms of the minimization of a quadratic cost function with kinetic energy and control effort terms. It is shown that for reasonable levels of control effort the optimal decentralized controller does not perform significantly better than a controller in which each local control loop is given an identical gain. It is also observed however that neither the optimal nor the equal gain controller is unconditionally stable. The stability of the equal gain controller is then discussed in terms of the generalized Nyquist criterion and it is observed that the coupling of the actuator and plate dynamics results in a system in which as the number of channels is increased the relative stability of the controller is significantly reduced. Finally the results of experimental trials are presented and compared with the predictions of the model.

[a]Electronic mail: onb@isvr.soton.ac.uk
[b]Electronic mail: sje@isvr.soton.ac.uk

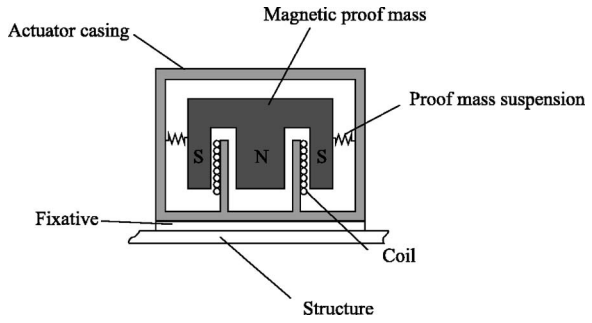


FIG. 1. Schematic of a concept inertial actuator.

II. PLATE MODEL

This section introduces the model of a clamped rectangular plate with inertial actuators and velocity sensors. We begin by assuming a modal formulation in which the transverse displacement of the plate $w(x,y,t)$ is approximated by a finite summation of M modal amplitudes $a_m(t)$ and mode shape functions $\psi_m(x,y)$,[10]

$$w(x,y,t) = \sum_{m}^{M} a_m(t)\psi_m(x,y). \tag{1}$$

The displacement at a finite set of N control locations described by the vector couple $\mathbf{x}_s$, $\mathbf{y}_s$ may be written in matrix form as

$$\mathbf{w}(\mathbf{x}_s,\mathbf{y}_s,t) = \mathbf{\Psi}(\mathbf{x}_s,\mathbf{y}_s)^T\mathbf{a}(t), \tag{2}$$

where $\mathbf{\Psi}(\mathbf{x}_s,\mathbf{y}_s)$ is matrix of dimensions M,N containing the values of mode shape functions evaluated at $\mathbf{x}_s$, $\mathbf{y}_s$, and $\mathbf{a}(t)$ is a column vector of the M modal amplitudes. The mode shape functions comprising $\mathbf{\Psi}(\mathbf{x}_s,\mathbf{y}_s)$ satisfy the imposed boundary conditions of the plate and the modal amplitude vector satisfies the second-order differential equation

$$\ddot{\mathbf{a}}(t) = \mathbf{D}\dot{\mathbf{a}}(t) + \mathbf{K}\mathbf{a}(t)\frac{4}{M_p}\mathbf{f}(t), \tag{3}$$

where M_p is the total mass of the plate and $\mathbf{D}$ and $\mathbf{K}$ are square matrices of dimensions M having the terms $-2\zeta_m\omega_m$ and $-\omega_m^2$, respectively, on the diagonal where ζ_m is the modal damping ratio and ω_m the resonance frequency of the mth mode. The vector of modal forces $\mathbf{f}(t)$ is due to the control forces which are provided by means of inertial actuators whose input is described by the vector $\mathbf{u}(t)$. Figure 3 shows a lumped parameter model of the concept inertial actuator presented in Fig. 1. Variables with the subscript p refer to the properties of the proof mass, its suspension, and displacement w_p. Variables with subscript c refer to the properties of the casing mass, the fixative properties, and the displacement

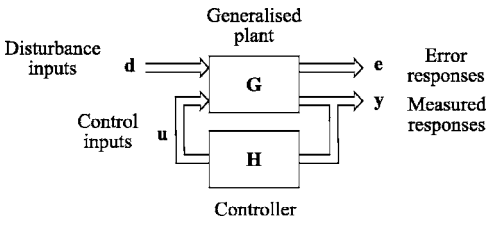


FIG. 2. Generalized plant model.

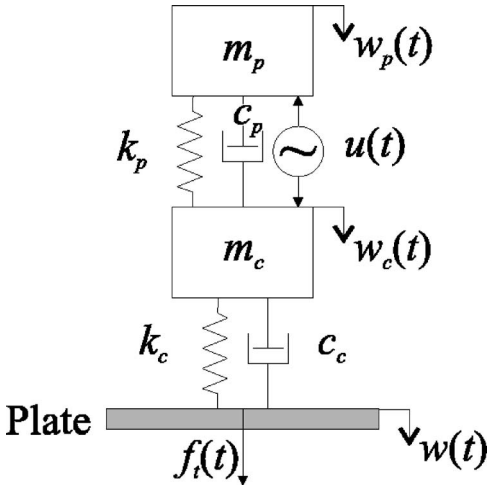


FIG. 3. Actuator model schematic.

of the casing w_c. It should be noted that the stiffness and damping of the actuator fixture, k_c and c_c, are included here to aid the modeling of the coupled system in state space. In practice this fixture will be very secure, which may be modeled by using values of k_c and c_c which result in a highly damped resonance having a natural frequency some way above the frequency range of interest. In this way the casing mass acts as though it is attached directly to the structure. The properties of the actuators used to obtain the results presented in this paper are given in Table I.

The force acting between the two masses is due to the current in the coil and therefore if the actuator is assumed to be driven within its linear range by a constant current amplifier this force is directly proportional to the control signal $u(t)$. For the purposes of simulation, the constant of proportionality or force constant is taken to be 1 N/A such that $\mathbf{f}_t(t)=\mathbf{u}(t)$. The vector of point forces acting on the structure due to N identical inertial actuators mounted at the control locations is given by

$$\mathbf{f}_t(t) = k_c(\mathbf{w}(\mathbf{x}_s,\mathbf{y}_s,t) - \mathbf{w}_c(t)) + c_c(\dot{\mathbf{w}}(\mathbf{x}_s,\mathbf{y}_s,t) - \dot{\mathbf{w}}_c(t)), \tag{4}$$

where $\mathbf{w}(\mathbf{x}_s,\mathbf{y}_s,t)$ is given by Eq. (2) and $\mathbf{w}_c(t)$ is a column vector of length N describing the displacements of the casing masses which satisfies the differential equation

TABLE I. Table of actuator properties.

Property	Value
m_p	0.032 kg
k_p	140 N/m
c_p	1.39 N s/m
m_c	0.050 kg
k_c	1×10^9 N/m
c_c	1×10^9 N s/m
Force constant	1 N/A

$$\ddot{\mathbf{w}}_c(t) = -\frac{k_p}{m_c}(\mathbf{w}_c(t) - \mathbf{w}_p(t)) + \frac{c_p}{m_c}(\dot{\mathbf{w}}_c(t) - \dot{\mathbf{w}}_p(t)) + \frac{1}{m_c}\mathbf{u}(t) - \frac{k_c}{m_c}(\mathbf{w}_c(t) - \mathbf{w}(t)) - \frac{c_c}{m_c}(\dot{\mathbf{w}}_c(t) - \dot{\mathbf{w}}(t)), \tag{5}$$

where $\mathbf{w}_p$ is a vector of proof mass displacements which itself satisfies

$$\ddot{\mathbf{w}}_p(t) = -\frac{k_p}{m_p}(\mathbf{w}_p(t) - \mathbf{w}_c(t)) - \frac{c_p}{m_p}(\dot{\mathbf{w}}_p(t) - \dot{\mathbf{w}}_c(t)) - \frac{1}{m_p}\mathbf{u}(t). \tag{6}$$

The vector of modal forces $\mathbf{f}(t)$ may now be expressed in terms of the vector of transmitted forces

$$\mathbf{f}(t) = \boldsymbol{\Psi}(\mathbf{x}_s, \mathbf{y}_s)\mathbf{f}_t(t). \tag{7}$$

Substitution of Eqs. (4), (7), and (2) into Eq. (3) yields

$$\ddot{\mathbf{a}}(t) = \left\{\mathbf{D} - \frac{4c_c}{m_p}\boldsymbol{\Psi}(\mathbf{x}_s, \mathbf{y}_s, t)^T\boldsymbol{\Psi}(\mathbf{x}_s, \mathbf{y}_s, t)\right\}\dot{\mathbf{a}} + \left\{\mathbf{K} - \frac{4k_c}{m_p}\boldsymbol{\Psi}(\mathbf{x}_s, \mathbf{y}_s, t)^T\boldsymbol{\Psi}(\mathbf{x}_s, \mathbf{y}_s, t)\right\}\mathbf{a} + \frac{4c_c}{m_p}\boldsymbol{\Psi}(\mathbf{x}_s, \mathbf{y}_s, t)\dot{\mathbf{w}}_c + \frac{4k_c}{m_p}\boldsymbol{\Psi}(\mathbf{x}_s, \mathbf{y}_s, t)\mathbf{w}_c. \tag{8}$$

Equations (5), (6), and (8) completely describe the dynamics of M modes of the plate with N inertial actuators. These may be cast into a time domain state space model of the form

$$\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}_u\mathbf{u}(t) + \mathbf{B}_d\mathbf{d}, \tag{9}$$

in which $\mathbf{x}(t)$ is a vector of system states, the matrix $\mathbf{A}$ describes the systems dynamics, and $\mathbf{B_u}$ and $\mathbf{B_d}$ describe how the control inputs $\mathbf{u}(t)$ and disturbance input $\mathbf{d}(t)$ act on the system states. The state vector takes the form

$$\mathbf{x} = \begin{bmatrix} \mathbf{a}(t) \\ \mathbf{w}_c(t) \\ \mathbf{w}_p(t) \\ \dot{\mathbf{a}}(t) \\ \dot{\mathbf{w}}_c(t) \\ \dot{\mathbf{w}}_p(t) \end{bmatrix}, \tag{10}$$

such that from Eqs. (5), (6), and (8) we derive

$$\mathbf{A} = \begin{bmatrix} \mathbf{0}_M & \mathbf{0}_{M,N} & \mathbf{0}_{M,N} & \mathbf{I}_M & \mathbf{0}_{M,N} & \mathbf{0}_{M,N} \\ \mathbf{0}_{N,M} & \mathbf{0}_N & \mathbf{0}_N & \mathbf{0}_{N,M} & \mathbf{I}_N & \mathbf{0}_N \\ \mathbf{0}_{N,M} & \mathbf{0}_N & \mathbf{0}_N & \mathbf{0}_{N,M} & \mathbf{0}_N & \mathbf{I}_N \\ \mathbf{K} - \frac{4k_c}{M_p}\boldsymbol{\Psi}\boldsymbol{\Psi}^T & \frac{4k_c}{M_p}\boldsymbol{\Psi} & \mathbf{0}_{M,N} & \mathbf{D} - \frac{4c_c}{M_p}\boldsymbol{\Psi} & \frac{4c_c}{M_p}\boldsymbol{\Psi} & \mathbf{0}_{M,N} \\ \frac{k_c}{m_c}\boldsymbol{\Psi}^T & \frac{-k_p - k_c}{m_c} & \frac{k_p}{m_c} & \frac{c_c}{m_c}\boldsymbol{\Psi}^T & \frac{-c_p - c_c}{m_c} & \frac{c_p}{m_c} \\ \mathbf{0}_{N,M} & \frac{k_p}{m_p} & -\frac{k_p}{m_p} & \mathbf{0}_{N,M} & \frac{c_p}{m_p} & -\frac{c_p}{m_p} \end{bmatrix}, \tag{11}$$

in which the arguments of $\boldsymbol{\Psi}(\mathbf{x}_s, \mathbf{y}_s, t)$ have been dropped for brevity and the subscripts of zero and identity matrices represent their dimensions, a single subscript denoting a square matrix. Also from Eqs. (5) and (6) we see that

$$\mathbf{B_u} = \begin{bmatrix} \mathbf{0}_{M,N} \\ \mathbf{0}_N \\ \mathbf{0}_N \\ \mathbf{0}_{M,N} \\ \frac{1}{m_c}\mathbf{I}_N \\ -\frac{1}{m_p}\mathbf{I}_N \end{bmatrix}. \tag{12}$$

We define the primary disturbance of the plate to be that due to a spatially and temporally uncorrelated pressure field such that $\mathbf{d}(t)$ is a set of uncorrelated signals which act directly on the plate modes. From Eq. (3) then

$$\mathbf{B_d} = \begin{bmatrix} \mathbf{0}_M \\ \mathbf{0}_{N,M} \\ \mathbf{0}_{N,M} \\ \frac{4}{M_p}\mathbf{I}_M \\ \mathbf{0}_{N,M} \\ \mathbf{0}_{N,M} \end{bmatrix}. \tag{13}$$

The output of the model, being the set of velocities at the control locations $\mathbf{y}(t)$, is related to the system states via

$$\mathbf{y}(t) = \mathbf{C}\mathbf{x}(t), \tag{14}$$

in which

$$\mathbf{C} = [\mathbf{0}_{N,M}\ \mathbf{0}_{N,N}\ \mathbf{0}_{N,N}\ \boldsymbol{\Psi}\ \mathbf{0}_{N,N}\ \mathbf{0}_{N,N}]. \tag{15}$$

Referring to the control model presented in Fig. 2 we see that for the system under control the vector of control signals $\mathbf{u}(t)$ is related to the vector of plate velocities $\mathbf{y}(t)$ and hence the system states via

$$\mathbf{u}(t) = -\mathbf{H}\mathbf{y}(t) = -\mathbf{H}\mathbf{C}\mathbf{x}(t), \tag{16}$$

such that Eq. (9) becomes

$$\dot{\mathbf{x}}(t) = [\mathbf{A} - \mathbf{B_u}\mathbf{H}\mathbf{C}]\mathbf{x}(t) + \mathbf{B_d}\mathbf{d}(t), \tag{17}$$

allowing us to define a new system matrix for the controlled system

$$\mathbf{A}_c = \mathbf{A} - \mathbf{B_u}\mathbf{H}\mathbf{C}, \tag{18}$$

such that the state equation becomes

$$\dot{\mathbf{x}}(t) = \mathbf{A}_c\mathbf{x}(u) + \mathbf{B_d}\mathbf{d}(t). \tag{19}$$

This model therefore describes the response of the plant of Fig. 2 both without control and under direct velocity feedback control with a feedback gain matrix. In the following section we define the effort limited kinetic energy cost function which may be used to optimize the controller.

III. CONTROLLER DESIGN

We describe the performance of the system by the quadratic cost function J given by

$$J = \int_0^\infty [\mathbf{e}(t)^T\mathbf{Q}\mathbf{e}(t) + \mathbf{u}(t)^T\mathbf{R}\mathbf{u}(t)]dt, \tag{20}$$

where $\mathbf{e}(t)$ is the vector of error responses depicted in Fig. 2. The first term of the cost function may be arranged to be proportional to the total kinetic energy of the plate by setting the error responses $\mathbf{e}(t)$ to be equal to the vector of system states $\mathbf{x}(t)$ and selecting $\mathbf{Q}$ such that

$$\mathbf{x}(t)^T\mathbf{Q}\mathbf{x}(t) = \frac{M_p}{8}\dot{\mathbf{a}}^T(t)\dot{\mathbf{a}}(t). \tag{21}$$

The second term of Eq. (20) is a weighted sum of the squared control signals or control effort term, $\mathbf{R}$ being the control effort weighting matrix of the form

$$\mathbf{R} = r\mathbf{I}_N. \tag{22}$$

The cost function may also be written as[11]

$$J = \mathrm{trace}(\mathbf{K}_l\mathbf{F}), \tag{23}$$

where $\mathbf{K}_l$ is the positive semidefinite solution of the continuous Lyapunov equation

$$\mathbf{0} = \mathbf{K}_l\mathbf{A}_c + \mathbf{A}_c^T\mathbf{K}_l + \mathbf{Q} + \mathbf{C}^T\mathbf{H}^T\mathbf{R}\mathbf{H}\mathbf{C} \tag{24}$$

and $\mathbf{F}$ is a matrix representing the primary disturbance, which for the case described in the previous section is given by

$$\mathbf{F} = \frac{16}{M_p^2}\begin{bmatrix} \mathbf{0}_{M,M} & \mathbf{0}_{M,N} & \mathbf{0}_{M,N} & \mathbf{0}_M & \mathbf{0}_{M,N} & \mathbf{0}_{M,N} \\ \mathbf{0}_{N,M} & \mathbf{0}_N & \mathbf{0}_N & \mathbf{0}_{N,M} & \mathbf{0}_N & \mathbf{0}_N \\ \mathbf{0}_{N,M} & \mathbf{0}_N & \mathbf{0}_N & \mathbf{0}_{N,M} & \mathbf{0}_N & \mathbf{0}_N \\ \mathbf{0}_{M,M} & \mathbf{0}_{M,N} & \mathbf{0}_{M,N} & \mathbf{I}_M & \mathbf{0}_{M,N} & \mathbf{0}_{M,N} \\ \mathbf{0}_{N,M} & \mathbf{0}_N & \mathbf{0}_N & \mathbf{0}_{N,M} & \mathbf{0}_N & \mathbf{0}_N \\ \mathbf{0}_{N,M} & \mathbf{0}_N & \mathbf{0}_N & \mathbf{0}_{N,M} & \mathbf{0}_N & \mathbf{0}_N \end{bmatrix}. \tag{25}$$

Levine and Athans[12] derive an expression for the matrix gradient of the cost function J with respect to the feedback gain matrix $\mathbf{H}$ as

$$\frac{dJ}{d\mathbf{H}} = 2\mathbf{R}\mathbf{H}\mathbf{C}\mathbf{L}\mathbf{C}^T - 2\mathbf{B}_u^T\mathbf{K}_l\mathbf{L}\mathbf{C}^T, \tag{26}$$

where $\mathbf{L}$ is the positive definite solution of

$$\mathbf{0} = \mathbf{L}\mathbf{A}_c^T + \mathbf{A}_c\mathbf{L} + \mathbf{F}. \tag{27}$$

TABLE II. Table of plate properties.

Property	Description	Value
l_x	Length	0.8 m
l_y	Width	0.5 m
h	Thickness	0.002 m
ρ	Density	2720 kg/m^3
M	Number of modes accounted for	53
$f_{\max}$	Natural frequency of highest mode	1100 Hz

A more complete derivation of these results may be found in Ref. 12. In order to optimize the gain matrix $\mathbf{H}$ the matrix gradient of Eq. (26) is evaluated based upon some initial gain matrix $\mathbf{H}_0$. The gain matrix is then updated by traversing the cost surface in the direction defined by the gradient matrix and by a predetermined step size in order to produce a gain matrix with a lower associated cost. Should a given step size result in a higher cost or an unstable system, the step size is reduced. The algorithm terminates when the step size falls below a set threshold. A full description of the algorithm is presented by Engels *et al.*[8] Although this technique has not been proved to converge on a global minimum of the cost function it has been found in practice that the same minimum is found regardless of the initial controller $\mathbf{H}_0$ provided this controller is stable and the control locations are not extremely close together.[13] For simplicity therefore we shall hereafter refer to the process of cost function minimization as optimization and the resulting controller as optimal. In this paper we are concerned with a decentralized controller in which the collocated velocity sensors and force actuators are connected in local feedback loops. The corresponding feedback gain matrix is therefore diagonal and as such all off-diagonal terms of the gradient matrix in each iteration of the optimization are set to zero prior to updating the feedback gain matrix.

The plate model used throughout is based upon a thin aluminum plate with clamped boundary conditions. The plate properties are presented in Table II. Note that all modes up to the frequency $f_{\max}$=1100 Hz have been used in the formulation resulting in 53 individual modes. Control is achieved via 24 equally spaced sensor/actuator pairs arranged on the plate as depicted in Fig. 4. The cost function described by Eqs.

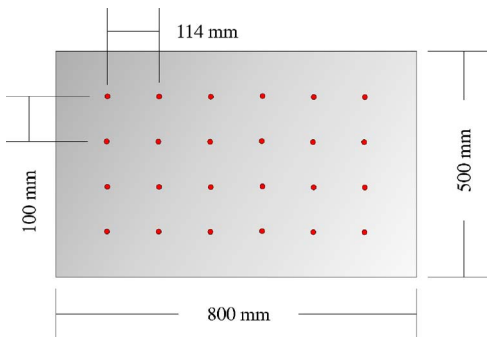


FIG. 4. Schematic of the plate showing the positions of the 24 sensor/actuator pairs.

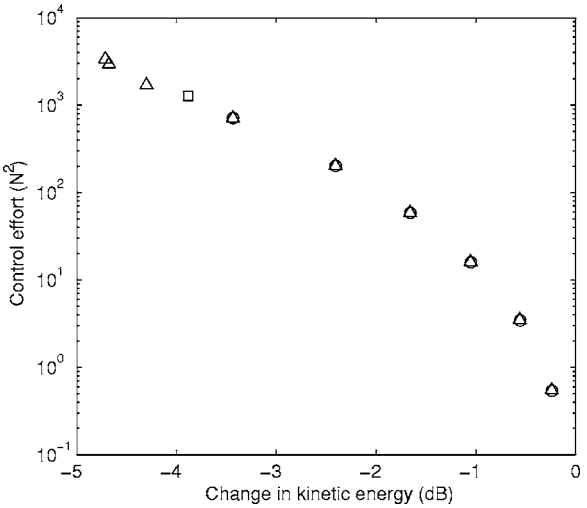


FIG. 5. Plot of control effort vs change in kinetic energy for optimal decentralized controllers (△), corresponding controllers of the form $h\mathbf{I}_N$ (○), and the $h\mathbf{I}_N$ controller with the maximum stable value of h (□).

(20)–(22) is minimized for various control effort weightings r, resulting in a set of feedback gain matrices with associated control effort and kinetic energy.

For each optimal controller we also present the results of a decentralized controller in which all gains are equal such that

$$\mathbf{H} = h\mathbf{I}_N, \tag{28}$$

where h is chosen to be the mean of the diagonal elements of the optimal gain matrix. It should be noted that for some optimal controllers with large associated control effort the corresponding controller with equal gains results in instability and as such is not presented. The reasons for the instability are discussed in the following section. Figure 5 shows a plot of the control effort against the change in kinetic energy relative to the passive system for optimal controllers obtained by minimizing Eq. (20) with various values of r and the equivalent equal gain controllers. It is clear from this plot that the optimal decentralized controller performs almost identically to the controller with equal decentralized gains. This is consistent with the observation that the difference between the largest and the smallest of the gains of the optimal decentralized gain matrices, i.e., the range of gains, is always less than 30% of their mean. It is however interesting to note that as the control effort weighting is reduced the range of gain values increases considerably, the lowest gains being associated with control loops in the center of the plate and the highest associated with those at the edges. For such controllers the range of gains approaches 140% of the mean. Also plotted in Fig. 5 is the control effort and kinetic energy of the equal gain controller in which h has been chosen to be the highest possible value for which the system remains stable. Thus, for all but the greatest reductions in kinetic energy which require very large control efforts, it is reasonable to employ a control strategy in which each local loop has an identical gain.

With this in mind we proceed with the equal gain controller and plot the kinetic energy frequency response of the plate without control and under control for values of h between approximately 3 and 62 which correspond to the points of Fig. 5 having control efforts between 16 and 1300 N^2. The frequency responses are plotted in Fig. 6 over two frequency ranges. We observe that while good control of the plate modes is achievable as the gain is increased, the kinetic energy of the plate between 5 and 15 Hz is significantly increased. The frequency at which this enhancement is observed corresponds to the resonance of the actuator proof mass and results in the eventual instability of the system for $h > 62$. This phenomenon is analyzed in the following section.

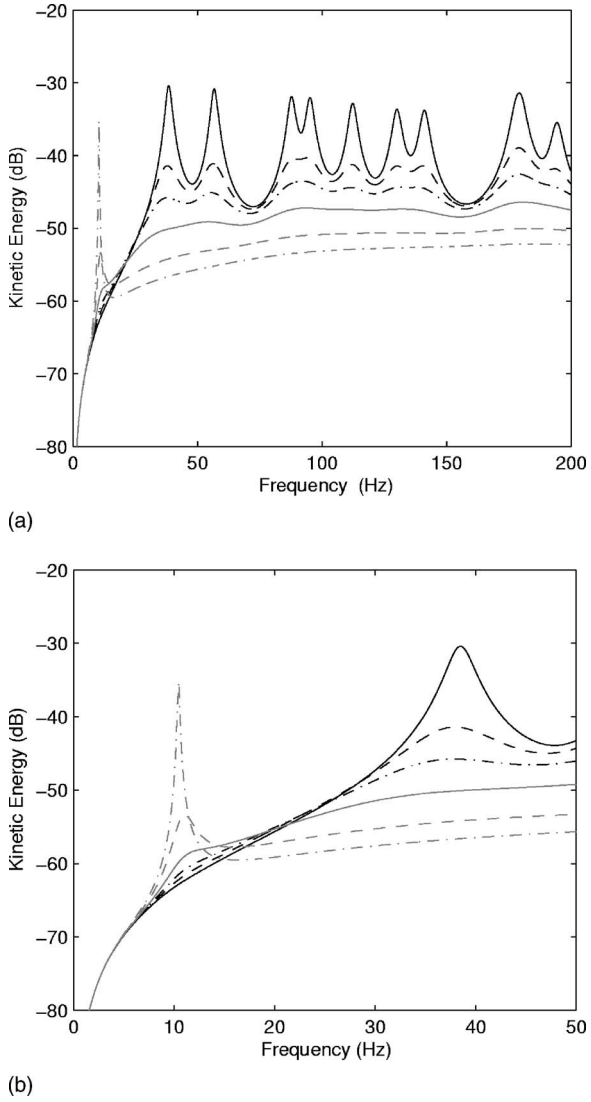


FIG. 6. Kinetic energy frequency responses for the uncontrolled plate (—) and with control for $h=3$ (– – –), $h=7$ (–·–), $h=17$ (—), $h=43$ (- - -), and $h=62$ (··· ···). Results are presented in the frequency range up to 200 Hz (a) and limited to 50 Hz (b).

IV. STABILITY CONSIDERATIONS

In order to discuss the reasons for the observed conditional stability of the equal gain controller it is convenient to formulate a model of plate and actuator system in the Laplace domain using impedance/mobility methods. The vector of velocities at the control locations may be described in terms of the plate mobility as

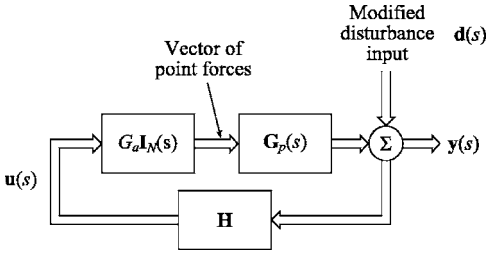


FIG. 7. Schematic of control model derived by impedance/mobility analysis.

$$\mathbf{y}(s) = \mathbf{y_d}(s) + \mathbf{\Gamma}_p(s)\mathbf{f}_t(s), \tag{29}$$

where $\mathbf{y_d}(s)$ is the Laplace transform of the vector of plate velocities at the control locations due to the disturbance input $\mathbf{d}(t)$, $\mathbf{\Gamma}_p(s)$ is the plate mobility matrix, and $\mathbf{f}_t(s)$ is the Laplace transform of the vector of forces acting on the plate due to the actuators $\mathbf{f}_t(t)$. This vector may be expressed in terms of the passive and active response of the actuator, thus

$$\mathbf{f}_t(s) = \mathbf{G}_a(s)\mathbf{u}(s) - \mathbf{Z}_a(s)\mathbf{y}(s), \tag{30}$$

where $\mathbf{G}_a(s)$ is the transfer function matrix of the actuators, $\mathbf{u}(s)$ is the Laplace transform of the control signal vector $\mathbf{u}(t)$, and $\mathbf{Z}_a(s)$ is the actuator input impedance matrix. In the calculation of $\mathbf{G}_a(s)$ and $\mathbf{Z}_a(s)$ the actuators are modeled as previously including the dynamics of the actuator fixative. Note that in the case of identical actuators, $\mathbf{G}_a(s)$ and $\mathbf{Z}_a(s)$ reduce to the scalar transfer functions multiplied by identity matrices $G_a(s)\mathbf{I}_N$ and $Z_a(s)\mathbf{I}_N$. Combining Eqs. (29) and (30) then yields

$$\mathbf{y}(s) = [\mathbf{I}_N - Z_a(s)\mathbf{\Gamma}_p(s)]^{-1}\mathbf{y_d}(s) + G_a(s)[\mathbf{I}_N - Z_a(s)\mathbf{\Gamma}_p(s)]^{-1}\mathbf{\Gamma}_p\mathbf{u}(s). \tag{31}$$

The feedback control system may thus be depicted as in Fig. 7, where

$$\mathbf{G}_p(s) = [\mathbf{I}_N - Z_a(s)\mathbf{\Gamma}_p(s)]^{-1}\mathbf{\Gamma}_p, \tag{32}$$

and

$$\mathbf{d}(s) = [\mathbf{I}_N - Z_a\mathbf{\Gamma}_p]^{-1}\mathbf{y_d}(s). \tag{33}$$

Note that in this model the input to $\mathbf{G}_p(s)$ is a vector of point forces acting on the entirely passive system consisting of the plate and undriven actuators and the output is a vector of collocated velocities. As such, if $\mathbf{G}_p(s)$ were to be the plant response in a decentralized control system, the system would be unconditionally stable. The instability observed previously must therefore be dictated by the actuator transfer function $G_a(s)$. The frequency response of a single actuator, $G_a(j\omega)$, is plotted in Fig. 8. It is clear that at frequencies above the natural frequency, at approximately 12 Hz, the transfer function approaches unity. Therefore the possibility of instability reduces with increasing frequency. The tendency of the system to go unstable then is due to the response of the actuator at low frequencies. In particular, this instability is due to the phase shift that occurs at the actuator resonance frequency.

In the single channel case the stability of the system may be investigated by plotting the frequency response function

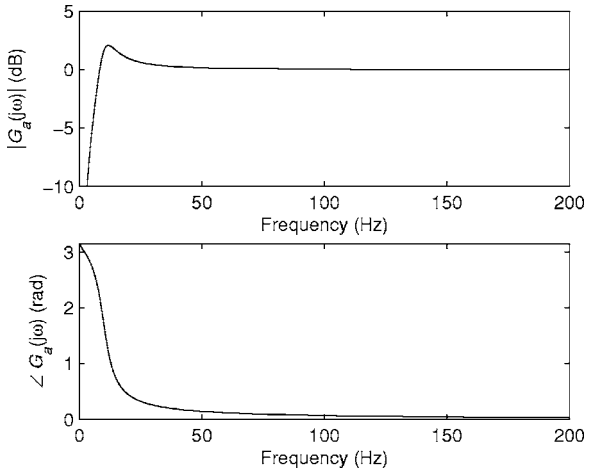


FIG. 8. Actuator frequency response function Bode plot.

$G_p(j\omega)G_a(j\omega)$ in the Nyquist plane as in Fig. 9. It is clear from this plot that the resonance associated with the actuator proof mass causes a loop on the left-hand side of the Nyquist plane and thus the gain margin of the system is finite. Not only does this place a limit on the maximum permissible gain in an output feedback configuration before instability, but also has important implications in extending the system to multiple channels.

In closing the feedback loop, the plant response at frequencies for which the open loop transfer function satisfies

$$|1 + G_a(j\omega)G_p(j\omega)h| < 1, \tag{34}$$

will not be attenuated but enhanced. In terms of the Nyquist plot of Fig. 9, any part of the locus which is inside the unit circle centered on the point −1, 0 will be enhanced by closing the loop. This has important implications when extending the system to multiple channels since subsequent channels will not only see a plant with a greater magnitude response at the actuator resonance but will also act to enhance the response seen by the existing actuators. This coupling of the actuators via the structure must be considered in the implementation of controllers using inertial actuators as it is not enough to assume that because each individual channel is

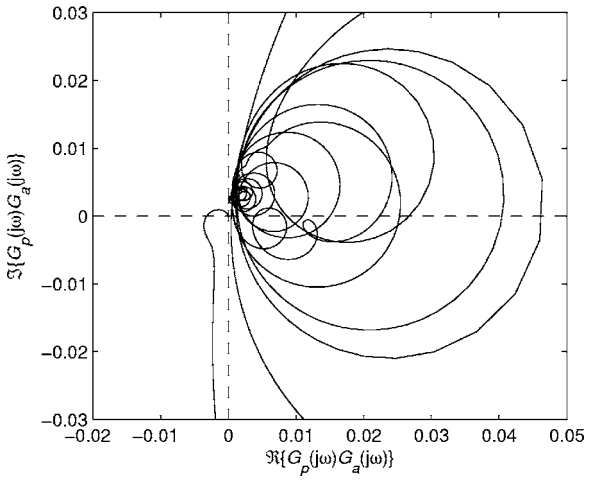


FIG. 9. Nyquist plot of the open loop transfer function for a single channel controller.

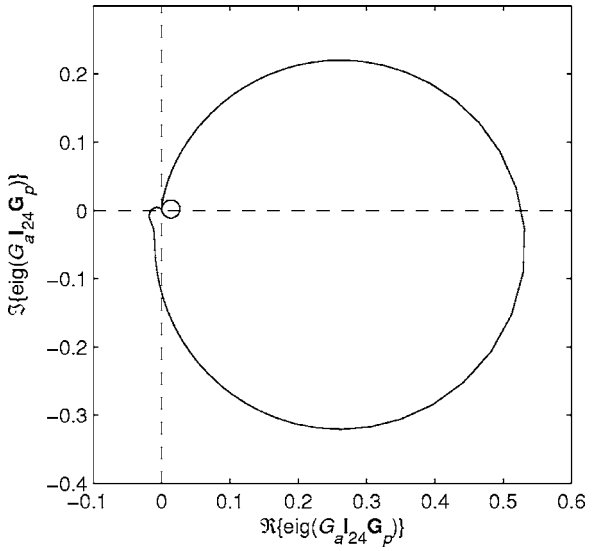


FIG. 10. A plot of the largest magnitude eigenvalue locus for the 24 channel system.

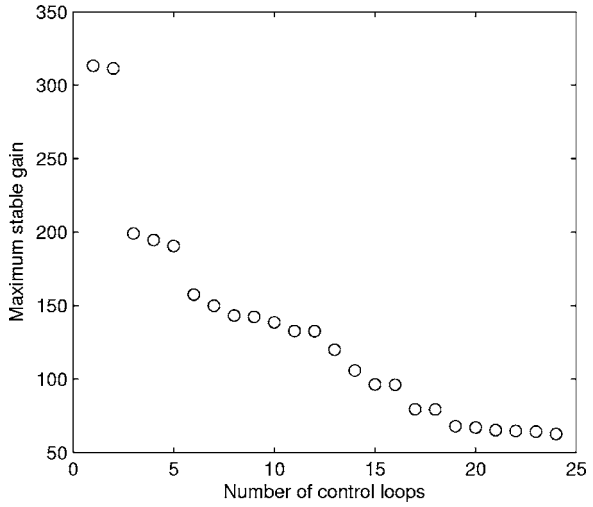


FIG. 11. Plot of the maximum stable gain of the equal gain controller as a function of increasing numbers of actuators, actuator locations are chosen to be a subset of those presented in Fig. 4.

stable for a gain h, the equal gain multichannel system will be stable with each loop taking the gain h. Indeed the gain margin and therefore the maximum stable gain of the equal gain multichannel system may be obtained using the generalized Nyquist criterion. First, an eigenvalue analysis of the open loop system is performed resulting in a set of eigenvalues and eigenvectors for each frequency. These are then sorted based upon the ordering of the eigenvectors[14] such that the eigenvalues represent loci which evolve with increasing frequency much as the scalar transfer function in the single channel example. This is shown for the largest eigenvalue of the 24 channel system in Fig. 10. The frequency at which the locus of the largest eigenvalue crosses the negative real axis, 10.5 Hz, is the frequency at which the greatest increase in the plate kinetic energy is observed in Fig. 6 and is also close to the natural frequency of the actuator. The gain margin of the equal gain controller is then assessed by taking the reciprocal of the absolute distance between this largest eigenvalue locus at the negative real axis and the $-1,0$ point. By performing the gain margin calculation for a unit controller one obtains the maximum stable gain h for the controller of the form given by Eq. (28). In this way we may plot the maximum stable gain of the multichannel equal gain controller as a function of the number of control loops, as shown in Fig. 11. The sensor/actuator locations used for each datum are based on an increasing subset of the locations described in Fig. 4 each addition to which is chosen randomly. It is clear from Fig. 11 that as more control loops are added to the system, the gain margin decreases significantly. In order to see how this reduction in relative stability affects the performance of the controller the reduction in the total kinetic energy of the plate is calculated for a controller with a 6 dB gain margin. The gain used in each controller is calculated by halving the maximum stable gain of each controller presented in Fig. 11. The performance for a fixed gain margin is plotted as a function of the number of actuators in Fig. 12. The two data sets correspond to the total effect of the control system including the change in kinetic energy due to the passive effect of the actuators and the change due to feedback alone. Interestingly we observe that as the number of actuators is increased above 12, the additional reduction in kinetic energy due to closing the control loops is actually reduced. We note from the discussion of the previous section, however, that by optimizing the gains of the decentralized controller the performance with many channels could be improved over the equal gain controller without compromising stability.

The stability of the multichannel control system, then, is dependent on the loop of the largest eigenvalue loci which appears on the left-hand side of the generalized Nyquist plane. The size and orientation of this loop is described by the product of the actuator dynamics and plate response in the frequency range of the actuator resonance. Simulations suggest that reducing either the plate response in this region by, for example, increasing the separation of the actuator resonance frequency and the natural frequency of the first

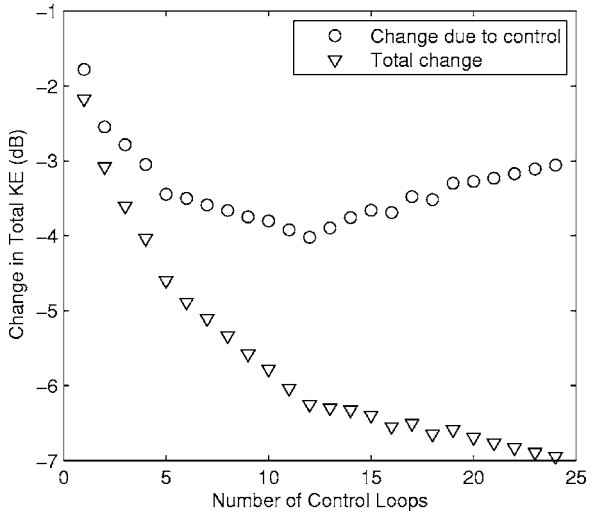


FIG. 12. Plot of the change in total kinetic energy of the plate due to active control (○) and the total change due to both passive and active effects (△).

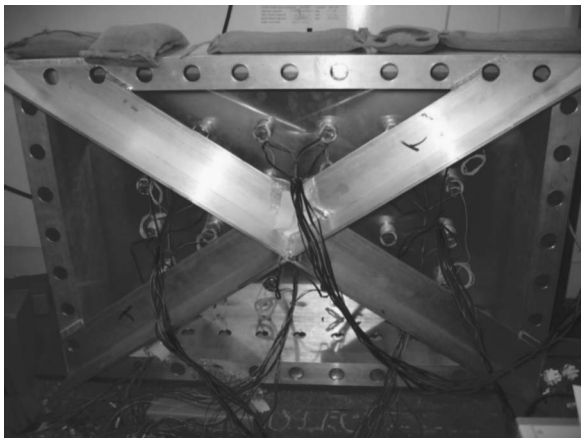


FIG. 13. Image of plate and supporting surround.

mode of the plate or increasing the mechanical damping of the proof mass suspension can significantly increase the performance to relative stability ratio.

V. EXPERIMENTAL SETUP AND RESULTS

This section discusses the design and implementation of a 24 channel decentralized velocity feedback controller using inertial actuators for the reduction of vibration of a plate.

The plate was clamped by means of an aluminum box section surround. This was designed with the assistance of a finite element package which was used to ensure that the natural modes of the surround were of significantly higher frequency than the highest plate mode of interest. The plate surround was then mounted on a large concrete block to eliminate any low frequency, whole body vibration of the structure. Figure 13 shows a photograph of the plate and surround. The inertial actuators are commercially available[15] and were attached to the plate using cyanoacrelate glue with a layer of adhesive copper tape to protect the actuator surface. They are driven with a multichannel constant current amplifier. Velocity signals were obtained by integrating the output of a Micro-Electro-Mechanical Systems (MEMS) accelerometer with an analog integrator circuit consisting of a first-order low pass filter with a very low frequency pole providing a magnitude response which falls as $1/\omega$ and $-\pi/2$ rad phase shift over the frequency range of interest. A first-order high pass filter with a cut-on frequency of 0.3 Hz was included to avoid very high gains and saturation of the electronics at low frequencies. The primary disturbance was achieved by means of an array of 16 loudspeakers each driven by a separate noise source such that the sound field acting on the plate is temporally and spatially uncorrelated.

The kinetic energy of the plate is approximated as the sum of the velocity power spectra at the control locations which, it can be demonstrated, is valid within the frequency range of interest. Figure 14 shows the approximation to the kinetic energy over two frequency ranges for the system without control and under control of a decentralized equal gain controller with five different gain levels. Comparison of these plots with those of Fig. 6 shows relatively good correlation, particularly in the resonant frequencies of the first few plate modes. The increase in general level at low frequencies

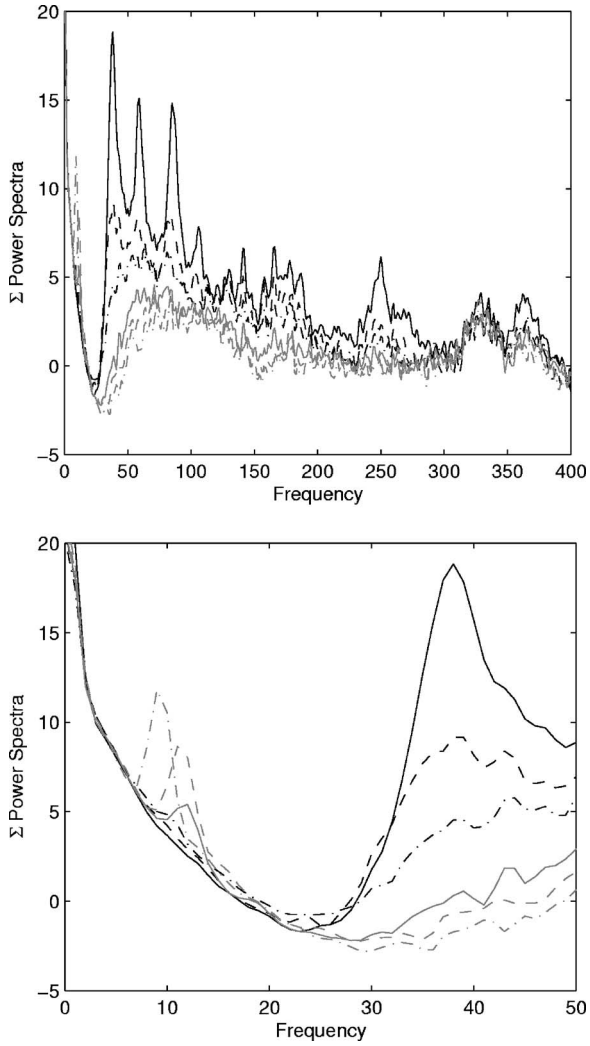


FIG. 14. Experimental results of the approximated kinetic energy over two frequency ranges without control (—) and for increasing control gains from lowest to highest: (– – –), (–·–), (—), (- - -), and (···).

not present in the simulations is attributed to the high gain of the integrator circuit amplifying the sensor noise. Of importance to this paper is the fact that as the gain of the controller is increased the kinetic energy associated with the modes of the plate is significantly reduced, up to 18 dB at the first resonance, whereas the kinetic energy at the actuator resonance is increased. This increase in the plate response results in instability if the gain is increased slightly above the highest value used to obtain the presented results. The instability primarily results in the stroke saturation of the actuator proof mass.

VI. CONCLUSIONS

This paper has introduced a time domain model of a plate structure with multiple inertial actuators. This model was then used to optimize the on-diagonal gains of a fixed feedback gain matrix for various control effort weightings. It was demonstrated that the performance of these optimal controllers for a given control effort is no better than the most simple decentralized controller in which each loop has an identical gain. It was observed, however, that as the control

effort and hence the gain of the controller is increased, systems with the equal gain controller become unstable. The mobility analysis of a single channel controller with an inertial actuator reveals that the controller tends to increase the response of the structure at frequencies around that of the proof mass resonance eventually going unstable. The implication of this in extending the system to multiple control loops, namely the coupling of proof masses via the structure, was highlighted and results demonstrating that as the number of control loops is increased the relative stability of the controller is reduced were presented. It was also shown that for the presented model the additional performance offered by closing the feedback loops of the multichannel controller actually decreased as the number of control loops was increased above 12. Finally the results of an implementation of a 24 channel decentralized controller were presented and the increase in the response of the plate at the actuator resonance with increasing control gain was observed.

ACKNOWLEDGMENT

The work of O.B. is supported by the Data and Information Fusion Defence Technology Center.

[1]M. J. Balas, "Feedback control of flexible systems," IEEE Trans. Autom. Control **AC-23**, 673–674 (1978).
[2]A. Preumont, *Vibration Control of Active Structures*, Solid Mechanics and its Applications, Vol. **50**, 1st ed. (Kluwer Academic, Ontario, 1997).
[3]S. J. Elliott, P. Gardonio, T. C. Sors, and M. J. Brennan, "Active vibroacoustic control with multiple local feedback loops," J. Acoust. Soc. Am. **111**, 908–915 (2002).
[4]P. Gardonio, E. Bianchi, and S. J. Elliott, "Smart panel with multiple decentralized units for the control of sound transmission. II. Design of the decentralised control units," J. Sound Vib. **274**, 193–213 (2004).
[5]L. Benassi, S. J. Elliott, and P. Gardonio, "Active vibration isolation using an inertial actuator with local force feedback control," J. Sound Vib. **276**, 157–179 (2004).
[6]C. Paulitsch, P. Gardonio, and S. J. Elliott, "Active vibration control using an inertial actuator with internal damping," J. Acoust. Soc. Am. **119**, 2131–2140 (2006).
[7]S. Griffin, S. A. Lane, C. Hansen, and B. Cazzolato, "Active structural-acoustic control of a rocket fairing using proof mass actuators," J. Spacecr. Rockets **38**, 219–225 (2001).
[8]W. P. Engels, O. N. Baumann, S. J. Elliott, and R. Fraanje, "Centralized and decentralized control of structural vibration and sound radiation," J. Acoust. Soc. Am. **119**, 1487–1495 (2006).
[9]J. Q. Sun, "Some observations on physical duality and colocation of structural control sensors and actuators," J. Sound Vib. **194**, 765–770 (1996).
[10]L. Meirovitch, *Dynamics and Control of Structures* (Wiley, New York, 1990).
[11]R. E. Kalman and J. E. Bertram, "Control system analysis and design via the second method of Lyapunov I. Continuous-time systems," J. Basic Eng. **82**, 371–393 (1960).
[12]W. S. Levine and M. Athans, "On the determination of the optimal constant output feedback gains for linear multivariable systems," IEEE Trans. Autom. Control **AC-15**, 44–48 (1970).
[13]W. P. Engels and S. J. Elliott, "Optimal velocity feedback control on a beam," in *Institute of Acoustics Spring Conference* (Institute of Acoustics, St. Albans, UK, 2004), Vol. **62**, Pt. 2, pp. 579–589.
[14]X. Huang, S. Elliott, and M. Brennan, "Active isolation of a flexible structure from base vibration," J. Sound Vib. **263**, 357–376 (2003).
[15]Micromega Dynamics (2006). Homepage. Available from http://www.micromega-dynamics.com/[15 June 2006].

Survey of noise in coal preparation plants

Jeffrey S. Vipperman[a)]
Department of Mechanical Engineering, University of Pittsburgh, Pittsburgh, Pennsylvania 15261

Eric R. Bauer and Daniel R. Babich
National Institute for Occupational Safety and Health, Pittsburgh Research Laboratory, P.O. Box 18070, 626 Cochrans Mill Road, Pittsburgh, Pennsylvania 15236

(Received 14 July 2005; revised 14 August 2006; accepted 1 September 2006)

In response to the continuing problem of noise induced hearing loss (NIHL) among mine workers, the National Institute for Occupational Safety and Health (NIOSH) has conducted numerous noise surveys in coal preparation plants. The research, consisting of worker dose monitoring, task observations, and equipment noise profiling, was completed in eight separate preparation plants. Worker dose monitoring was conducted for three shifts in most cases. Workers experiencing higher than allowable doses were task-observed for one full shift to correlate dose to noise source(s). Finally, noise levels on all floors, and in lunch rooms and control rooms, were characterized. Results indicate that only workers who routinely spend a significant portion of their shift in the plants (away from the control rooms) are susceptible to overexposure from noise. Certain pieces of equipment (screens, centrifuges, sieve bends) are the loudest primary noise sources responsible for the worker noise exposures. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2372587]

PACS number(s): 43.50.Jh, 43.50.Gf, 43.50.Qp [DKW] Pages: 197–205

I. INTRODUCTION

Prolonged exposure to noise over a period of several years can cause noise induced hearing loss (NIHL). While NIHL is the most common occupational disease in the United States today, with 30 million workers exposed to excessive noise levels,[1] the problem is particularly severe in all areas of mining (surface, processing plants, and underground). An early NIOSH analysis of NIHL in miners revealed an alarming prevalence of severe hearing loss.[2] For example, by age 60, over 70% of miners had a hearing loss of more than 25 dB, and about 25% had a hearing loss of more than 40 dB. A more recent, 1996 analysis of NIHL in miners, also performed by NIOSH, showed an apparent worsening of NIHL.[3] This recent analysis of a private company's 20 022 audiograms indicated that the number of miners with hearing impairment increased exponentially with age until age 50, at which time 90% of the miners had a hearing impairment.[3] In addition to government researchers, academics have reported that the "…policies and practices for preventing occupational hearing loss among miners are inadequate …".[4]

The Federal Coal Mine Safety and Health Act of 1969 established requirements for protecting coal miners from excessive noise and subsequently, the Federal Mine Safety and Health Act of 1977 broadened the scope to include all miners, regardless of mineral type.[5] Data from more than 60 000 full shift Mine Safety and Health Administration (MSHA) noise surveys show that the noise exposure of selected mining occupations has decreased since the 1970s, although the percentage of miners considered overexposed under current MSHA noise regulations remains high.[6] MSHA found that the percentage of coal miners with noise exposures exceeding federal regulations, and unadjusted for the wearing of hearing protection, was 26.5% and 21.6% for surface and underground mining, respectively.

Despite the extensive work that ensued in the 1970s and 1980s, NIHL is still a pervasive problem, as outlined in Ref. 7. MSHA has published new Noise Health Standards for Mining.[8] One of the changes is the adoption of a provision similar to OSHA's Hearing Conservation Amendment. MSHA concluded in a survey that if an OSHA-like hearing conservation program were adopted, hypothetically, 78% of the coal miners surveyed would be required to be in a hearing conservation program.[9] Other requirements of the new regulations are a Permissible Exposure Level (PEL) of 90 dBA $L_{TWA(8)}$, no credit for the use of personal hearing protection, and the primacy of engineering and administrative controls for noise exposure reduction.

Improvements in both mining and noise monitoring equipment necessitate that new data be taken in order to base noise control decision making. In many cases, the existing data are specific to machine type and were obtained for characterizing noise source sound power rather than exposure assessment. There is also a great range in noise levels for a given occupation. For example, noise levels for continuous mining machine operators have a modal value of 90 dBA and a range that varies from 80 to 105 dBA.[10] Yet, at present, there is insufficient information to explain this variation in exposure for this and other mining occupations. Specifically, noise level data are needed that provide a time exposure history for workers in addition to information on noise sources. Such information will provide the basis for targeting and selecting engineering controls, in combination with administrative controls and personal protection equipment, to reduce noise exposures among the mining workforce.

[a)]Electronic mail: jsv@pitt.edu

This study presents a survey and assessment of noise exposures in eight coal preparation plants from five states. Results concentrate on the sources of noise exposure, both in terms of occupations and plant location. Generalizations about the extent of the noise problem in preparation plants as well as the noise generating potential of the cleaning equipment are drawn.

II. SURVEY METHODS

A four-phase approach was taken to analyze noise in coal preparation plants, which included worker dose monitoring, task observations, equipment noise profiling, and reverberation time measurements. Each task is described in greater detail in the following sections, along with information on plant selection.

A. Plant selection

The number of coal preparation plants in the U.S. totaled 212 for the last census in 2000.[11] The average plant capacity is 1000 tons per hour (TPH), with 57 plants exceeding 1000 TPH, and 10 exceeding 2000 TPH. A sample of eight preparation plants with capacities between 300–2000 TPH was chosen for this study. The plants were located across the states of Pennsylvania, Kentucky, Virginia, Illinois, and West Virginia. They are owned by five different major coal companies that are members of the Noise Partnership, a consortium that was formed to reduce NIHL in the mining industry. The partnership includes members from government (NIOSH), regulation (MSHA), industry [Bituminous Coal Operators Association (BCOA)], labor [United Mine Workers of America (UMWA)], and professional associations. Participation in the surveys was voluntary for the plants, but 100% of the plants contacted participated.

B. Equipment and process

Nearly all coal preparation plants employ the same basic coal cleaning processes of crushing, screening, separation, flotation, and dewatering (drying). In general, the raw feed enters the plant on one of the upper floors, and flows downward by gravity through the cleaning process. Eventually, the clean coal and waste are deposited on separate belts on one of the lower floors for removal to the clean coal storage area and refuse area, respectively. Finally, the bottom floor usually contains the sumps and pumps used to recycle water and separation media back to the appropriate equipment on the upper floors. Specific equipment found in most of the plants included sieve bends, magnetic separators, froth flotation cells, banana screens, drain and rinse screens, deslime screens, heavy media cyclones, coal spirals, centrifuges, clean coal and refuse conveyors, crushers, heavy media vessels, vacuum filters, and pumps. In addition, since compressed air is used in the cleaning process, the plants have compressor rooms.

C. Dosimetry

Quest model Q-400 dosimeters were mounted on workers just prior to the start of the shift. Microphones were installed at mid-shoulder as recommended by MSHA,[12] and contained a wind screen. The Q-400 is a type II instrument and has two independent dosimeters that were programmed to log A-weighted $L_{AV}, L_{\mathrm{MAX}}, L_{\mathrm{MIN}}$, and L_{PK} (Ref. 13) at 10-s intervals. One dosimeter was set to record the MSHA permissible exposure limit (PEL) and the other measured the 8-h, A-weighted, equivalent sound level, L_{Aeq8}, which could be used for subsequent analyses. $L_{TWA(8)}$ represents the A-weighted constant sound level that over 8 h would result in the recorded shift dose.[13] The MSHA PEL uses a 5 dB exchange rate (ER), a 90 dBA criterion level (L_C), a 90 dBA threshold (L_{TH}) level, and an 8 h criterion time, T_C. Thus, a full day's exposure (100% dose) would occur if a worker is exposed to a steady noise at the 90 dBA criterion level for 8 h. Mean shift time, $\bar{T}$, and dose, $\bar{D}$, were computed for the measured data. The mean $L_{TWA(8)}$ was computed from the average dose as

$$\overline{L_{TWA(8)}} = L_C + Q \log\left(\frac{\bar{D}}{100}\right), \tag{1}$$

where $L_C = 90$ dBA is the criterion level, $Q = \log(ER)/\log(2)$ or 16.61 for the 5 dB ER. Since shift times varied from plant to plant, the average L_{AV} was also computed as

$$\overline{L_{AV}} = L_C + Q \log\left(\frac{8\bar{D}}{100\bar{T}}\right). \tag{2}$$

The above means were used to compute following standard deviations: $S_{L_{TWA(8)}}, S_T$, and $S_{L_{AV}}$.

Asymmetric standard deviations for dose were computed from $\overline{L_{TWA(8)}}$ values as

$$S_D^{\pm} = \left(\frac{100\bar{T}}{T_C}\right) 10^{(\overline{L_{TWA(8)}} \pm S_{L_{TWA(8)}} - L_C)/Q}. \tag{3}$$

This unconventional practice prevented the dose dispersions from being negative, as would occur if computed in the usual way. The M samples of L_{AV} data were also postprocessed to produce cumulative dose plots for each shift as

$$D(k) = 100 \frac{T_S}{T_C} \sum_{i=1}^{k} 10^{(L_{AV}(i) - L_C)/Q}, \tag{4}$$

where $T_S = 10$ s is the sampling time, and i, k are discrete time indices.

D. Task observations

Detailed time-at-task observations were conducted for each job classification. Note that for most jobs, the worker is required to move throughout the plant performing various tasks. A crew of technicians and mining engineers monitored three shifts for each job classification. As the worker moved or changed activities, the shift time and activity were logged, along with any comments from the observer. The observation resolution can be as fine as 10 s, as with the dosimeter logging. These observations were later used to determine the noise exposures for certain activities as well as exposures within the plant.

E. Equipment noise profiling

Sound level measurements (L_{eq}) were made throughout the plant in order to determine the distributed noise levels. A measurement grid was established, with measurements occurring approximately every 1.5–6 m, depending upon the density of the noise sources within the floor. Two different instruments were used to measure data: Quest model 2900 (type II) sound level meters (SLM) and/or a Brüel and Kjær (B&K) 2260 investigator (type I). Measurements typically included *A*-weighted, *C*-weighted, and linear overall sound levels and linear 1/3-octave band levels, which were *A*-weighted in postprocessing. The instruments were mounted on a tripod, with the microphones 1.5 m from the floor (approximate ear height), angled at 70° from horizontal, and facing the noise source as per the manufacturer's recommendations. A slow response time with an averaging time of 30 s was also employed. The instruments were field-calibrated at the start and end of each shift and undergo annual NIST-traceable calibrations. Postprocessing included spatially interpolating linear sound pressures to create noise contour plots, and determining the average L_{eq} in the vicinity of specific types of equipment. Sound levels were taken from grid points nearest to the equipment, typically within a few meters. Average values were computed from linear L_{eq} quantities as was done in Eq. (2). Standard deviations are computed from the dBA levels.

F. Reverberation time measurements

Reverberation time measurements were conducted in two typical plants while they were shut down for extended maintenance. The measurements could be used to determine whether treating walls of the plant with absorptive materials could be expected to significantly reduce the noise levels in the plant. Tests were completely automated in accordance with the ISO-140 standard[14] by the BZ7404 building acoustics software package installed on the B&K 2260 Investigator. A B&K 2716 bridging amplifier drove the B&K 4296 omnidirectional sound source in order to generate high levels of pseudorandom noise at one location of the plant floor. At another floor location, the B&K 2260 Investigator measured the decay in all 1/3-octave bands after abruptly switching off the noise source. The instrument measures either T_{20} or T_{30} reverberation times (time for the noise to decay 20 or 30 dB, respectively, in each 1/3-octave band) and extrapolates to the more characteristic T_{60} reverberation time (time to decay by 60 dB). Three combinations of source/receiver locations were measured on each floor where either the source or the receiver was moved between measurements.[14] At each source/receiver location a total of 5 decay times were measured and averaged together by the instrument to find the average T_{60} time.

The Sabine formula[15] was used to compute the overall change in reverberation times when absorptive treatments covered varying percentages of the walls. These were converted to the Sabine absorption coefficient, α, by also taking into consideration the room volumes and surface areas. Next the room constants, both with (R_1) and without (R_0) acoustical treatments, were computed from the absorption coefficients. Assuming distances far from the source, the theoretical maximum reduction in sound pressure was computed as

$$\Delta L_P \approx -10 \log\left(\frac{R_1}{R_0}\right). \tag{5}$$

TABLE I. Characteristics of coal preparation plants surveyed.

Plant No.	State	Raw feed (TPH[a])	Reject (%)	No. of floors	Total floor area, (ft^2)
1	PA	1000	31	8	58 200
2	WV	2000	52	6	81 850
3	PA	2000	33	12	142 650
4	VA	1000	27	7	63 350
5	PA	1000	21	10	51 400
6	KY	300	38	5	15 505
7	KY	2000	51	5	75 084
8	IL	1500	38	5.5	31 582

[a]Tons per hour.

III. SURVEY RESULTS

A. Coal preparation plant characteristics

All the plants were of similar construction; steel I-beams covered by either single ply corrugated steel sheeting or two-ply sheeting with insulation in between the plies, and concrete or steel grating floors. Table I summarizes the specific characteristics of each plant, including raw feed capacity, reject (rock), number of floors, and total square footage. Reject is potentially important since rock is harder than coal and thus typically produces more noise.

B. Worker classifications

In general there are four specific job classifications. The first class of worker is the "control room operator," who is positioned in the enclosed control room and is surrounded by video monitors and alarms. Typically, this person remains in the control room the entire shift. A second classification is a "mechanic/electrician." This worker is responsible for equipment maintenance throughout the plant and surrounding facilities. His time in the plant varies depending on plant maintenance needs during a particular shift. The worker that typically spends the entire shift, except for lunch and breaks, in the plant is the "plant attendant." This worker's duties can be located throughout the entire plant or on specific floors. This person is generally responsible for checking equipment operation, making measurements of process variables, and cleaning. The fourth job class is the "utility worker" whose job is to help the plant attendants or mechanic/electricians. Effort was made to determine the total number of workers in each occupation class for the eight plants, and is presented in Table II. Note that the same workers may have been surveyed across different shifts. Also, if a sample was considered atypical (e.g., partial shift monitoring), it was not included in the analysis.

TABLE II. Plant jobs surveyed by plant.

Occupation	Number of employees (number surveyed)								
	Plant 1	Plant 2	Plant 3	Plant 4	Plant 5	Plant 6	Plant 7	Plant 8	Total
Control Rm. Oper.	3(0)	3(1)	4(1)	3(0)	4(1)	3(2)	3(1)	2(2)	25(8)
Elec./Mech.	6(0)	4(0)	17(3)	9(0)	21(0)	1(0)	3(0)	2(0)	63(3)
Utility Man	9(0)	3(0)	9(4)	9(0)	8(1)	0(0)	1(0)	2(3)	41(8)
Plant attendants	3(1)	9(3)	3(3)	3(2)	4(3)	3(2)	19(7)	2(4)	46(25)
Total Workers	21(1)	19(4)	33(11)	24(2)	37(5)	7(4)	26(8)	8(9)	175(44)

C. Worker noise exposure

Table III provides a statistical summary for three of the job categories, including the number of samples, N, mean dose, and asymmetric standard deviation of dose. The Mechanic/Electrician job class did not have enough samples ($N=3$) to be statistically relevant, and hence statistical data are not reported for this class.

D. Noise profiles

The noise was categorized as continuous with high overall levels generally experienced throughout the plant. Figure 1 presents a noise contour plot of the overall A-weighted L_{eq} measurements from a representative floor. Levels are consistently high and range from 95 to 100 dBA, as indicated by the scale to the right of the figure. Figure 2 is an example of a much larger floor, with varying-noise levels (87–104 dBA) and a specific area of higher noise caused by a group of clean-coal screen-bowl centrifuges. Each plant typically had one or two floors with more modest levels (90 dBA) of noise. A fairly typical 1/3-octave band spectrum of the noise in the vicinity (within 1–2 m) of a raw coal screen is shown in Fig. 3, where the linearly weighted data are represented by the black bars and the A-weighted data are represented in grey. Screens are found to have very relatively high sound levels.

Figure 4 provides information on the average equivalent sound levels recorded in the vicinity of various pieces of equipment surveyed in the plants. They are sorted from quietest to loudest. The equipment type and total number of unique locations within the eight plants where each type of equipment was found (N) are given in the x axis labels. Standard deviations are also included on the plot where appropriate.

E. Dose/source relationships

The logged L_{AV} values recorded by the dosimeter were converted to a cumulative dose plot using Eq. (4). The resulting plots were then annotated using the information ob-tained from the task observations. A representative cumulative dose plot for a utility man serving as a plant attendant is shown in Fig. 5 and a cumulative dose plot for a mechanic is provided in Fig. 6. No equivalent plot is given for the control room operator, since their exposures are generally small.

F. Reverberation time and noise control

The measured areas and volumes for the six measured floors are given in Table IV. Each floor varies in the number and type of equipment, as well as room surface area and overall volume. It is typical for floor plans to get smaller with respect to the building height. Also given in the last column of Table IV, is the range for the average absorption coefficient for each floor in the bandwidth of 250–4000 Hz. Figure 7 graphically depicts the T_{60} and absorption coefficient calculations across the 1/3 octave bands for one of the floors.

Sample calculations based on Eq. (5) were made to determine the potential effectiveness of adding absorptive materials to the walls of the six measured plant floors. The results were predicted from applying an EAR E-100SM aluminum-faced, urethane-foam, absorbing-material to various percentages (15%, 40%, and 60%) of floor surface areas. The absorption coefficients for the material are {0.81, 0.61, 0.73, 0.71, 0.69} in the {250, 500, 1000, 2000, 4000} Hz 1/1-octave bands, respectively. In many cases much of the ceiling and wall area was amenable to being covered with material, making the 60% area coverage practical. Predicted reductions range from 0.9 to 10.6 dB across the bands. The reductions were then arithmetically averaged across the 250–4000 Hz 1/1-octave bands and are presented in Fig. 8.

IV. DISCUSSION

Dose and noise survey data from the eight coal preparation plants suggest that overexposures to noise can easily occur, particularly in plant attendants. Plant attendants are found to experience significantly higher exposures than other job categories when examining the data given in Table III. The mean value nearly equals the citable range and one stan-

TABLE III. Data summary for job classes.

Job class	N	$\bar{D}(\%)$	$S_D^+(\%)$	$S_D^-(\%)$	$\bar{T}_i$ (h)	$\overline{L_{TWA(8)}}$ (dBA)	$\overline{L_{AV}}$ (dBA)
Plant attendant	22	129	165	101	8:49	91.1	92.0
Control room operator	7	25.4	60.6	17.9	10:23	80.1	78.2
Utility man	7	66.6	102	39.6	9:31	85.5	85.8

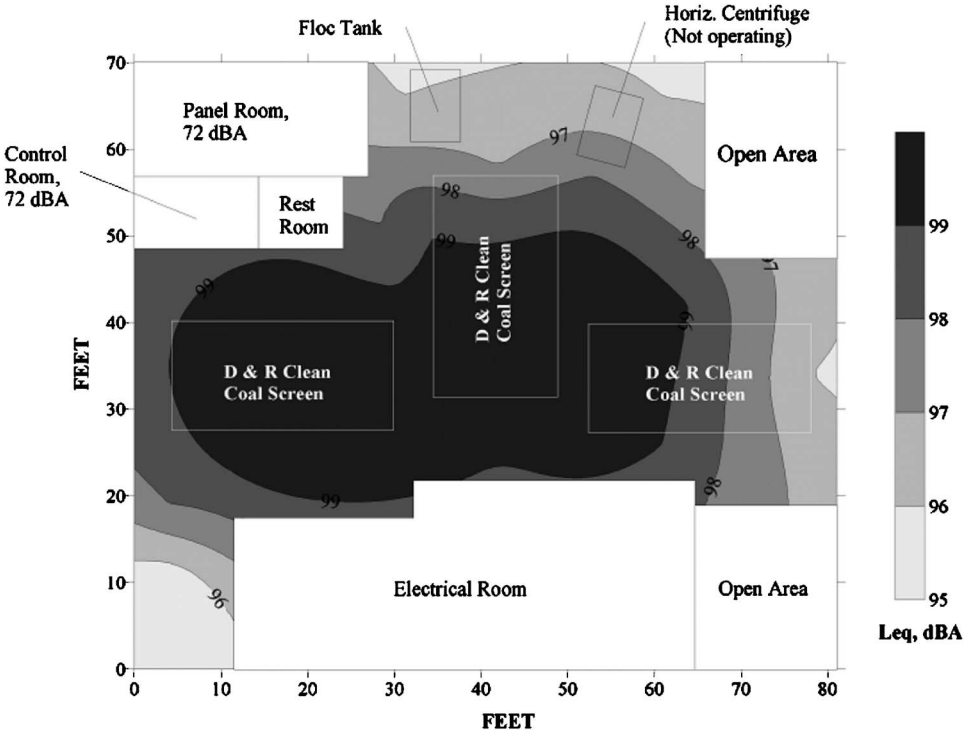


FIG. 1. Example of a high noise floor containing drain and rinse (D&R) screens (Plant 5).

dard deviation is more than double the citable dose. Citations are not issued until the dose exceeds 132%, allowing for the type II accuracy (2 dB) of most noise dosimeters. In contrast, it is found that the control rooms can be considered a "quiet area," having only a 22.4% average dose with a positive standard deviation of 60.6% such that $(\bar{D}+S_D^+)<132\%$. The highest recorded dose for all eight control room operators is 99.6%, which also occurred for the longest shift of 12:26 (hours:minutes). In addition, this particular subject was atypical in that he had responsibilities that required him to periodically leave the control room and work in noisier environments. The doses for all utility men surveyed were all

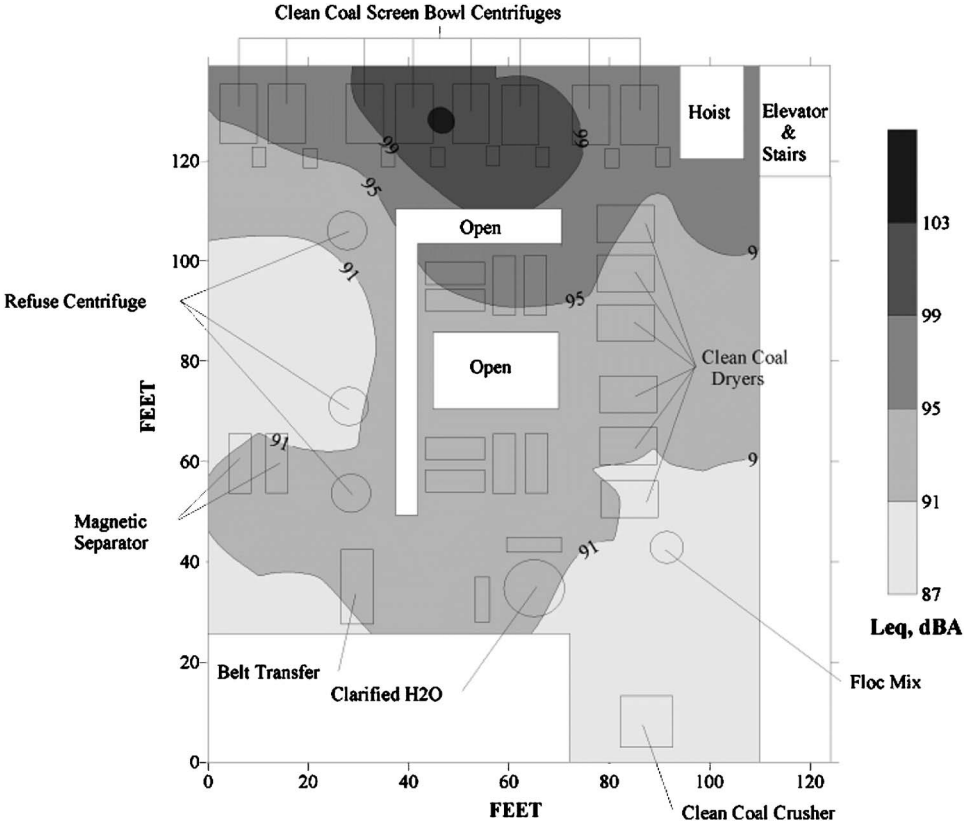


FIG. 2. Example of a varying noise floor.

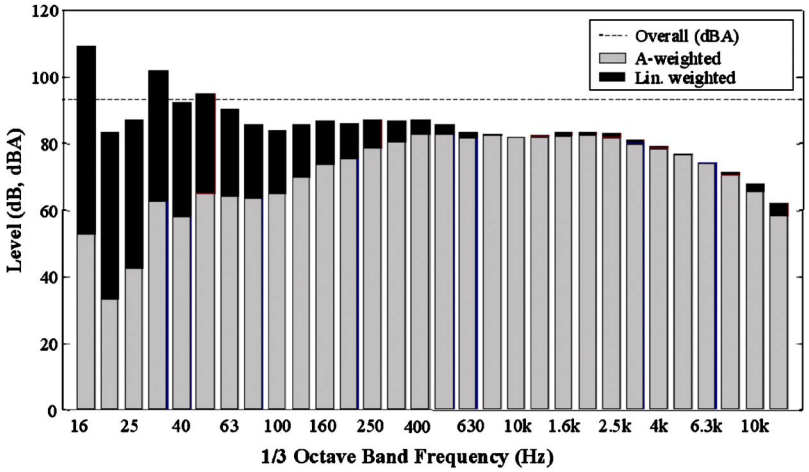


FIG. 3. Linear and *A*-weighted 1/3-octave bands measured near raw coal screens.

below 105% and averaged 63.5%. Doses within a single positive standard deviation are just below the 132% citable level.

Some of the variability in doses was a result of the varying shift durations. Some plants work a standard 8 h shift, while others work a 12 h shift. This fact was reflected in the relatively large standard deviations in shift durations for the control room operator (2:04), plant operator (1:59), and utility man (2:11). The three mechanics surveyed all worked 8 h shifts, producing a relatively low standard deviation of 0:09. The time-weighted averages given in Table III provide a normalized representation of the measured doses, since they were defined as the equivalent continuous 8 h exposure levels that would result in the recorded shift dose. A 100% dose is equivalent to $L_{TWA(8)}=L_C=90$ dBA. L_{AV} measurements are independent of measurement duration and may be compared across work shifts of different lengths. Note that $L_{AV}=L_{TWA(8)}$ for a 100% dose over an 8 h shift and $L_{AV}>L_{TWA(8)}$ for shifts under 8 h and vice versa.

The high variability in the noise dose data ($60.6\% \leqslant S_D^+ \leqslant 411\%$) for each job suggests that the practice of single-shift monitoring is not adequate for characterizing doses for coal preparation plant workers. The variance can be attributed to several factors including plant processes, shift time, plant location, equipment serviced, individual worker traits, production, age of the plant, and plant construction. Another important variable that must be taken into consideration was the quality of plant maintenance (e.g., effects of "squeaky belts"), which was found to vary considerably between plants. Equipment noise levels were found to be approximately 8 dBA higher in two such cases. With the exception of shift time, no attempt was made to quantify the effects of these variables.

The characterization of noise sources in the plants was a complicated task for several reasons. First, the large number of pieces of equipment and their close proximity to each other made separating specific noise sources difficult at times. Next, the openness of the building allowed noise to propagate between floors. Finally, the measured noise came from multiple air-borne and structure-borne paths. Air-borne noise was present as direct noise, generated by the equipment, the process, and motors, and as reflected noise from the reverberant field (building walls, floors, etc.). Structure-borne noise paths resulted from equipment vibration and transfer of that vibration to the buildings structural components, which in turn radiate into the surrounding area. Noise profiles consisting of area noise sampling on noise contour plots for all plant floors revealed areas of high and low noise levels and permitted identification of the equipment generally responsible for the noise. Although levels varied considerably throughout the plant, the patterns of noise in any one location were judged to be stable.

Figure 4 provides information about the range of sound levels within the working area surrounding a particular type

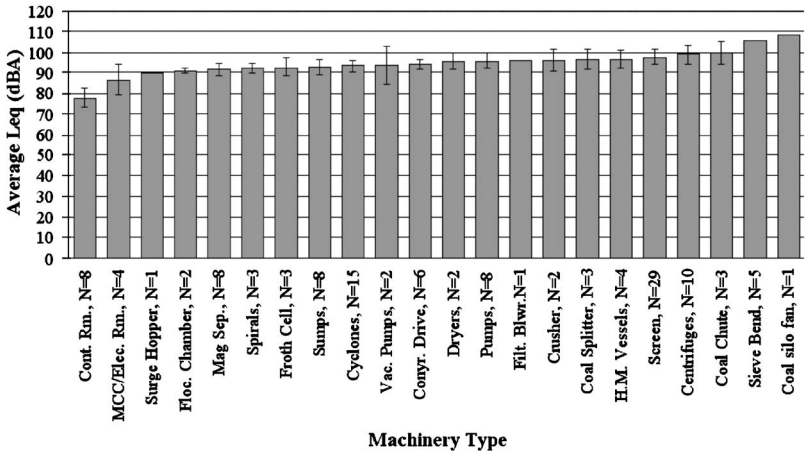


FIG. 4. Average overall equivalent sound levels in the vicinity of various types of mining equipment with 90 percentile confidence bounds shown.

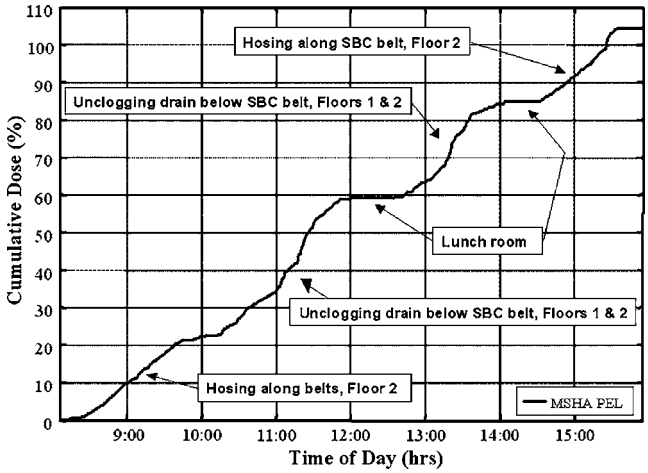


FIG. 5. Cumulative dose plot for utility man.

of equipment. Average equivalent sound levels measured in the vicinity of all pieces of equipment are ≥90 dBA, except the control room and the motor control center/electrical rooms (unmanned). Thus, spending an 8 h shift in close proximity to most all coal processing machines would result in 100% or more dose accumulation. Eleven of the machines produce overall sound levels above 95 dBA and two are above 100 dBA (sieve bends, 106.5 dBA and a coal silo fan, 108.2 dBA), which would accumulate dose at a much greater rate.

Noise contour plots were provided to the management of each plant. This information can be used to develop a plan for engineering or administrative controls that will mitigate or avoid worker exposures to these areas. Figures 1 and 2 illustrated that levels tend to be generally high enough throughout the plant, and that there are a few "quiet" areas within the plants where workers could spend a whole shift without overexposure. The wide range of levels (87–104 dBA) in Fig. 2 and the specific area of higher noise near the centrifuges at the top of the figure illustrate that the noise from one type of equipment can influence the sound field in the adjacent area containing quieter equipment. Where such influence was obvious, data were omitted from the averages in Fig. 4, but otherwise the effects would be difficult to quantify, since the continuous circuit operation of the plants does not permit equipment to be operated individually while loaded with coal. Overall *A*-weighted levels can easily exceed 100–110 dBA in the vicinity of equipment requiring attention, illustrating the importance of properly maintaining equipment.

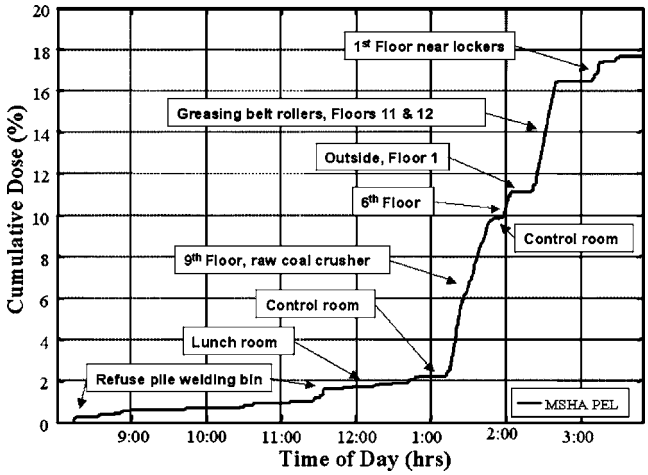


FIG. 6. Cumulative dose plot for mechanic.

TABLE IV. Surface areas, volumes, and average absorption coefficients for each measured floor.

Plant	Floor	Surface area (m^2)	Volume (m^3)	Avg. absorp. coeff. (0.25–4 kHz)
3	5	4400	9400	0.185–0.372
3	8	2300	3900	0.148–0.280
3	11/12	1500	2700	0.150–0.187
5	3	970	1040	0.103–0.197
5	7	1100	1400	0.119–0.182
5	10	1100	1400	0.087–0.141

The typical 1/3-octave band noise spectrum given in Fig. 3 illustrates that the noise was dominant in the low-frequency bands, particularly the 16 and 31.5 Hz 1/3-octave bands, which are the subharmonic and driving frequency of most electric motors throughout the plant. Linearly weighted sound levels on most floors were routinely over 100–110 dB in the 16 and 31.5 Hz bands. While this high amplitude, low-frequency sound can be uncomfortable, it is the *A*-weighted sound that is currently used to predict NIHL.[13,16] The *A*-weighted machinery noise has the highest contributions from 500 to 4000 Hz, which are also the bands that are most contributive to hearing loss.

The cumulative dose plots in Figs. 5 and 6 corroborate that dose is accumulated at the highest rates when working in the vicinity of equipment. No accumulation of dose indicates that the noise levels were below the threshold for the MSHA PEL. The utility man in Fig. 5 spent much of his shift on floors 1 and 2, and received a similar rate of dose while on each floor, albeit at a higher dose while unclogging drains between 11:00 and 11:50 and 1:00 and 1:35. Although the mechanic in Fig. 6 received only an 18% dose, the plot reveals that he received nearly 45% of his dose while repairing the raw coal crusher on floor 9 and approximately 25% while greasing belt rollers on floors 11 and 12. Had he spent more time in these or other loud areas, his shift dose would have been considerably greater.

Reverberation time measurements (Table 4) may also be helpful in determining if treating the entire plant would be helpful for reducing noise levels and exposures. Of the six floors, floor 5 from the first plant (data not presented) had the highest absorption coefficients (0.185–0.372), which was partially attributable to a layer of sand that had accumulated from sandblasting activities. The second plant had smaller absorption coefficients in general ($\bar{\alpha} \in [0.087-0.197]$) than the first plant ($\bar{\alpha} \in [0.148-0.372]$). Floor 10 of the second plant was the acoustically hardest floor, with absorption coefficients ranging from 0.087 to 0.141. The 1/3-octave measured T_{60} and $\bar{\alpha}$ spectrums for this floor are given in Fig. 7. Note from Fig. 7 that the reverberation time decreases and

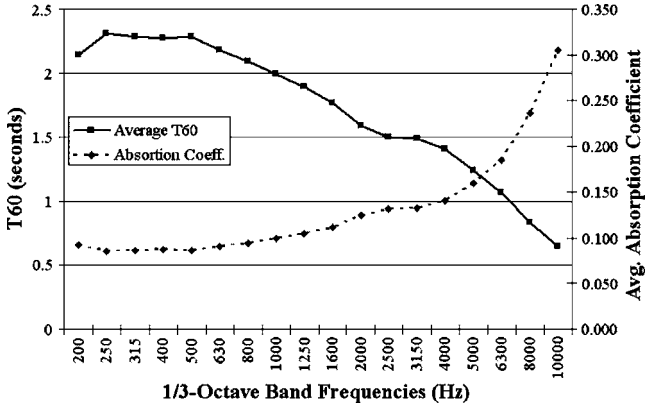


FIG. 7. Example measured T_{60} and computed average absorption coefficient for floor 10 of the second plant where T_{60} measurements were performed.

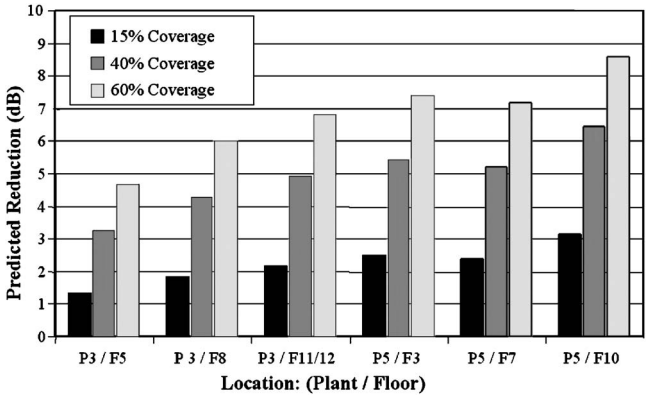


FIG. 8. Average predicted decrease in the reverberant field.

the absorption coefficient increases with frequency, since absorptive materials, whether inherent or introduced, work best at high frequencies.

The plant floors were found to be acoustically "hard" to "medium," and thus some improvements in noise levels were expected with treatments. Figure 8 shows that *predicted* reductions can be as high as 8.5 dB. The large predicted reductions are encouraging, however reductions from adding absorptive materials rarely exceed 3–5 dBA in practice.[17] Better control than expected was predicted at lower octave bands, a result of the good low-frequency performance of the treatment material and relatively small initial absorption coefficients in those bands. Any absorptive materials selected must hold up to harsh, industrial conditions and meet flammability requirements governed by MSHA.

V. CONCLUSIONS

Noise levels and worker noise exposures in eight coal preparation plants were assessed as part of a cross-sectional survey of noise in the mining industry being conducted by NIOSH. Assessment techniques included noise dosimetry, task observations, contour mapping of noise fields, and reverberation time measurements. The maps were provided to plant management to be incorporated into their noise management programs. Overall noise levels were found to range from 75.9 to 115 dBA throughout the plant. The open construction of the plant provided many direct paths for noise to propagate between floors. Most areas of the plant (except control rooms, electrical rooms, and motor control centers) were found to have noise levels in excess of 90 dBA, suggesting the noise overexposure will occur if a full 8 h shift is spent within the plant. Plant maintenance was found to be an important concern, with A-weighted levels increasing by 8 dB in two cases as a result of squeaky belts or bearings. These results are corroborated by the dosimetry and task observations of individual workers, although much higher dose accumulation rates occur when proximal to equipment. For most occupations, the shift-to-shift variability in noise dose is large. Most of the variation is due to differences in exposure levels as the worker moves about the plant or shift lengths (8 h versus 12 h shifts). Depending upon a worker's location, the average dose accumulation rates varied between 0.25–31 %/h with instantaneous values being much higher. Two of the four plant worker categories (plant attendant and mechanic/electrician) had reported shift doses above 132% (the MSHA citable level). The high variability in dose indicates that the current practice of partial or single shift monitoring for compliance may not adequately characterize worker noise exposures. Reverberation time (T_{60}) measurements were performed on six floors from two representative plants. Floors were found to be fairly acoustically "hard," suggesting that absorptive-treatments may provide significant control.

ACKNOWLEDGMENTS

The authors would like to thank the following NIOSH employees who participated in the data collection activities: Ellsworth Spencer, Tom Ozanich, Gerald Morrow, and Mike DiMartino and a very special thanks to Dr. Mary Prince, Mr. Dennis Giardino, and Mr. John Earshen for their helpful suggestions. The findings and conclusions in this report are those of the authors and do not necessarily represent the views of the National Institute for Occupational Safety and Health.

[1]John R. Franks, Mark R. Stephenson, and Carol J. Merry, "Preventing occupational hearing loss - A practical guide," Technical Report No. 96-110, National Institute for Occupational Safety and Health, June 1996.
[2]National Institute for Occupational Safety and Health, Survey of Hearing Loss in the Coal Mining Industry, Publication 76-172, 70 pp., June 1976.
[3]J. R. Franks, "Analysis of audiograms for a large cohort of noise-exposed miners," National Institute for Occupational Safety and Health, Internal Report, Cincinnati, OH, 1996, 7 pp.
[4]J. L. Weeks, "Policies and practices for preventing occupational hearing loss among miners are inadequate," in *Proceedings of the Hearing Conference III/XX*, Cincinnati, OH, March 1995, pp. 16–19.
[5]Federal Register, Code of Federal Regulations (CFR) governing noise exposure in mining, CFR 30, Subchapter O, Part 70, Subpart F: Noise Standards for Underground Coal Mines; Subchapter 0, Part 71, Subpart D: Noise Standards for Surface Work Areas of Underground Coal Mines and Surface Coal Mines; Subchapter O, Part 55, Section 55.5: Metal and Nonmetal Open Pit Mines; Section 56.5: Sand, Gravel, and Crushed Stone Operations and 57.5: Metal and Nonmetal Underground Mines. U.S. Department of Labor, 1997.
[6]J. P. Seiler, M. P. Valoski, and M. A. Crivaro, Noise Exposure in U.S. Coal Mines, U.S. Department of Labor, Mine Safety, and Health Administration, Informational Report No. IR 1214, 1994, 46 pp.
[7]Federal Register, Health Standards for Occupational Noise Exposure in Coal, Metal, and Nonmetal Mines; Proposed Rule. 30 CFR, Parts 56, 57, 62, 70, and 71, Vol. **61**, No. 243. U.S. Department of Labor, Mine Safety

and Health Administration, Tuesday, December 17, 1996, pp. 66348–66469.
[8]Federal Register, *Health Standards for Occupational Noise Exposure in Coal, Metal, and Nonmetal Mines; Final Rule*. 30 CFR Parts 56 and 57 *et al*., Vol. **64**, No. 176. U.S. Department of Labor, Mine Safety and Health Administration, September 13, 1999, pp. 49548–49634.
[9]J. P. Seiler and D. A. Giardino, "The effect of threshold on noise dosimeter measurements and interpretation of their results," U.S. Department of Labor, Mine Safety and Health Administration, Informational Report No. IR 1224, 16 pp., 1994.
[10]Mine Safety and Health Administration, Environmental Noise Report, District Summaries for the Fiscal Year 1997.
[11]S. Fiscor and J. Lyles, "Prep plant population reflects industry," Coal Age **105**(10), 31–38 (2000).
[12]D. A. Giardino and J. P. Seiler, "Uncertainties associated with noise dosimeters in mining," J. Acoust. Soc. Am. **100**, 1571–1576 (1996).
[13]*The Noise Manual*, 5th ed., edited by E. H. Berger, L. H. Royster, J. D. Royster, D. P. Driscoll, and M. Layne (AIHA, Fairfax, 2000), Chaps. 3 and 7.
[14]ISO-140:1998. Acoustics – Measurement of sound insulation in buildings and of building elements – Part 4: Field measurements of airborne sound insulation between rooms. International Organization for Standardization, Geneva, Switzerland, 1998.
[15]Leo L. Beranek and István L. Vér, *Noise and Vibration Control Engineering* (Wiley Interscience, New York, 1992).
[16]ANSI S3.44-1996. American National Standard: Determination of occupational noise exposure and estimation of noise-induced hearing impairment. American National Standards Institute, New York, NY, 1996.
[17]David A. Bies and Colin Hansen, *Engineering Noise Control: Theory and Practice* (Taylor and Francis, New York, 2003).

Porous layer impedance applied to a moving wall: Application to the radiation of a covered piston

Olivier Doutres,[a] Nicolas Dauchez, and Jean-Michel Génevaux
Laboratoire d'Acoustique UMR CNRS 6613, Univ. du Maine, 72095 Le Mans cedex 9, France

(Received 27 June 2006; revised 6 September 2006; accepted 6 September 2006)

Modeling a porous layer mounted on a vibrating structure using acoustic impedance is investigated in this paper. It is shown that the use of surface impedance usually measured with the impedance tube method can provide an inaccurate estimation of the acoustic pressure radiated by the covered structure. The paper focuses on the derivation of an impedance, denoted the "transfer impedance," which describes accurately the dynamic movement of the porous layer. Biot's theory is used in the model to account for deformations in the thickness of the layer. Experimental validation is performed using a circular piston covered by a foam or a fibrous layer, radiating in an infinite half space. The radiation model including the transfer impedance shows close agreement with experimental data. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2359233]

PACS number(s): 43.50.Gf, 43.40.Rj, 43.20.Rz [KA] Pages: 206–213

I. INTRODUCTION

The reduction of the sound radiated by vibrating structures is usually performed using covering materials. If the structure is immersed in a heavy acoustic fluid, as in marine applications, the covering decouples the surrounding fluid from the vibrations of the structure due to deformation in the thickness.[1–3] In the case of light surrounding fluid, as in building or transport applications (car, aircraft, train), porous materials such as polymer foams and fibrous materials are widely used because they provide sound absorption and radiation attenuation due to interaction between the porous frame and the saturating fluid.[4]

The behavior of a layer attached to a vibrating structure can be described by using a simple model based on an impedance formulation. This model, known as "locally reacting," assumes that the motion at a point of the covering surface is independent of the motion of any other point of the surface area. When considering immersion in an acoustically heavy fluid,[5,6] this impedance is usually written in the form

$$Z_t = \frac{p}{v_p - v}, \tag{1}$$

where p is the pressure exerted by the fluid on the fluid-covering interface, v the velocity of the same covering surface and v_p is the velocity of the base plate (equal to the velocity on the other covering surface). In these studies, the covering is assumed to behave like massless distributed springs and the impedance is derived only by means of its thickness and bulk modulus. This simple model does not account for wave effects in the layer; however, it provides a good description of the main vibratory phenomena.

For acoustic radiation problems relating to vibrating structures immersed in light fluid, the impedance formulation is frequently used because of its apparent simplicity. However, according to the authors' knowledge, a clear ambiguity appears regarding the definition of the impedance to be applied to the moving structure. For example, in their book, Morse and Ingard[7] give an expression for the far field radiated pressure of a nonrigid boundary plane. The active and passive regions of the boundary plane are characterized by the same acoustic impedance which is defined as the structure reaction to an incoming pressure wave. This impedance is usually called the surface impedance and is given in the following form:

$$Z_s = \frac{p}{v}, \tag{2}$$

where p and v are, respectively, the acoustic pressure and the particle velocity at the nonrigid boundary surface. Mangulis[8,9] has already pointed out that the active and passive regions of a nonrigid plane should not be characterized by the same impedance. The case of a circular piston set in an infinite baffle was investigated, considering the characteristic impedances as vibratory impedances defined in terms of their mass per unit area. The same ambiguity in the definition of the acoustic impedance exists for applications dealing with porous coverings. Suzuki, Maruyama, and Ido,[10] investigated the effect of porous coverings on the pressure radiated by vibrating structures in a vehicle cabin in terms of an acoustic impedance as in the form of Eq. (2). In the case of porous coverings, p and v of Eq. (2) are the pressure and the normal component of velocity in the free air close to the front face of the material. This impedance applied to a porous layer is generally experimental data which can be obtain using a common method such as an impedance tube.[11] However, by using this impedance, the moving boundary condition on the vibrating surface is not taken into account. In the same way, Piot, Ville, and Auger[12] modeled the behavior of an absorbent material in dynamic conditions using the surface impedance Z_s. More recently, Shao and Mechefske[13] used a surface impedance independent of the frequency to account for absorbing materials mounted on vibrating cylindrical ducts, but, as in the previous models, the moving

[a]Author to whom correspondence should be addressed; electronic mail: olivier.doutres@univ-lemans.fr

boundary condition is not clearly demonstrated in that case.

The problem considered in this paper focuses on the ambiguity of acoustic impedance applied to a porous material attached to a moving structure. An impedance as in Eq. (1), usually used for heavy fluid configurations, is implemented in the case of a covered boundary moving in a light acoustic fluid: the moving behavior of the covering is taken into account. The first part of the paper presents the acoustic boundary conditions linked to the impedances Z_s and Z_t. Then, both impedances are derived for two kinds of porous material commonly encountered in industrial applications. Both materials differ by their mechanical properties: one is a relatively rigid foam layer and the other, a limp fibrous layer. A one-dimensional model based on Biot's theory[14] is used to take into account wave propagation in the solid and fluid phases of the porous media. Using Biot's theory is of great importance for the moving boundary configuration because the frame of the porous layer is directly excited by the vibrating structure. In order to validate the porous modeling using Z_t and illustrate the consequences of using Z_s instead of Z_t in an acoustic problem with moving covered boundaries, the derived impedances are applied to the radiation problem of a piston covered with the porous layers and set in an infinite baffle. Finally, these simulations are compared to acoustic radiation measurements.

II. IMPEDANCE LINKED TO COVERED BOUNDARIES

A. Definition of the boundary conditions

Let us consider an acoustic domain Ω limited by the nonrigid boundary $\Gamma=\partial\Omega$ as illustrated in Fig. 1. The normal to the boundary Γ, outwardly directed, is denoted n. The harmonic pressure field p is governed by the homogeneous linear wave equation (the time factor $e^{j\omega t}$ is omitted throughout this paper),

$$\nabla^2 p + k^2 p = 0 \quad \text{in } \Omega, \tag{3}$$

where the wave number is $k=\omega/c_0$, with ω as the angular frequency and c_0 the speed of sound in the medium. In the case of an acoustic problem involving a porous covering, the boundary condition can be expressed in terms of a harmonic acoustic impedance. If the backing is a rigid and impervious wall [see Fig. 2(a)], the impedance applied to Γ, denoted in this paper as the "surface impedance" Z_s, is usually given as in Eq. (2). This impedance expresses the coupling between the scalar pressure and the z component of particle velocity in the free air close to the front face of the porous layer, respectively $p(0)$ and $v(0)$. Substituting the normal particle velocity v in Eq. (2) for the one-dimensional Euler equation

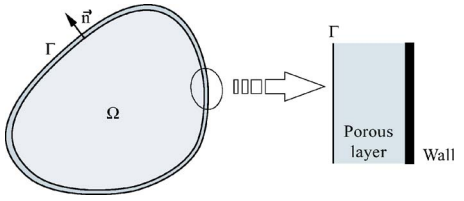


FIG. 1. System consisting of an acoustic domain Ω limited by a nonrigid boundary Γ.

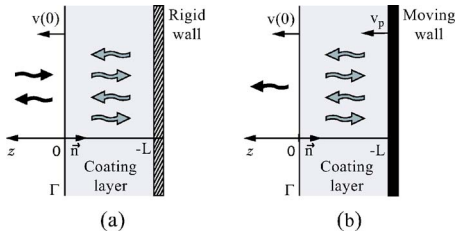


FIG. 2. Boundary condition: (a) static covered wall, (b) moving covered wall.

$$v = -\frac{1}{j\omega\rho_0}\frac{\partial p}{\partial z}, \tag{4}$$

gives the associated boundary condition

$$\frac{\partial p}{\partial z} + jk\beta_s p = 0 \quad \text{on } \Gamma, \tag{5}$$

where $\beta_s=\rho_0 c_0/Z_s$, with ρ_0 as the acoustic fluid density. This boundary condition is known as the homogeneous mixed Dirichlet-Neumann boundary condition or Robin boundary condition. The acoustic impedance Z_s to be implemented in this boundary condition is easily measured, using the impedance tube method, for example. In addition, this impedance is mainly used to derive the absorption coefficient of a porous layer backed by a rigid wall.

Let us now consider the covered wall moved by a harmonic velocity of magnitude v_p [see Fig. 2(b)]. As usually encountered in literature dealing with covered plates immersed in heavy fluid, the impedance applied to the boundary can be written in terms of the relative velocity between the velocity of the wall v_p and the normal component of the fluid velocity at the fluid-porous interface as written in Eq. (1). This formulation involves a transfer of energy from the moving wall to the acoustic medium through the covering and will thus be denoted in this paper as the "transfer impedance."

By substituting the normal particle velocity v in Eq. (1) for Eq. (4), the associated boundary condition is written as an inhomogeneous mixed boundary condition

$$-\frac{\partial p}{\partial z} + jk\beta_t p = j\omega\rho_0 v_p \quad \text{on } \Gamma, \tag{6}$$

where $\beta_t=\rho_0 c_0/Z_t$. The source component on the right hand side of Eq. (6) appears because the impedance is expressed in terms of the wall velocity v_p. As a consequence, it may be difficult to measure the impedance, although it is usually considered as experimental data.

However, in most of the applications met in the literature which involve porous coverings, the acoustic impedance Z_s characteristic of Eq. (5) is also used in Eq. (6) instead of Z_t. This may be explained by the fact that only Z_s is easily obtained by experiment. However, substituting Z_t for Z_s may give an inaccurate estimation of the acoustic pressure generated in the domain Ω. This is illustrated in the following, first, by deriving both impedances with a one-dimensional model based on Biot's theory, and then, by implementing these impedances in a radiation problem of a covered piston.

TABLE I. Properties of the porous materials.

Porous	A	B
Thickness: L (mm)	18.9	29
Air flow resistivity: σ (N s/m^4)	57 000	105 000
Porosity: ϕ	0.97	0.95
Tortuosity: α_∞	1.54	1
Viscous length: Λ (μm)	24.6	35.1
Thermal length: Λ' (μm)	73.8	105.3
Skeleton density: ρ_1 (kg/m^3)	46	17
Young's modulus at 5 Hz: E (kPa)	214	1.4
Structural loss factor at 5 Hz: η	0.115	0.1
Poisson's ratio: ν	0.3	0

B. Derivation of the acoustic impedance for a locally reacting porous material

To derive the transfer impedance Z_t [Eq. (1)], it is necessary to calculate the acoustic pressure and porous surface particle velocity induced by the wall movement v_p. As a result of the "locally reacting" assumption, a simple one-dimensional model may be used. It assumes that only plane waves propagate in the z direction. It is interesting to note that this model is equivalent to the transfer matrix method[4,15] where the multilayer was excited by plane waves with normal incidence.

The theoretical one-dimensional model of the porous layer is based on Biot's theory[4,14] to take into account vibration of both solid and fluid phases induced by the wall vibration. According to Biot's theory two longitudinal waves can propagate at the same time in both phases. The one mainly supported by the frame is called the frame-borne wave and the one mainly supported by the saturating fluid is called the airborne wave.[4] A rotational wave may also be supported by the frame but it is not excited in the case of a one-dimensional problem. The motion of the poroelastic media is described by the macroscopic displacement of the solid and fluid phases represented by the scalars u^s and u^f, respectively. Four waves contribute to the displacement of one fluid or solid particle: two waves propagating in the z direction and two waves propagating in the $-z$ direction [see Fig. 2(b)]. Since the transfer impedance will be applied to a piston radiating in an infinite half space, we consider only one acoustic wave propagating from the porous interface to the semi-infinite acoustic domain. This assumption allows us to use the simple relation between acoustic pressure and particle velocity on the porous surface: $p(0)/v(0)=Z_0$, where Z_0 is the characteristic impedance of the acoustic fluid.

The boundary conditions at the fluid-porous interface ($z=0$) are: continuity of the normal stress in the solid and fluid phases with the external pressure

$$\sigma_z^s(0) = -(1-\phi)p(0), \tag{7}$$

$$\sigma^f(0) = -\phi p(0), \tag{8}$$

where ϕ is the porosity of the porous material, and continuity of the total flow

$$j\omega[(1-\phi)u^s(0) + \phi u^f(0)] = v(0). \tag{9}$$

At $z=-L$ the velocity of the fluid and the velocity of the frame are both equal to the wall velocity

$$j\omega u^s(-L) = j\omega u^f(-L) = v_p. \tag{10}$$

From Eqs. (7)–(10), it is possible to calculate amplitude of the five waves and derive the transfer impedance.

The surface impedance Z_s is derived using the same formalism with the appropriate boundary conditions. In this configuration, an acoustic plane wave hit the porous layer at $z=0$ and Eqs. (7)–(9) are used to express the continuity of stress and total flow at this interface. On the surface of the porous layer in contact with the backing, the displacement of the air and the frame are equal to zero

$$u^s(-L) = u^f(-L) = 0. \tag{11}$$

The impedances Z_s and Z_t are derived for two kinds of porous materials, the characteristics of which are listed in Table I. Material A is a foam layer with a stiff skeleton and material B is a light fibrous layer with a very soft skeleton, both fairly resistive to airflow. Figures 3 and 4 show the simulation of the transfer impedance Z_t and the surface impedance Z_s for both materials.

When the wall is covered with material A (Fig. 3), it appears that the impedances Z_t and Z_s are similar as far as

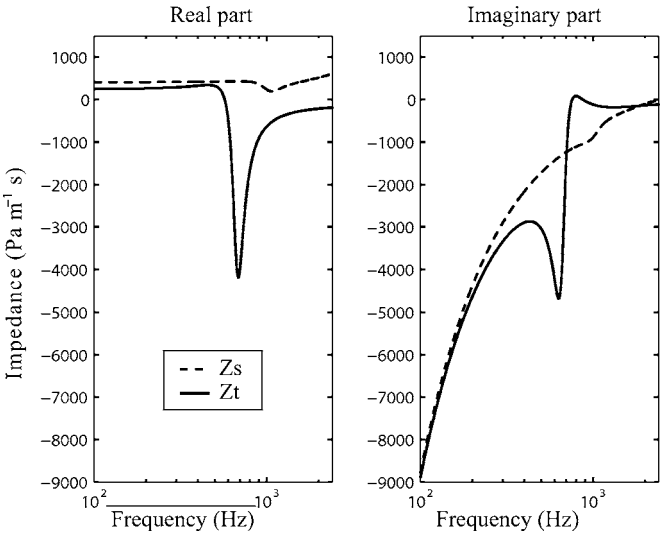


FIG. 3. Material A: real and imaginary part of the (- -) surface impedance Z_s and (—) transfer impedance Z_t.

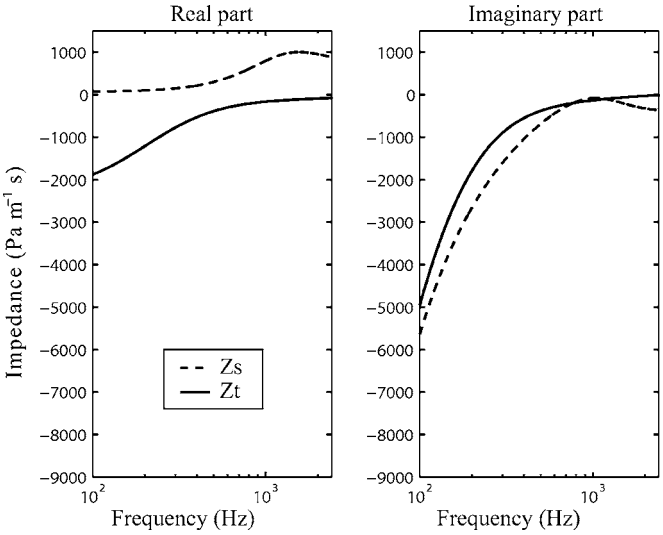


FIG. 4. Material B: real and imaginary part of the (- -) surface impedance Z_s and (—) transfer impedance Z_t.

500 Hz. Above this frequency, the real part of the transfer impedance becomes negative. Now, if the wall is covered with material B (Fig. 4), the imaginary part of both impedances is seen to be similar in the entire frequency range. However, the real parts are completely different, the real part of the transfer impedance being negative in the whole frequency range.

C. Discussion

For a better understanding, the vibratory behavior of the porous layers (A and B) is investigated for both kinds of excitation (depicted in Fig. 2).

Let us first consider the case of material A attached to the moving wall. Figure 5(a) (solid curve) shows the velocity ratio between the frame velocity at the air-porous interface and the wall velocity v_p at $z=-L$. In the low frequency range, the velocity ratio is equal to one (0 dB), i.e., the material follows the wall movement and its thickness remains constant. At higher frequencies (up to 500 Hz), the thickness of the porous layer does not remain constant during excitation and the velocity ratio reaches a maximum around 1100 Hz. Figure 6(a) shows that this maximum appears around the first quarter-wavelength resonance frequency f_r corresponding to the frame-borne wave: the Biot frame-borne wavelength is four times the layer thickness L. This resonance frequency may be approached by simply considering the properties of the frame *in vacuo*[16]:

$$f_r \approx \frac{1}{4L}\sqrt{\frac{E\frac{(1-\nu)}{(1+\nu)(1-2\nu)}}{\rho_1}}, \tag{12}$$

where L is the thickness, E, ν, and ρ_1 are, respectively, the Young modulus, the Poisson ratio and the density. For the given parameters (see Table I), f_r is evaluated at 1047 Hz for material A and it is observed at around 1100 Hz in Fig. 6(a). Considering next the static wall configuration, Fig. 5(a) (dashed curve) shows the ratio between the frame velocity at the air-porous interface and the velocity of the acoustic fluid $v(0)$ at $z=0$. The low velocity ratio amplitude in the entire frequency band shows that the frame of the layer can be considered motionless, even at the resonance frequency of the frame. From the observations carried out in the low frequency range, it is possible to conclude that the acoustic behavior of material A, which is due to the fluid movement in the pores, is similar for both configurations because there is negligible deformation in the thickness of the layer. The frame's thickness tends to be constant at low frequencies because the wavelength of the frame-borne wave is long compared to the thickness of the layer as seen in Fig. 6(a). This explains the similarities observed below 500 Hz between the impedances Z_s and Z_t. However, at higher frequencies, the vibratory behavior of the porous frame exerts a greater influence on the acoustical behavior of the layer and it is different for both configurations: the frame of the layer is excited much more by mechanical loading than by acoustical loading.

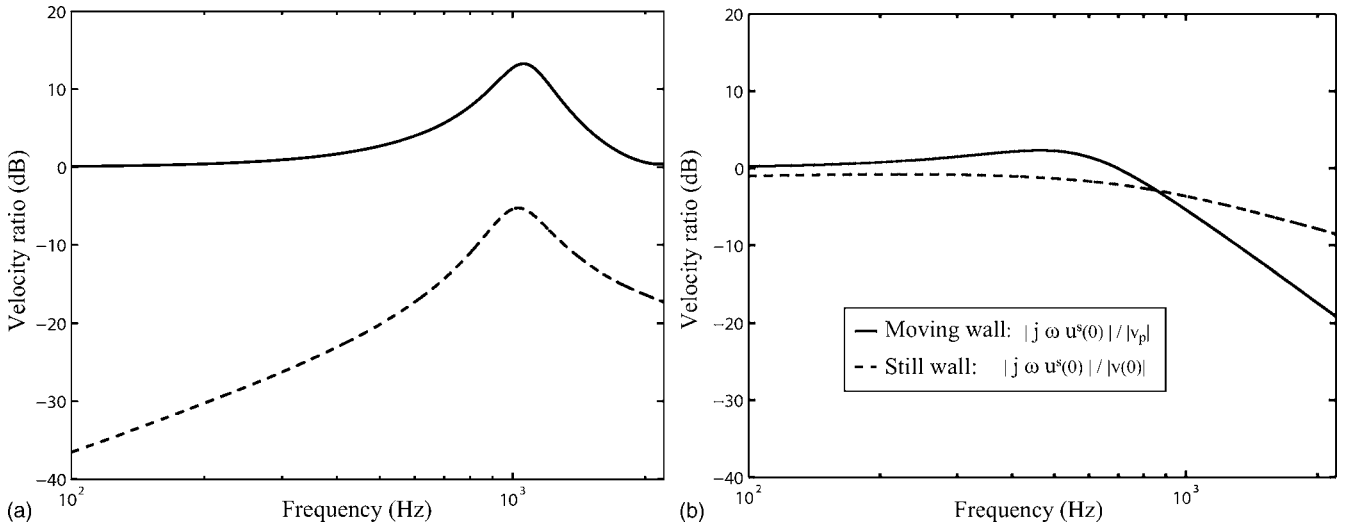


FIG. 5. Ratio between the velocity of the frame at the air-porous interface and: (- -) the velocity of the fluid at the porous surface $v(0)$ (still wall model), (–) the velocity of the piston v_p (moving wall model): (a) Material A; (b) Material B.

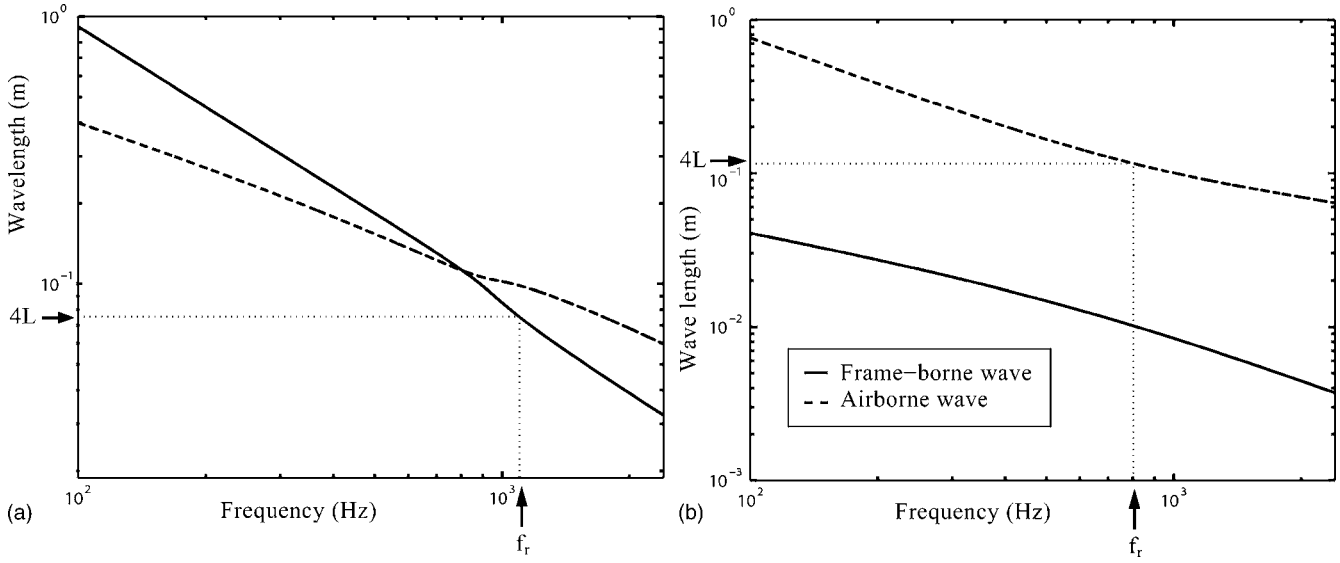


FIG. 6. Biot wavelengths: (a) Material A; (b) Material B.

Let us now consider the configuration in which the wall is covered with material B. If the source is the wall vibration [Fig. 5(b) solid curve], it is seen that the frame follows the wall movement at low frequencies. In the mid frequency range, a slight increase in the velocity ratio is observed, related in this case to the first $\lambda/4$ resonance of the airborne wave [see Fig. 6(b)]. At higher frequencies, the velocity ratio decreases until it reaches −20 dB at 2400 Hz which means that the frame tends to become motionless at the surface of the layer. Thus, material B allows a better decoupling between the vibration of the wall and the external media than material A because the relative velocity between both phases is greater. In the still wall configuration [Fig. 5(b) dashed curve] at low frequencies, it is seen that the ratio of the velocity of the frame at the air-porous interface to the velocity of fluid particle is close to 0 dB. This means that the airflow drags the fibers and induces a deformation in the thickness of the fibrous layer. At higher frequencies, the velocity ratio decreases and the frame also tends to become motionless, but at a slower rate than in the case of the moving wall.

To sum up, the surface impedance Z_s is close to the transfer impedance Z_t in a restrictive frequency range where the frame can be considered "rigid" for both static and moving backing configurations. This occurs when the frame-borne wavelength is greater than the porous layer thickness. In this paper, this only applies to material A below 500 Hz.

III. ACOUSTIC RADIATION OF A COVERED PISTON

In order to compare simulations with experimental results, the acoustic radiation of a circular piston covered with a porous layer and set in an infinite baffle is investigated. The aim is to illustrate the consequences of using the surface impedance Z_s instead of the transfer impedance Z_t in an inhomogeneous mixed boundary condition.

A. Theory

The radiation efficiency factor σ_R is used to characterize the acoustical effect of the covering. It is defined as the ratio of the radiated acoustic power, denoted as W, over the vibratory power

$$\sigma_R = \frac{W}{\rho_0 c_0 S_P \frac{|v_p|^2}{2}}, \tag{13}$$

where S_P is the piston surface, ρ_0 and c_0 are the ambient density and velocity of sound in air, and v_p is the amplitude of the piston velocity. The radiated acoustic power W can be derived by integrating the far field acoustic intensity $I(\boldsymbol{r})$ expressed in spherical coordinates (r, θ, φ) over a hemisphere of radius r,

$$W = \int_0^{2\pi} \int_0^{\pi/2} I(\boldsymbol{r}) r^2 \sin\theta d\theta d\varphi. \tag{14}$$

Calculus of the intensity uses the asymptotic form of the far field pressure $p(\boldsymbol{r})$ as:

$$I = \frac{|p(\boldsymbol{r})|^2}{2\rho_0 c_0}. \tag{15}$$

Mangulis[8,9] and Feit and Duncan[17] have investigated the pressure radiated from a rigid piston set in a nonrigid baffle. A similar approach is presented in the following. However, since we are interested in the influence of the covering applied to a moving structure, the baffle is taken as being perfectly reflecting.

Let us consider a circular piston in the infinite rigid baffle at $z=0$ (see Fig. 7). The surface of the piston is denoted by S_P, and the surface of the baffle by S_B. In the acoustic domain, the pressure p is a solution of the harmonic Helmholtz equation [Eq. (3)]. The boundary conditions are: Sommerfeld condition

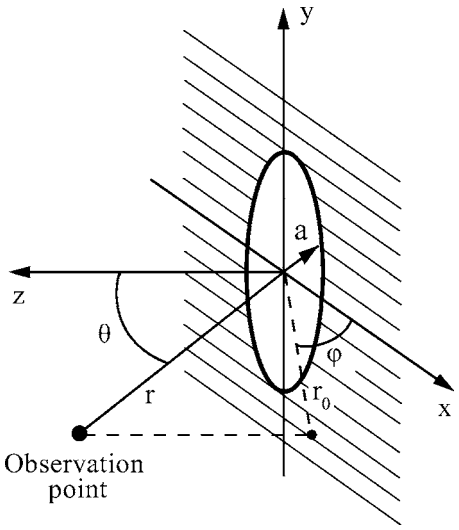


FIG. 7. Geometry of system and coordinates defining observation point.

$$\lim_{r\to\infty} r\left(\frac{\partial p}{\partial r} + jkp\right) = 0, \tag{16}$$

homogeneous Neumann condition on S_B

$$\frac{\partial p}{\partial z} = 0, \tag{17}$$

and inhomogeneous mixed condition on S_P

$$-\frac{\partial p}{\partial z} + jk\beta_t p = j\omega\rho_0 v_p, \tag{18}$$

where $\beta_t = \rho_0 c_0 / Z_t$ and Z_t is the transfer impedance of Eq. (1) expressed on the surface of the covering.

The pressure field is derived using the surface integral equation. The Green function (at $z=0$) chosen to satisfy the Sommerfeld boundary condition and the boundary condition

$$\frac{\partial G}{\partial z} = 0 \tag{19}$$

on both S_B and S_P is

$$G(\boldsymbol{r},\boldsymbol{r}_0) = \frac{e^{-jk|\boldsymbol{r}-\boldsymbol{r}_0|}}{2\pi|\boldsymbol{r}-\boldsymbol{r}_0|}. \tag{20}$$

Next, the solution of Eq. (3) subjected to the specified boundary conditions is given by

$$p(\boldsymbol{r}) = j\omega\rho_0 \int_{S_P} v(\boldsymbol{r}_0) G(\boldsymbol{r},\boldsymbol{r}_0) dS_{\mathbf{P}}, \tag{21}$$

where v is the fluid velocity at the fluid-porous interface. This velocity v can be expressed in terms of the piston velocity v_p using the boundary condition of Eq. (18). First, the normal derivative of the acoustic pressure is replaced by the particle velocity using Eq. (4):

$$j\omega\rho_0 v + jk\beta_t p = j\omega\rho_0 v_p. \tag{22}$$

Then, a radiation condition at the fluid-porous interface is used to write the acoustic pressure in terms of the particle velocity, $Z_{\text{rad}} = p/v$. Note that, since the plane wave approximation is used in the porous model to describe wave propagation (see Sec. II B), the acoustic pressure is related to fluid velocity by the impedance of the acoustic fluid Z_0. Furthermore, by definition, piston velocity v_p is uniform on S_p.

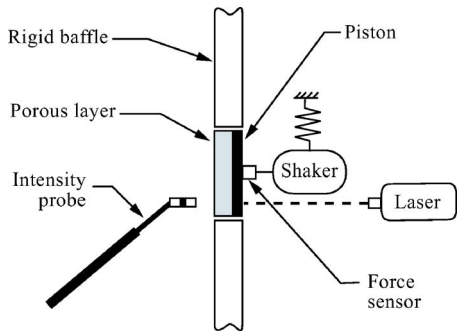


FIG. 8. Experimental setup schema to measure structure radiation efficiency.

Equation (21) can then be rewritten in the following form:

$$p(\boldsymbol{r}) = j\omega\rho_0 \frac{Z_t}{Z_t + Z_{\text{rad}}} \int_{S_P} v_p G(\boldsymbol{r},\boldsymbol{r}_0) dS_{\mathbf{P}}. \tag{23}$$

This gives the relation between the far field pressure radiated by the bare system $p'(\boldsymbol{r})$ and by the covered system $p(\boldsymbol{r})$,

$$p(\boldsymbol{r}) = \frac{Z_t}{Z_t + Z_{\text{rad}}} p'(\boldsymbol{r}) \tag{24}$$

where $p'(\boldsymbol{r})$ is given by Morse and Ingard[7] for a circular piston of radius a

$$p'(\boldsymbol{r}) = jk\rho_0 c_0 v_p \frac{e^{-jkr}}{2\pi r} \pi a^2 \frac{2J_1(ka\sin\theta)}{ka\sin\theta}. \tag{25}$$

The multiplicative term of Eq. (24) has no influence on the far field directivity pattern. It acts mainly as an harmonic magnitude correction due to the presence of the covering.

B. Experimental validation

Radiation efficiency measurements of a piston with and without covering were performed. The flat piston was set in a rigid baffle (see Fig. 8) and radiated in an anechoïc room. The excitation was performed using a shaker driven with a white noise signal. The piston is honeycomb plate with a radius of 75 mm. Because this structure is extremely stiff and light, its first natural frequency resonance are shifted above 2500 Hz. Two types of covering materials were tested: a stiff foam layer 18.9 mm thick and a very soft fibrous layer 29 mm thick. The properties of these materials are given in Table I (see Sec. II B). The viscoelastic characteristics of the porous frames, Young's modulus E and loss factor η, were measured at low frequency using the quasi-static method.[18,19]

The vibratory power of the plate was determined from the quadratic normal velocity measured at the piston surface by a laser vibrometer. The acoustic radiated power was determined from intensity measurements on porous surface using an intensity probe made from two 1/2 in. microphones

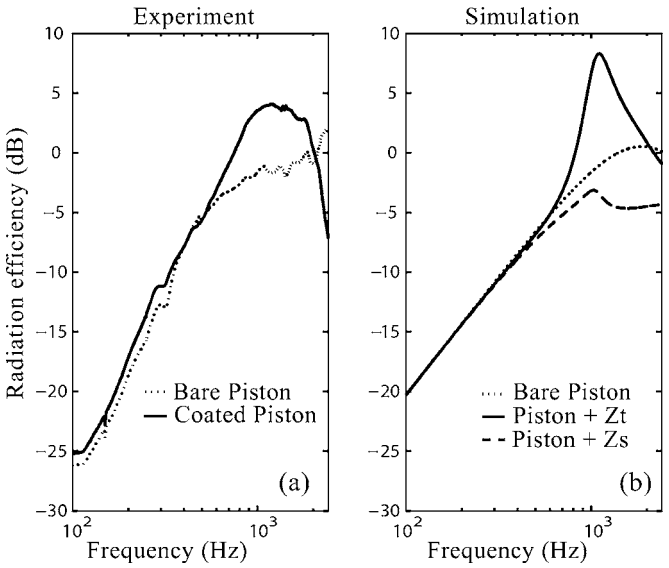


FIG. 9. Acoustic radiation efficiency of a bare and covered (material A) circular piston set in a rigid baffle: (a) experiment; (b) simulation.

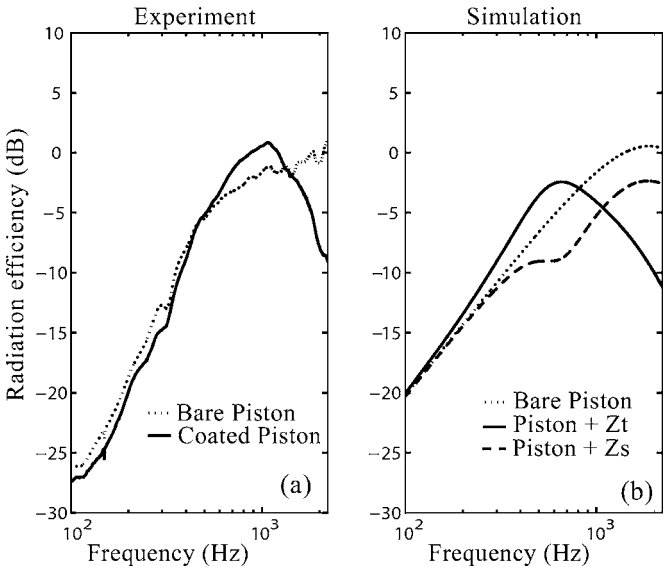


FIG. 10. Acoustic radiation efficiency of a bare and covered (material B) circular piston set in a rigid baffle: (a) experiment; (b) simulation.

12 mm apart. The lateral dimensions of the piston are large enough in relation to the thickness of the porous layer for the effect of the edges to be ignored.

Figure 9(a) shows the radiation efficiency measurement of the bare piston and the radiation efficiency measurement of the piston covered with material A. Three distinct zones are observed. In the low frequency range, the porous layer has little or no effect on the radiation efficiency. In the mid frequency range, an amplification of the acoustic radiation is observed around the first resonance frequency of the frame f_r (Eq. (12)). This phenomenon which deteriorates the insulation efficiency of the covering has already been observed by Harrison *et al.*[20] in the case of vibration reduction using insulating mounts. Above 2000 Hz, the radiation efficiency decreases due to losses and frame decoupling.

Figure 9(b) shows the simulated data when both the transfer impedance Z_t, which describes properly the moving boundary condition, and the surface impedance Z_s (see Sec. II A) are applied in the inhomogeneous mixed boundary condition [Eq. (18)]. There is seen to be close agreement between the model including transfer impedance Z_t (solid line) and the measurements over the entire frequency range considered, i.e., the radiation peak is predicted and a decrease is observed at higher frequencies. Note that the predicted radiation peak due to the frame resonance is sharper than the experimental one because the porous model only accounts for dissipations in one direction in the thickness of the porous layer. Furthermore, constant viscoelastic properties E and η measured at 5 Hz are used in the model, but are expected to increase at higher frequency.[18,21,22] Thus, with more appropriate values of viscoelastic parameters, the simulated peak would be higher in frequency and lower in amplitude.

Figure 10(a) shows the radiation efficiency measurement of the piston covered with material B. It is seen that the same phenomena observed experimentally with material A occur with material B. An increase in radiation occurs around 800 Hz due, in this case, to the $\lambda/4$ airborne wave resonance in the thickness of the layer. At higher frequencies, the radiation efficiency strongly decreases compared to the radiation of the bare piston. The soft material B allows a better decoupling between the piston and the external media than material A. Note that only the model including the impedance Z_t closely agrees with experimental data [see Fig. 10(b)].

In the frequency range where the covering has a significant effect, using Z_s instead of Z_t is improper: the radiation peak is not predicted and the high frequency drop does not agree with the experimental data.

IV. CONCLUSION

In this paper, the impedance to be applied to a moving structure covered with a porous layer has been investigated. The well known surface impedance Z_s which can easily be measured in an impedance tube is often applied to vibrating structures to account for the covering effects. However, it is shown that the impedance must be expressed in terms of the base movement to account for the real behavior of the layer. This impedance, called transfer impedance and denoted as Z_t, is derived using Biot's theory and compared to the surface impedance. It is shown that these impedances are similar as long as the longitudinal wavelengths are large compared to the layer thickness. In the case of a covered piston set in an infinite rigid baffle, the acoustic radiation peak and the decrease at higher frequencies due to the presence of the porous layer can only be predicted using the transfer impedance: the surface impedance measured in an impedance tube is not suitable in this case.

ACKNOWLEDGMENTS

The authors are grateful to the CTTM (Centre de Transfert de Technologie du Mans) where the measurements were performed and the CNRS and Région Pays de la Loire for

their financial support. The authors would also like to thank Professor Michel Bruneau for his useful comments on the manuscript.

[1]G. Maidanik, "Use of decoupling to reduce the radiated noise generated by panels," J. Sound Vib. **81**(2), 165–185 (1982).
[2]B. Laulagnet and J. L. Guyader, "Sound radiation from finite cylindrical coated shells, by means of asymptotic expansion of the three-dimensional equations for coating," J. Acoust. Soc. Am. **96**(1), 277–286 (1994).
[3]A. Berry, O. Foin, and J. P. Szabo, "Three-dimensional elasticity model for a decoupling coating on a rectangular plate immersed in a heavy fluid," J. Acoust. Soc. Am. **109**(6), 2704–2714 (2001).
[4]J. F. Allard, *Propagation of Sound in Porous Media: Modelling Sound Absorbing Materials* (Elsevier Applied Science, London, 1993).
[5]B. Laulagnet and J. L. Guyader, "Sound radiation from finite cylindrical coated shells covered with a compliant layer," Trans. ASME, J. Vib. Acoust. **113**, 267–272 (1991).
[6]O. Foin, A. Berry, and J. P. Szabo, "Acoustic radiation from an elastic baffled rectangular plate covered by a decoupling coating and immersed in a heavy acoustic fluid," J. Acoust. Soc. Am. **107**(5), 2501–2510 (2000).
[7]P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (McGraw–Hill, New York, 1968).
[8]V. Mangulis, "Radiation of sound from a circular rigid piston in a nonrigid baffle," Int. J. Engng. Sci. **2**(1), 115–126 (1964).
[9]V. Mangulis, "Comments on: Numerical evaluation of the radiation impedance for a piston in a nonrigid baffle," J. Acoust. Soc. Am. **45**, 1045–1046 (1969).
[10]S. Suzuki, S. Maruyama, and H. Ido, "Boundary element analysis of cavity noise problems with complicated boundary conditions," J. Sound Vib. **130**(1), 79–91 (1989).
[11]A. F. Seybert and D. F. Ross, "Experimental determination of acoustic properties using a two-microphone random-excitation technique," J. Acoust. Soc. Am. **61**, 1362–1370 (1977).
[12]A. Piot, J. M. Ville, and J. M. Auger, "Rayonnement acoustique dans un conduit par un piston recouvert d'un matriau absorbant," *Proc. of the International Symposium on the Prediction of the Noise Emitted by Vibrating Structures*, CETIM-Senlis, March 26–28 (1991).
[13]W. Shao and C. K. Mechefske, "Analysis of the sound field in finite length infinite baffled cylindrical ducts with vibrating walls of finite impedance," J. Acoust. Soc. Am. **117**(4), 1728–1736 (2005).
[14]M. A. Biot, "The theory of propagation of elastic waves in a fluid-saturated porous solid. I. Low frequency range. II. Higher frequency range," J. Acoust. Soc. Am. **28**, 168–191 (1956).
[15]B. Brouard, D. Lafarge, and J. F. Allard, "A general method of modelling sound propagation in layered media," J. Sound Vib. **183**(1), 129–142 (1995).
[16]J. F. Allard, C. Depollier, P. Guignouard, and P. Rebillard, "Effects of a resonance of the frame on the surface impedance of glass wool of high density and stiffness," J. Acoust. Soc. Am. **89**, 999–1001 (1991).
[17]D. Feit and M. E. Duncan, "Numerical evaluation of the radiation impedance for a piston in a nonrigid baffle," J. Acoust. Soc. Am. **43**, 885–886 (1968).
[18]N. Dauchez, M. Etchessahar, and S. Sahraoui, "On measurement of mechanical properties of sound absorbing materials," *Second Biot Conference on Poromechanics*, Grenoble, France, Sept. 14–16 (2002), pp. 197–202.
[19]M. Etchessahar, S. Sahraoui, L. Benyahia, and J. F. Tassin, "Frequency dependence of elastic properties of acoustic foams," J. Acoust. Soc. Am. **117**(3), 1114–1121 (2005).
[20]M. Harrison, A. O. Sykes, and M. Martin, "Wave effects in isolation mounts," J. Acoust. Soc. Am. **24**, 62–71 (1952).
[21]T. Pritz, "Unbounded complex modulus of viscoelastic materials and the Kramers-Kronig relations," J. Sound Vib. **279**, 687–697 (2005).
[22]O. Doutres and N. Dauchez, "Characterization of porous materials viscoelastic properties involving the vibroacoustical behavior of coated panels," *Proc. of the International Symposium on the Acoustics of Poro-Elastic Materials (SAPEM)*, Lyon, France, Dec. 7–9 (2005).

Properties of transparent sound-absorbing panels for use in noise barriers

Francesco Asdrubali[a)]
Department of Industrial Engineering, University of Perugia, via G. Duranti 67, 06125 Perugia, Italy

Giulio Pispola[b)]
Department of Industrial Engineering, University of Perugia, via G. Duranti 67, 06125 Perugia, Italy

(Received 7 June 2006; revised 20 October 2006; accepted 20 October 2006)

Sound absorption and optical transparency are among the most useful properties of noise barriers. While the latter is required to reduce visual impact and for aesthetical reasons, the former is required whenever conditions of multiple reflections and presence of close, high receivers occur. The technical feasibility of a transparent, sound-absorbing panel for outdoor antinoise devices is investigated in this paper. An analysis of acoustical performance of multiple perforated plates is performed employing an existing theory for microperforated absorbers under normal incidence and diffused sound field. An optimization of the geometrical parameters is carried out on the basis of the European classification criteria of noise barriers for roadways. An optimized three-layer configuration can achieve sound-absorption properties similar to nontransparent products with only a limited loss of visual transparency and appropriate mechanical strength. Experimental data obtained with an impedance tube on small test samples made of transparent polycarbonate and in a reverberation room on full-scale prototypes are reported, showing a rather good agreement with the theoretical predictions. The optical performance of a multilayered configuration is evaluated also.
 [DOI: 10.1121/1.2395916]

PACS number(s): 43.50.Gf, 43.55.Ev, 43.20.Mv [KA] Pages: 214–221

I. INTRODUCTION

Poroelastic materials are widely used in noise barriers, since they fulfill many requirements in terms of both acoustic and nonacoustic performances. However, they are not suitable when properties of light transparency are also requested. The possibility to integrate sound absorption and light transparency was originally proposed for indoor applications by Fuchs and Zha,[1] who designed perforated acrylic glazing to be mounted in front of windows in the German Parliament in Bonn. More recently, Kang and Brocklesby[2] explored the use of microperforated plates in silencers for ventilating window systems. However, there is no published evidence of previous investigations concerning outdoor noise barriers. In recent years, an increasing demand for multifunctionality in the design of noise barriers has been noticed: among the major requirements there are visual impact, sound absorption on the receiver side, and durability. The former is usually tackled employing transparent materials (e.g., polycarbonate, polymethylmethacrylate, stratified glasses) both for aesthetical and safety reasons. The use of sound-reflective surfaces can however deteriorate the overall acoustic performance of noise barriers in certain conditions,[3] e.g., parallel barriers and presence of close, high receivers on the roadside. Such conditions are likely to coexist in high-density urban areas where also visual impact often has to be taken into account. Furthermore, the use of fibrous materials within the barrier panels can result in a poor durability of the acoustic performance due to weathering and fiber quality deterioration.

This paper investigates the application of multiple perforated transparent panels to noise barriers, both from theoretical and experimental points of view. The model originally adopted by Maa[4] for the acoustic surface impedance of a microperforated panel (MPP) is employed here for the optimization of the geometrical parameters of the perforations, with particular concern to traffic noise abatement. The adopted values of some of the geometrical parameters depend on the specific application and represent constraints for the design optimization: e.g., layer thickness cannot be too small, in order to guarantee a reasonable mechanical strength for the whole panel. The use of multiple resonators was explored to achieve a good absorption performance in a broad frequency range. The absence of porous materials within the cavities can also help to enhance the robustness of sound-absorption properties.

Section II briefly introduces the well-known theory of sound propagation in small tubes and perforated panels. In Sec. III the sound-absorption performance of single and multiple perforated panels is discussed. In Sec. IV the authors report comparisons between numerical and experimental data for several optimized configurations suitable for outdoor applications. Finally, Sec. V is devoted to the experimental evaluation of the optical performance of a configuration with multiple layers.

II. THEORETICAL REMARKS

The thermo-viscous effect constitutes the basic principle of a microperforated panel (MPP).[4] When a sound wave

[a)]Electronic mail: fasdruba@unipg.it
[b)]Electronic mail: pispola.unipg@ciriaf.it

strikes a hole in the micro perforated panel, the velocity distribution of the oscillatory flow across the hole cross section is not uniform. A distinct velocity gradient in the radial direction is present and is associated with shear stresses which result in energy dissipation into heat. Such phenomenon is not entirely adiabatic as part of this heat is transmitted to the panel itself, depending on the viscous and thermal properties of the filling fluid and on the thermal conductivity of the panel.

In the case of an MPP constituted by a lattice of cylindrical tubes, with radius r_0 and length t both much smaller than the wavelength of the incident sound, the normalized surface impedance, z_{MPP}, can be expressed in terms of the wave number in air, k, the tube length t, the porosity, p, i.e., the ratio of the perforated area to the overall area of the panel, and the perforation constant, κ, proportional to the ratio of the hole radius and the viscous boundary layer thickness,[4,5]

$$\kappa = r_0 \sqrt{\frac{\rho_0 \omega}{\mu}}, \tag{1}$$

$$z_{\mathrm{MPP}} = r + j\omega m = \frac{jkt}{p}\left[\frac{1}{\Theta(\kappa')} + \frac{16}{3\pi}\frac{r_0}{t}\frac{1}{\Theta(\kappa)\psi(p)}\right], \tag{2}$$

$$\Theta(x) = 1 - \frac{2}{x\sqrt{-j}}\frac{J_1(x\sqrt{-j})}{J_0(x\sqrt{-j})}, \tag{3}$$

where ρ_0 is the static fluid density, μ the dynamic viscosity, ω the angular frequency, r the acoustic resistance, m the mass reactance, and J_0 and J_1 are the Bessel functions of the first kind and zeroth- and first order, respectively. All the quantities used in Eqs. (1)–(3) and following should be provided in SI units. The semiempirical polynomial correction function, ψ, dependent on the parameter p, has been here adopted, following Melling,[5] to take into account interference phenomena between neighboring holes. In addition, the model by Sivian[6] has been employed for the end corrections of the resistance and the mass reactance. The heat transfer from the fluid to the tube walls is accounted for by employing a fictitious value for the air viscosity[4,5] ($\mu' \approx 2\mu$) and therefore for the perforation constant, κ', in Eq. (2).

III. SOUND ABSORPTION OF PERFORATED PANELS

Employing Eq. (2) and the electro-acoustic analogy, it is straightforward to predict the surface impedance, $z_{\mathrm{MPPA},\theta}$ of a MPP mounted at distance D from a rigid surface (MPPA, microperforated panel with an air gap; see Fig. 1). In the case of a plane wave striking at an angle θ with respect to the surface normal,[4]

$$z_{\mathrm{MPPA},\theta} = z_{\mathrm{MPP}} + z_D = r + j\left(\omega m - \frac{\cot(kD\cos\theta)}{\cos\theta}\right). \tag{4}$$

Equation (4) regards the MPP as a locally reacting element and the air cavity as a nonlocally reacting one. Wu[7] suggested to employ a local reaction assumption also for the cavity when the perforated panels are partitioned, and in this case the cavity impedance, z_D, in Eq. (4) does not depend on the incidence angle and corresponds to the normal incidence case (θ=0).

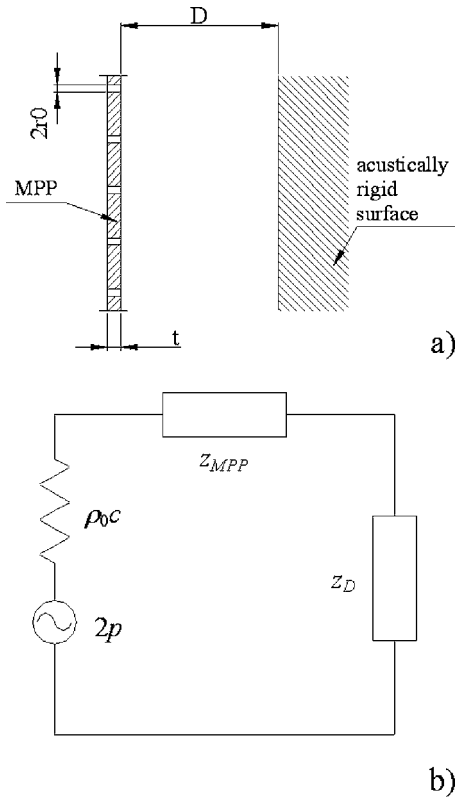


FIG. 1. Schematic drawing (a) and electro-acoustic equivalent circuit (b) of a perforated panel with an air gap (MPPA).

As pointed out by Kang *et al.*,[8] Eq. (4) should also take into account the impedance of the panel itself, regarded as a membrane acting in parallel with the MPP impedance. However, the model adopted here does not include this effect, considering the relatively high surface density of the employed panels (>2.4 kg/m^2; see Sec. IV). On the other hand, the fluid-structural modal coupling phenomenon studied by Lee *et al.*[9] can be observed at some frequencies from the experimental data obtained in the impedance tube (see Sec. IV): such effect is not taken into account by the described model, as it is strictly dependent on the measurement technique and is likely not to be significant in full-scale systems.

Equation (4) can be used to compute the random (diffuse) incidence absorption coefficient, α_r, with the assumption that sound incidence follows the cosine law,

$$\alpha_\theta = \frac{4\,\mathrm{Re}(z_{\mathrm{MPPA},\theta})\cos\theta}{[1+\mathrm{Re}(z_{\mathrm{MPPA},\theta})\cos\theta]^2 + [\mathrm{Im}(z_{\mathrm{MPPA},\theta})\cos\theta]^2}, \tag{5}$$

$$\alpha_r = 2\int_0^{\pi/2} \alpha_\theta \cos\theta \sin\theta d\theta. \tag{6}$$

The above expression can be used to compute the diffuse incidence absorption coefficient in the 18 third-octave bands between 100 and 5000 Hz, $\alpha_{r,i}$, so that a single-number rating absorption performance of a noise barrier, DL_α, can be determined according to the European standard EN 1793-1,[10]

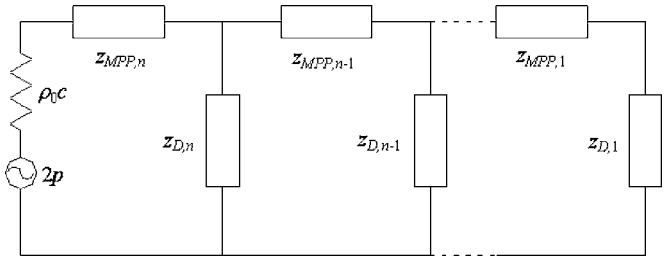


FIG. 2. Electro-acoustic equivalent circuit of a multiple system with n perforated panels (n-MPPA).

$$DL_\alpha = -10 \log_{10} \left| 1 - \frac{\sum_{i=1}^{18} \alpha_{r,i} 10^{0.1L_i}}{\sum_{i=1}^{18} 10^{0.1L_i}} \right|. \tag{7}$$

The above quantity represents the difference between the incident and reflected sound-pressure levels, in decibels, for the normalized traffic noise source with spectrum L_i, as defined by EN 1793-3.[10] The single-number rating was employed here to optimize the geometrical parameters, as it is used in Europe as a criterion to classify the acoustical performance, according to the standards concerning noise barriers to be installed along transport infrastructures.

The effective absorption range of a resonant system with one perforated layer is still too narrow to be suitable against broadband noise sources (e.g., road and railway traffic). In order to enhance the acoustic absorption of the MPP, Maa[4] introduced the concept of the double resonator: two perforated layers are mounted with an air gap in between, while the inner layer is in front of an acoustically rigid surface. This can be extended to an arbitrary number of perforated layers, n, separated by air gaps and with a rigid surface at one side (n-MPPA, Fig. 2). The surface acoustic impedance for such a system, $z_{n\text{-MPPA}}$, is calculated with the following recursive formula:

$$\begin{aligned} z_{n-\text{MPPA}} &= z_{\text{MPP},n} + \left(\frac{1}{z_{D,n}} + \frac{1}{z_{n-1}} \right)^{-1}, \\ &\cdots \\ z_2 &= z_{\text{MPP},2} + \left(\frac{1}{z_{D,2}} + \frac{1}{z_1} \right)^{-1}, \\ z_1 &= z_{\text{MPP},1} + z_{D,1}, \end{aligned} \tag{8}$$

where $z_{\text{MPP},i}$ is the surface impedance of the ith perforated layer, computed by Eq. (2), and $z_{D,i}$ is the cavity reactance of the ith air cavity with width D_i. The electro-acoustic analogy used here is not fully correct to calculate the impedance of a multiple-layer absorber, as it assumes that each air cavity is loaded by a rigid surface impedance.[11] Such a limitation can be overcome by using the acoustic transmission analysis (ATA) described by Lee *et al.*,[11] which takes into account the effective loading impedance, $z_{c,i}$, of each air gap. The equivalent formula of Eq. (8) developed by the ATA approach is the following:

TABLE I. Geometrical characteristics of the single perforated layer samples.

Sample code	r_0 [mm]	t [mm]	p [%]
MPP1	0.5	2	5.1
MPP2	0.5	2	2.1
MPP3	0.5	2	0.8

$$\begin{aligned} z_{n-\text{MPPA}} &= z_{\text{MPP},n} + \frac{z_{c,n-1} \cosh(jkD_n) + \sinh(jkD_n)}{z_{c,n-1} \sinh(jkD_n) + \cosh(jkD_n)}, \\ &\cdots \\ z_{c,2} &= z_{\text{MPP},2} + \frac{z_{c,1} \cosh(jkD_2) + \sinh(jkD_2)}{z_{c,1} \sinh(jkD_2) + \cosh(jkD_2)}, \\ z_{c,1} &= z_{\text{MPP},1} + z_{D,1}. \end{aligned} \tag{9}$$

Equation (9) assumes the continuity of particle velocity through the perforated layers:[11] this can be considered reasonably accurate when the thickness is much smaller than the wavelength, as in the present case. Obviously, the two methods provide equivalent results for a single-layer configuration. In the case of an oblique incidence wave the surface acoustic impedances of the cavities in Eqs. (8) and (9) have to be modified accordingly to the adopted propagation assumption in the cavity, local or nonlocal (extended) reaction. All the predictions for multiple layer configurations reported in Sec. IV are performed by the acoustic transmission analysis. A multiple resonator solution is able to widen the effective absorption frequency range mainly towards the lower frequencies but also to the higher one. The acoustic designer has to look for the best compromise between the increased acoustic performance of a multiple system and the rise of costs and complexity. Moreover, it has to be considered that multiple transparent layers undoubtedly reduce the overall optical characteristics of the panel (see Sec. V).

IV. EXPERIMENTAL RESULTS FOR NORMAL AND DIFFUSE INCIDENCE

Several specimens of perforated panels manufactured from polycarbonate sheets [density 1200 kg/m^3, thermal conductivity 0.20 W/(m K)] were tested. The choice of such polymeric material is justified by its good mechanical and optical properties, chemical stability, and its commercial availability. Table I reports the main characteristics of the samples. The hole radius was kept constant (0.5 mm), being constrained by the employed manufacturing method (punching). While, according to Maa's theory,[4] the optimum design of an MPP is achieved with a thickness of the order of the hole diameter, in this study the layer thickness could not be too small to assure a good mechanical strength to the panel (for common panel dimensions and without internal partitions, a minimum thickness of 2 mm was estimated for the employed material).

Measurements were initially performed using the transfer function method described in ISO 10534-2.[12] The experimental apparatus was a B&K type 4206 standing wave tube.

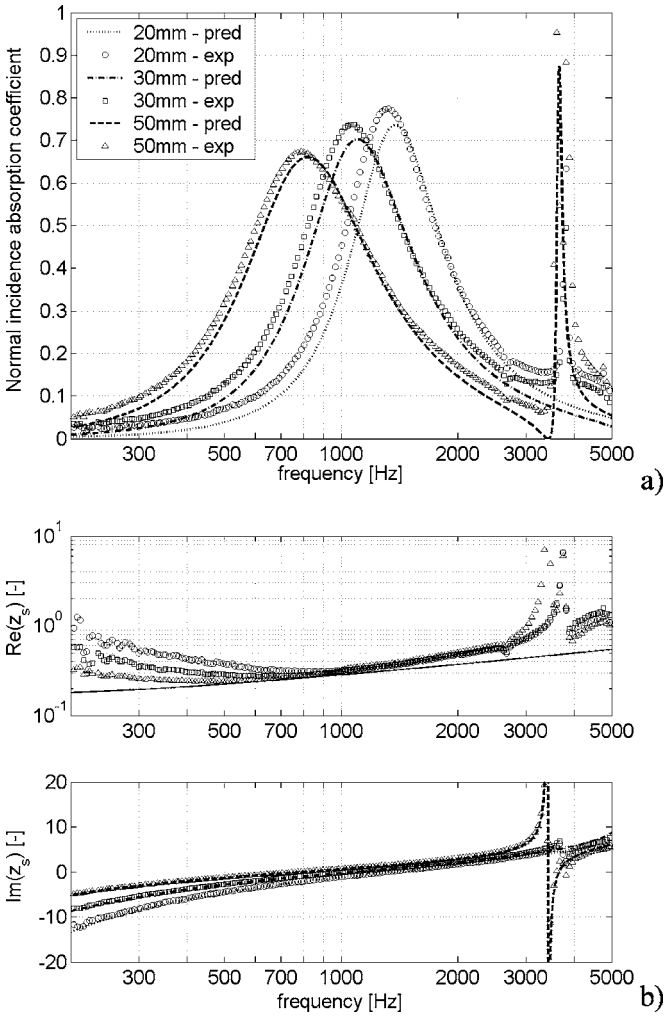


FIG. 3. Normal incidence absorption coefficient (a) and normalized surface impedance (b) for sample MPP1 (see Table I) mounted at different distances from a rigid surface: comparison between theoretical prediction and impedance tube experimental data.

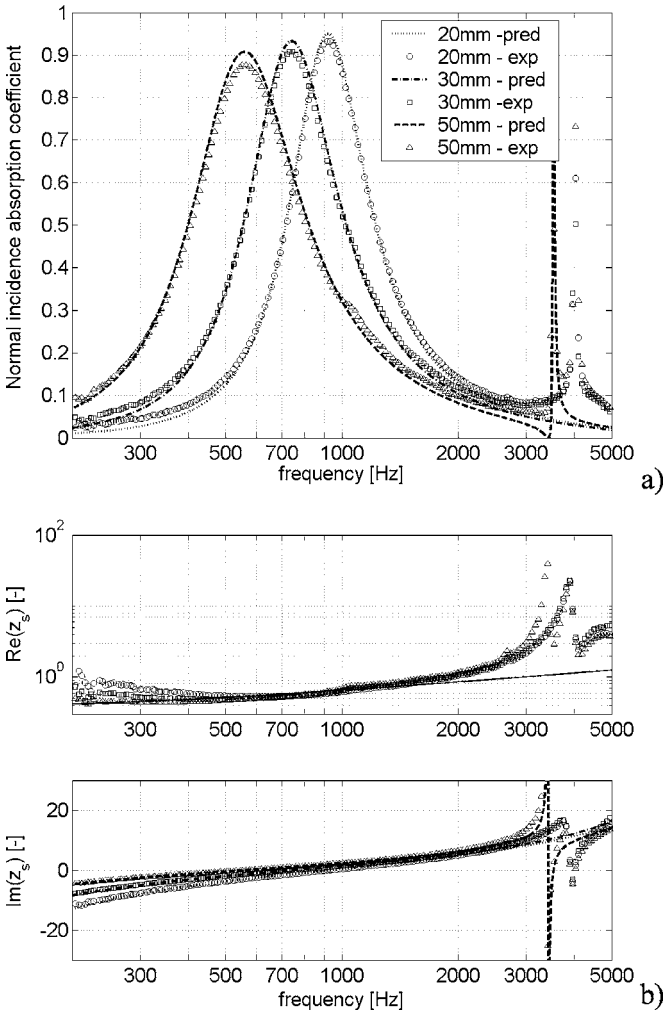


FIG. 4. Normal incidence absorption coefficient (a) and normalized surface impedance (b) for sample MPP2 (see Table I) mounted at different distances from a rigid surface: comparison between theoretical prediction and impedance tube experimental data.

The measurement setup included a multichannel acquisition system B&K PULSE type 3560, two B&K type 4939 1/4 in. condenser microphones with type 2670 preamplifier, B&K type 7206 signal amplifier, and a PC equipped with a specifically designed software. Considering the tube internal diameter (29 mm) and the microphone spacing (20 mm), the expected working frequency range of the apparatus was between 200 and 6400 Hz. Throughout the measurements, sound-pressure levels within the tube were kept sufficiently low (below 100 dB) to avoid nonlinear effects.

Figures 3–5 present the comparisons among numerical and experimental data in terms of the normalized surface impedance and normal incidence absorption coefficient. The obtained agreements in terms of absorption coefficient appear satisfying for the range of air gaps considered in this work. The data suggest that the discrepancy between measurements and predictions noticeable around 4 kHz could be explained by the fluid-structural coupling between the specimen and the backing cylindrical cavity in the impedance tube.[9] All the samples were cut using the same rotating blade and had the diameter slightly larger than the sample holder. This guaranteed that the samples were mounted firmly in the tube, avoiding any side gaps and having the mechanical boundary conditions at the edges close to clamped.

The measured and predicted data for the imaginary part of the surface impedance show very good agreement (e.g., see Fig. 3). The imaginary part in the surface impedance is mainly controlled by the size of the air gap and this effect is accurately captured by the presented model. On the other hand, the real part of the impedance is mainly linked to visco-thermal effects in the perforations and its behavior is better captured in the case of samples with lower values of porosity (e.g., see Fig. 3 in comparison with Fig. 5). This may be due to the interference between neighboring holes which may become pronounced particularly in the low-frequency regime. Some significant discrepancies can be noticed at high frequencies (e.g., see Fig. 5), where the mechanical impedance of the layer become comparable with the acoustical impedance because of the coupling effect; this effect also depends on the value of the acoustical impedance itself.[9]

Figure 6 shows a comparison between theoretical prediction and experimental data for a system with three perforated panels (sample TMPP1, see Table II for geometrical

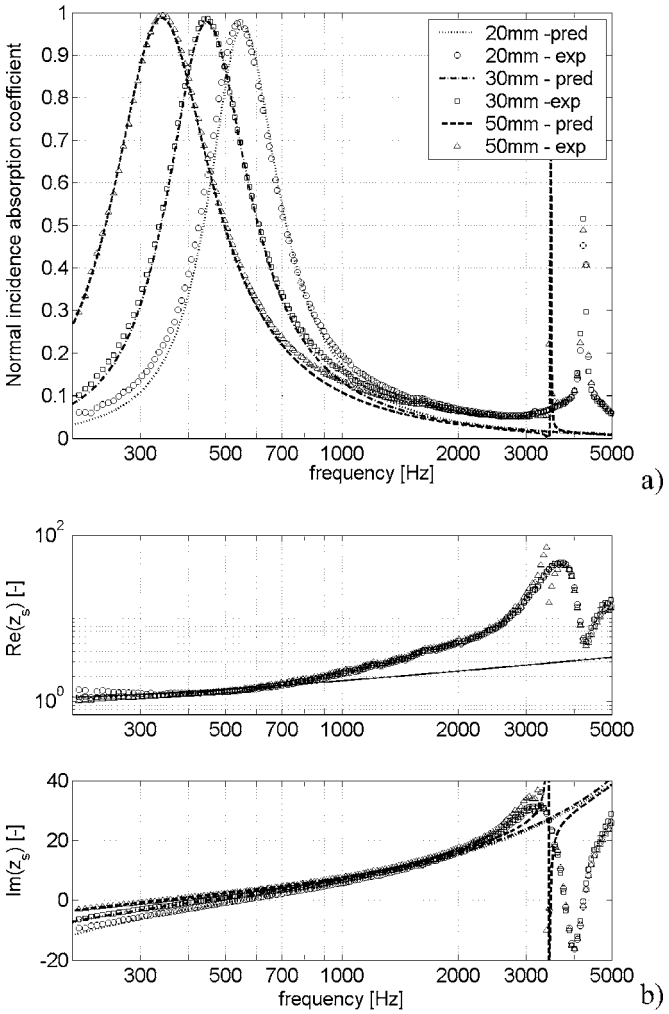


FIG. 5. Normal incidence absorption coefficient (a) and normalized surface impedance (b) for sample MPP3 (see Table I) mounted at different distances from a rigid surface: comparison between theoretical prediction and impedance tube experimental data.

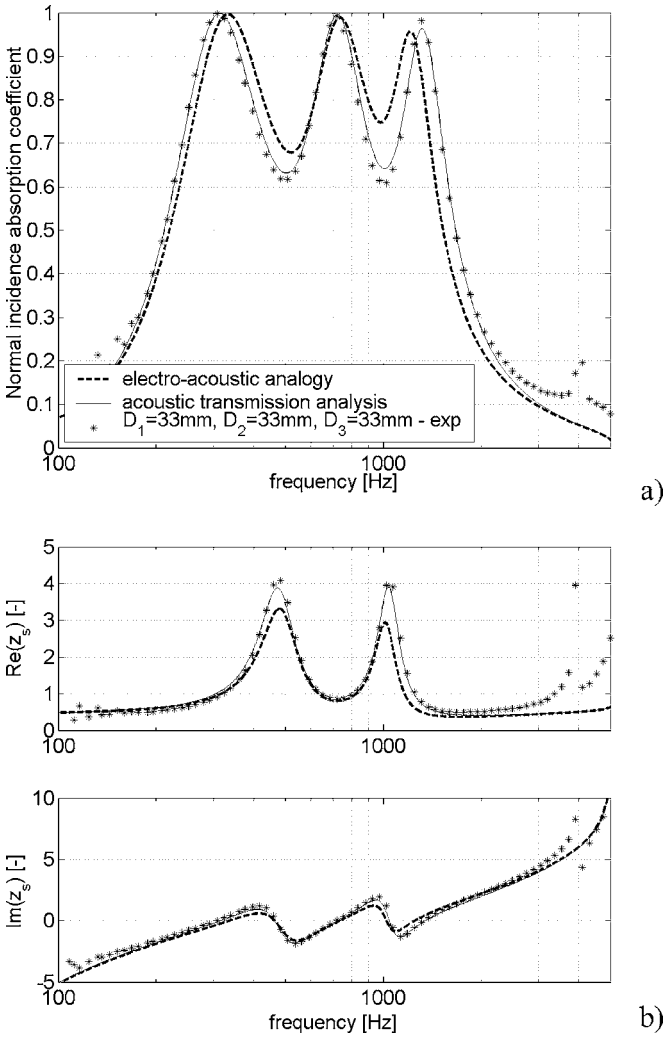


FIG. 6. Normal incidence absorption coefficient for sample TMPP1 (see Table II): comparison between theoretical predictions and impedance tube experimental data.

parameters), proving that as expected the ATA model can be considered more reliable than the electro-acoustic analogy for multilayered configurations. It has also been noticed that, through an optimization of the geometrical parameters, it is possible to extend the effective absorption frequency range to three octaves and place it within the most interesting frequency regime (between 400 and 2500 Hz for the normalized road traffic noise[10]) to achieve the maximum value of the single-number rating, $DL_{\alpha,}$.

In the present study, the hole radius and the layer thickness were fixed for the mentioned technological reasons; therefore, the investigation was carried out on the effects of the remaining parameters, i.e., number of layers, porosities, and air gap widths. It was noticed that a decreasing porosity of the perforated panels from the source side to the rigid backing can considerably increase the acoustic performance. This can be expected because of the positive gradient of the material's flow resistance.[11] The width of air gaps has strong influence on the positions of the resonance peaks, which implies that the single-number rating of sound absorption DL_α shows a local maximum depending on the air gap widths. Anyway, the analysis performed on a two-layer system with similar porosities (Fig. 7) demonstrates that the single-number rating is only slightly changing over a rather large range of widths. Similar results have been observed for three-layer systems. Therefore, it has been predicted that, with the mentioned geometrical parameters, a three-layer configuration with decreasing porosities can be able to achieve top-range performance, without excessive mounting complications.

Full-scale prototypes of transparent sound-absorbing panels were then designed, on the basis of the modeling procedure described in Sec. III, and manufactured in order to

TABLE II. Geometrical characteristics of the perforated layers in multiple layer samples (layers are numbered from receiver to source side).

Sample code	Layer no.	r_0 [mm]	t [mm]	p [%]
TMPP1	1	0.5	2	0.8
	2	0.5	2	2.1
	3	0.5	2	5.1
TMPP2	1	0.5	3	0.8
	2	0.5	3	2.1
	3	0.5	3	5.1

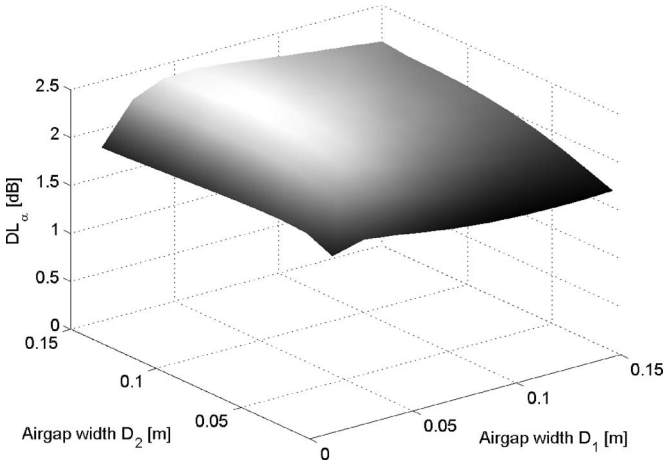


FIG. 7. Calculated dependence of the single-number rating, DL_α[dB] on the air gap widths for a system with two perforated layers. Geometrical parameters: r_1=0.5 mm, t_1=3 mm, p_1=2.1%, r_2=0.5 mm, t_2=3 mm, p_2=5.1%.

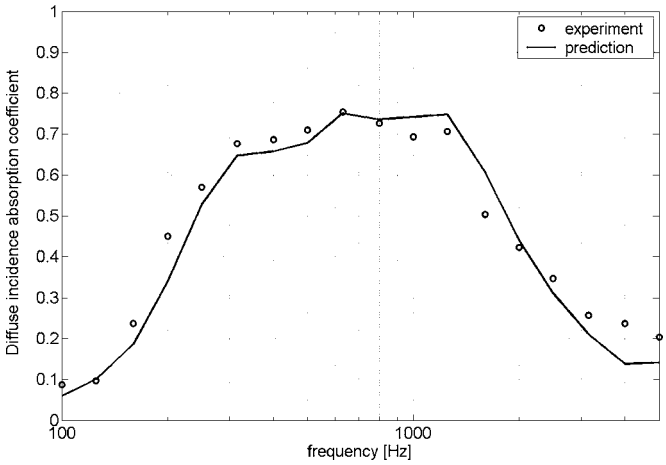


FIG. 9. One-third-octave band diffuse incidence absorption coefficients for sample TMPP2 (see Table II, $D_1=D_2=D_3=40$ mm). Comparison between theoretical prediction (nonlocally reacting cavity) and reverberation room experimental data.

achieve the acoustical performance comparable with that observed in the case of standard, nontransparent products. Figure 8 shows a prototype of a noise barrier panel constituted by three 3-mm-thick perforated layers and one 5 mm-thick solid layer, all made up of polycarbonate (sample TMPP2; see Table II).

The prototypes were tested in the reverberation room of the Acoustics Laboratory of the University of Perugia. The measurement procedure follows the interrupted noise method described by the standard ISO 354[13] even though the volume of the employed reverberation room (120 m^3) is below the standard limit. Measures have been taken to guarantee a sufficient accuracy: the floor surface area covered by the test specimens (6.4 m^2)) and the sound-diffusing treatment of the room (4.2 m^2 of suspended gypsum-board plane diffusers) were chosen accordingly. A two-channel data acquisition system was used to record the sound-pressure decays and to drive an omnidirectional (dodecahedral) source with random noise. Four source positions were chosen; for each source position, four microphone positions were employed and the acquisition repeated four times, resulting in 64 measurements per session as an outcome. The decay curves recorded in each single source-microphone position were averaged; then, the reverberation times in one-third-octave bands were estimated with a PC-controlled interpolation procedure from the averaged decay curve (starting decay level: −5 dB, dynamic range: 20 dB). The reverberation times computed for the different positions were finally averaged. Through several repetitions of the measurements the relative standard uncertainty on the absorption coefficient within the frequency range 315-3150 Hz was estimated below 10%.

FIG. 8. Prototype of the transparent sound-absorbing panel employed for the reverberation room tests. Panel dimensions are 1.04×2.05 m.

Figure 9 reports a comparison between the experimental data and the theoretical prediction for sample TMPP2 (Table II); three panels (each of size 1.04×2.05 m) were manufactured with sound-reflecting edges. Prediction was calculated by the acoustic transmission analysis and assuming a nonlocally reacting air cavity. A slight underestimation of the calculated data occurs below 800 Hz, probably due to the partitioning effect of the boundary edges of the panel. Nonetheless, a rather good agreement can be noticed: an average single-number rating of 3.9(±0.6) dB was experimentally observed, with a discrepancy from the prediction of about 0.1 dB.

Partitioning the air cavities and increasing the sound-absorption properties of the edges may significantly affect the performance and also contribute to enhance the overall panel absorption. Figure 10 theoretically demonstrates such a

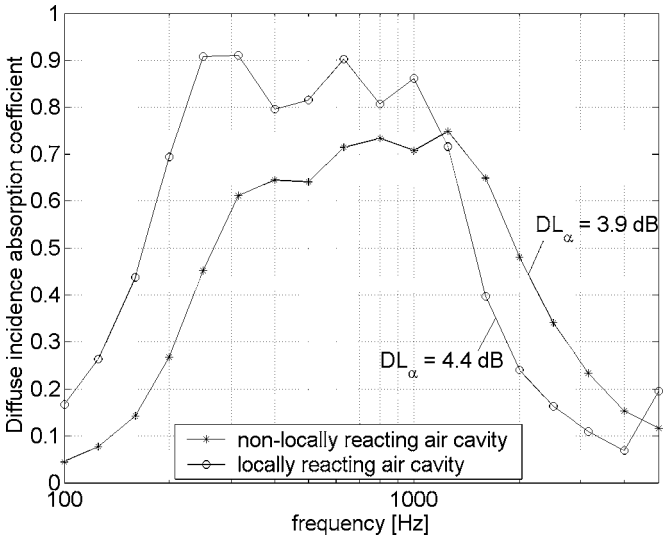


FIG. 10. Comparison between theoretical predictions for diffuse incidence sound absorption employing the assumptions of locally and nonlocally reacting cavity. Geometrical parameters are those of Fig. 9.

possibility for a three-layer configuration corresponding to that of sample TMPP2: the single-number rating shows an increase of 0.5 dB in the presence of a partitioned cavity. This is also confirmed by the experimental results recently reported by Yairi *et al.* on a MPP backed by a paper honeycomb panel.[14] This has to be taken into account for the design stage of a full barrier.

V. EVALUATION OF THE OPTICAL PROPERTIES

A separate experimental campaign aimed at evaluating the optical performance, namely the transmittance and reflectance in the visible range, was carried out on a three-layer configuration (sample TMPP1; see Table II) to verify the feasibility of achieving acceptable acoustical and optical properties at the same time.

A calibrated spectrophotometer Varian model Cary 2300 was used together with an integrating sphere in the wavelength range 300–800 nm, with a 10-nm wavelength step.[15] For each sample, the measure was repeated in three different positions and the average over the three series of data calculated.

Figure 11 reports a comparison between the normal incidence optical transmittance measured on the multiple layer configuration and on a single 5-mm nonperforated polycarbonate layer. The triple MPP shows the expected effect of lowering the optical transparency with respect to the single layer. However, the transmittance factor, τ_v, in the visible range,[16] which is an average of the measured transmittance, $\tau_v(\lambda)$, weighted by the variation of human sensitivity within the 380–780-nm range of light wavelengths, can still be considered adequate (a value of 0.52 has been measured), at least in terms of visual impact of the panel. No significant deviation was observed between data acquired in different positions, showing the small influence of perforations on the normal incidence transparency. The measured reflectance for the multiple layer configuration was very low (below 8% in the whole range), avoiding any possible glaring phenomena.

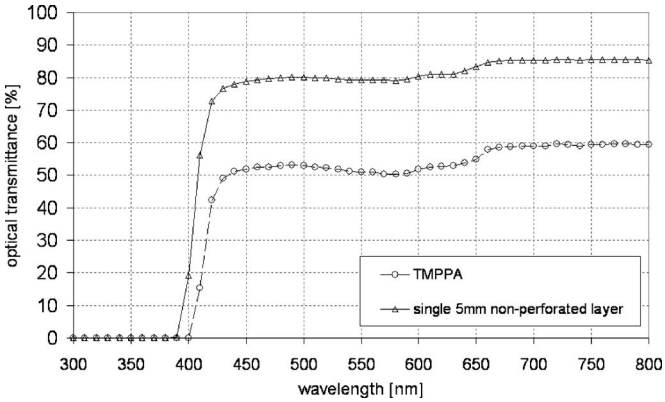


FIG. 11. Optical transmittance in the visible range vs wavelength of sample TMPP1 (see Table II), compared with a single 5-mm nonperforated layer.

VI. CONCLUSIONS

There is an increasing demand for multifunctional design solutions for noise barriers, especially for barriers along transport infrastructures: sound absorption, optical transparency, durability, and lightness are among those. The use of transparent materials is very common whenever aesthetical and visual impact issues are of concern. While ensuring good sound-insulating performances, transparent panels are of little help when sound-absorbing properties are also requested; this is not uncommon for noise barriers to be installed along both sides of roads and railways, close to high buildings.

The paper demonstrates the technical feasibility of manufacturing noise barrier panels using perforated transparent layers in a multiple resonator configuration without porous materials, in order to achieve acoustical sound-absorbing performance similar to those of commercial nontransparent products, together with rather good optical characteristics. The well-established theory of microperforated absorbers—along with the electro-acoustic analogy and the acoustic transmission analysis—was adopted to design panels with optimized sound-absorbing properties with respect to traffic noise, in terms of the single-number rating DL_α used in the European countries for noise barrier quality assessment. The optimization was carried out taking into account the technological constraints of mechanical strength of the panels, which has the main effect of limiting the minimum thickness.

A series of experiments was carried out, both in an impedance tube and in a reverberation room, on single and multiple perforated layer systems made of perforated polycarbonate sheets. The results were compared with theoretical predictions, achieving good agreement, at least for engineering purposes. The predictions showed the potential of a three-layer configuration with layers of decreasing porosity (starting from the source side) and evidenced the limited effect of the air gap widths on the single-number rating.

The measured absorption performance of an optimized configuration with three 3-mm perforated layers (4 dB for DL_α) is comparable with that of standard commercial non-

transparent panels of similar thickness (13 cm). At the same time, adequate optical characteristics as far as light transparency can still be observed.

ACKNOWLDGMENTS

The authors are indebted to Vincenzo Tognaccini and Roberto Pignattini for sample manufacturing, and all the team at the Laboratory of Acoustics of the University of Perugia for the support provided during the measurements in the reverberation room. The spectrophotometer data were kindly provided by Giorgio Baldinelli.

[1]H. V. Fuchs and X. Zha, "Acrylic-glass sound absorbers in the plenum of the Deutscher Bundestag," Appl. Acoust. **51**(2), 211–217 (1997).
[2]J. Kang and M. W. Brocklesby, "Feasibility of applying micro-perforated absorbers in acoustic window systems," Appl. Acoust. **66**(6), 669–689 (2005).
[3]G. R. Watts and N. S. Godfrey, "Effects on roadside noise levels of sound absorptive materials in noise barriers," Appl. Acoust. **58**(4), 385–402 (1999).
[4]D.-Y. Maa, "Theory and design of microperforated-panel sound-absorbing construction," Sci. Sin. **18**, 55–71 (1975).
[5]T. H. Melling, "The acoustic impedance of perforates at medium and high sound pressure levels," J. Sound Vib. **29**(1), 1–65 (1973).
[6]L. J. Sivian, "Acoustic impedance of small orifices," J. Acoust. Soc. Am. **7**(2), 94–101 (1935).
[7]M. Q. Wu, "Micro-perforated panels for duct silencing," Noise Control Eng. J. **45**, 69–77 (1997).
[8]J. Kang and H. V. Fuchs, "Predicting the absorption of open weave textiles and micro-perforated membranes backed by an air space," J. Sound Vib. **220**(5), 905–920 (1999).
[9]Y. Y. Lee, E. W. M. Lee, and C. F. Ng, "Sound absorption of a finite flexible micro-perforated panel backed by an air cavity," J. Sound Vib. **287**(1–2), 227–243 (2005).
[10]EN 1793, "Road traffic noise reducing devices—Test method for determining the acoustic performance," Parts 1 and 3, European Committee for Standardization, Brussels, Belgium, 1997.
[11]F.-C. Lee and W.-H. Chen, "Acoustic transmission analysis of multi-layer absorbers," J. Sound Vib. **248**(4), 621–634 (2001).
[12]ISO 10534-2, "Acoustics—Determination of sound absorption coefficient and impedance in impedance tubes—Part 2: Transfer-function method," International Organisation for Standardisation, Geneva, Switzerland, 1998.
[13]ISO 354, "Acoustics—Measurement of sound absorption in a reverberation room," International Organisation for Standardisation, Geneva, Switzerland, 2003.
[14]M. Yairi, K. Sakagami, M. Morimoto, and A. Minemura, "Acoustical properties of microperforated panel absorbers with various configurations of the back cavity," on the CD-ROM: *The 12th International Congress on Sound and Vibration (ICSV12)*, 11–14 July 2005, Lisbon, Portugal, available from the Instituto Superior Técnico (IST).
[15]F. Asdrubali, G. Baldinelli, L. Frezzini, "Evaluation of the energy and lighting performances of innovative transparent materials in buildings," on the CD-ROM: *The 10th European Lighting Conference Lux Europa 2005*, Berlin, Germany, 2005.
[16]EN 410, "Glass in building—Determination of luminous and solar characteristics of glazing," European Committee for Standardization, Brussels, Belgium, 1998.

Local feedback control of light honeycomb panels

Chinsuk Hong[a] and Stephen J. Elliott
Institute of Sound and Vibration Research, University of Southampton, University Road, Southampton, Hampshire, SO17 1BJ, United Kingdom

(Received 31 January 2006; revised 18 September 2006; accepted 6 October 2006)

This paper summarizes theoretical and experimental work on the feedback control of sound radiation from honeycomb panels using piezoceramic actuators. It is motivated by the problem of sound transmission in aircraft, specifically the active control of trim panels. Trim panels are generally honeycomb structures designed to meet the design requirement of low weight and high stiffness. They are resiliently mounted to the fuselage for the passive reduction of noise transmission. Local coupling of the closely spaced sensor and actuator was observed experimentally and modeled using a single degree of freedom system. The effect of the local coupling was to roll off the response between the actuator and sensor at high frequencies, so that a feedback control system can have high gain margins. Unfortunately, only relatively poor global performance is then achieved because of localization of reduction around the actuator. This localization prompts the investigation of a multichannel active control system. Globalized reduction was predicted using a model of 12-channel direct velocity feedback control. The multichannel system, however, does not appear to yield a significant improvement in the performance because of decreased gain margin.
 [DOI: 10.1121/1.2384843]

PACS number(s): 43.50.Ki, 43.40.Vn, 43.40.Rj [KAC] Pages: 222–233

I. INTRODUCTION

Interior cabin noise control is a crucial issue in aircraft design since the noise affects the passenger comfort directly (Mixson and Powell, 1984; Mixson and Wilby, 1995). The high levels of noise and vibration are generated by many sources, such as engines, propellers, jets, airflow, and so on (Gardonio, 2002; Gardonio and Elliott, 1999; Maury *et al.*, 2002). The majority of the noise and vibration is transmitted through the fuselage and the trim panel into the cabin. Many passive reduction techniques are employed, such as barriers and absorption materials for the airborne noise, and vibration isolators and damping materials for the structure-borne noise. In order to achieve the best performance, based on the cost effectiveness and the weight requirement, a combination of the reduction of the airborne noise and the structure-borne noise must be achieved. These conventional techniques are of limited use at low frequencies, however, particularly with the weight requirements in aircraft. From the classification of the main sources of interior noise (Bhat, 1999; Mixson *et al.*, 1978; Pope *et al.*, 1987a, b; Sulc *et al.*, 1982) and the identification of the transmission paths of airborne noise and structure-borne noise (Mixson *et al.*, 1978; Pope *et al.*, 1987a, b; Unruh, 1988, 1989), the performance and limitations of the passive treatments are well-known (Barton and Mixson, 1981; Vaicatis, 1980).

Passenger comfort is still influenced by noise at low frequencies, and several approaches have been investigated for controlling this low-frequency noise actively (Fuller *et al.*, 1996; Gardonio, 2002; Gardonio and Elliott, 1999; Nelson and Elliott, 1992). This work started with active noise control (ANC) of the cabin, followed by active vibration control (AVC), and then active structural acoustic control (ASAC) of the fuselage and/or the trim panel. It is believed that AVC and ASAC can achieve global noise reduction with much lower dimensionality than ANC. When actuators are placed on the fuselage, however, they can cause damage to the structure due to fatigue. To avoid this problem, placing the actuators on noncritical structures, for instance, interior trim panel of aircraft, has also been studied.

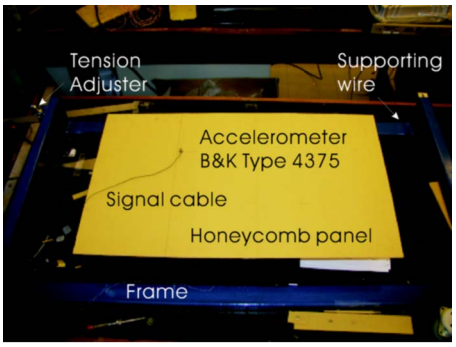
The trim panel in the aircraft is connected to the fuselage with resilient mounts. They are generally honeycomb structures because of the design requirements for high load capacity (Nilsson and Nilsson, 2002) and lightweight (Saito *et al.*, 1997). Honeycomb structures are also used in many other engineering applications, such as automobiles, railways (Saito *et al.*, 1997), and underwater vehicles. Their high bending stiffness, in combination with their light weight (Nilsson and Nilsson, 2002), however, leads to an increase in their vibration and sound transmission/radiation properties and conventional passive treatments are of limited use. The active control of trim panels thus remains an open and important research area.

The purpose of this study is to begin to study the control of noise and vibration for resiliently mounted honeycomb structures. Although practical trim panels have a complicated geometric shape, a simple lightweight honeycomb panel system is studied here for simplicity.

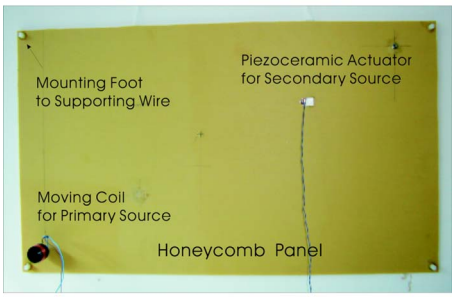
II. PANEL SYSTEM STUDIED

The panel system studied in this paper is shown in Fig. 1 and consists of a resiliently mounted light honeycomb panel, a piezoceramic actuator, and an accelerometer closely spaced on the opposite side of the panel. The parameters of the honeycomb panel are summarized in Table I. The piezoceramic actuator is bonded on the lower skin of the panel at

[a]Electronic mail: csh@isvr.soton.ac.uk



(a)



(b)

FIG. 1. (Color online) Resiliently mounted honeycomb panel under a single-channel active feedback control system. (a) panel system; (b) the panel under the test having plastic feet at the corners, the coil for the primary source at (0.05, 0.05) and sensor/piezoceramic actuator at $(0.7L_x, 0.7L_y)$.

$(0.7L_x, 0.7L_y)$, with the coordinates taken with respect to the lower left-hand corner of Fig. 1(b). The dimension of the piezoceramic patch is 13×13 mm^2, the thickness 1 mm, and its mass is 1.3 g. The accelerometer (B&K 4375), whose mass is 2.5 g, is mounted on the upper skin of the panel at the same location of the piezoceramic actuator as shown in Fig. 1. The resilient mounts supporting the honeycomb panel are achieved using the tensioned wires shown in Fig. 1(a). The plastic feet located at the coordinates (0.01, 0.01), $(L_x - 0.01, 0.01)$, $(L_x - 0.01, L_y - 0.01)$, and $(0.01, L_y - 0.01)$ shown in Fig. 1(b) are used for mounting the panel on the wires. A moving coil is attached on the panel for a primary force disturbing the panel. To excite as many modes as possible, it is positioned near the corner, at (0.05,0.05).

III. THEORETICAL MODELING

The theoretical model takes into account the dynamics of the honeycomb panel with the lumped masses due to the moving coil for the primary source and the actuator-sensor pair since their masses are comparable to those of panel's mass. When the actuator and the sensor are positioned on the opposite skins at the same location over the panel, they are locally and dynamically coupled (Hong, 2005; Malker, 1999). The local dynamic coupling is included in the theoretical model using a single degree of freedom system (Hong, 2005). The honeycomb panel is modeled using the classical plate theory (CLPT) (Reddy, 1996), assuming each layer of the honeycomb panel to be homogeneous, which results in an equivalent thin plate neglecting the effect of the shear deformation and the rotary inertia. The equivalent panel model can be expressed as

$$D_{11}\left(\frac{\partial^4 w}{\partial x^4} + 2\frac{\partial^4 w}{\partial^2 x \partial^2 y} + \frac{\partial^4 w}{\partial^4 y}\right) + I_0 \frac{\partial^2 w}{\partial t^2} = q(x,y,t) - \frac{\partial T_{xx}(x,y,t)}{\partial x} - \frac{\partial T_{yy}(x,y,t)}{\partial y} - \frac{\partial T_{xy}(x,y,t)}{\partial x} - \frac{\partial T_{yx}(x,y,t)}{\partial y}, \quad (1)$$

where T_{xx}, T_{yy}, T_{xy}, and T_{yx} are applied moments and I_0 is given by

$$I_0 = \rho^{(c)} h_c + 2\rho^{(s)} h_s. \quad (2)$$

The equivalent bending rigidity is given by (Nilsson and Nilsson, 2002)

$$D_{11} = \frac{E}{(1-\nu^2)}\left(\frac{h_c^2 h_s}{2} + h_c h_s^2 + \frac{2h_s^3}{3}\right). \quad (3)$$

When the moments are generated by an isotropic piezoceramic actuator, $T_{xx}=T_{yy}$ and $T_{xy}=T_{yx}=0$. $q(x,y,t)$ is the sum of the externally applied force, $F_p(x,y,t)$, the interaction forces due to the mounts, $F_s(x,y,t)$, lumped masses, $F_m(x,y,t)$, and lumped mass-spring-damper single degree of freedom systems, $F_a(x,y,t)$, so that

$$q(x,y,t) = F_p(x,y,t) + F_s(x,y,t) + F_m(x,y,t) + F_a(x,y,t). \quad (4)$$

The model consists of the equivalent panel and the lumped elements due to sensors, actuators, resilient mounts, and the local coupling model as shown in Fig. 2. This model represents the active honeycomb panel, resiliently mounted at the corners, with the feedback sensors and actuators. The single degree of freedom mass-spring-damper systems are included to take into account the local behavior of sensor/actuator on the honeycomb structure. The solution of this honeycomb panel system is presented in the Appendix and evaluated in the next section.

TABLE I. Parameters of the resiliently mounted honeycomb panel.

Notation	Unit	Value	Description
L_x	m	0.7	Length of the panel
L_y	m	0.4	Width of the panel
h_s	mm	0.5	Thickness of the skin
h_c	mm	3.0	Thickness of the core
ρ	kg/m^3	253.5	Average density of the panel
$\rho^{(s)}$	kg/m^3	507	Density of the skin
$\rho^{(c)}$	kg/m^3	169	Density of the core
M_p	kg	0.212	Total mass of the panel
E	Pa	14.8×10^9	Young's modulus
ν	–	0.3	Poisson's ratio

A. Dynamics of resiliently mounted panels

The eigenvalue problem can be set up from Eq. (A7) to obtain the natural behavior of the resiliently mounted panel, that is,

$$[-\omega^2 \mathbf{M}_{gg} + \mathbf{K}_{gg}]\mathbf{w}_g = \mathbf{0}. \quad (5)$$

For the nontrivial solution for $\mathbf{w}_g$, the natural frequencies can be obtained from the eigenvalue problem, Eq. (5), and the

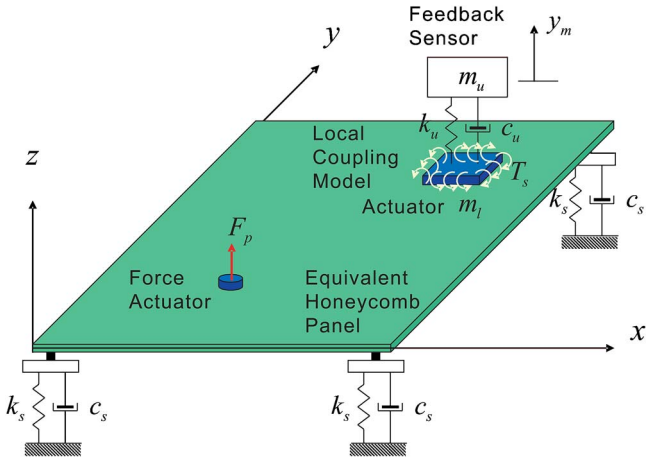


FIG. 2. (Color online) Simplified theoretical model consisting of the equivalent honeycomb panel, the lumped masses due to sensors such as accelerometers and force transducers, and the local coupling model.

mode shapes can be evaluated using the eigenvectors such that

$$\Theta_{g,pq} = \sum_m \sum_n w_{g,mn}^{(pq)} \Theta_{mn}, p, \; q = 1,2,3, \ldots , \tag{6}$$

where $\Theta_{g,pq}$ is the mode shape function of order pq of the resiliently mounted panel with the lumped elements, $w_{g,mn}^{(pq)}$ is the mnth coefficient of the pqth eigenvector, and Θ_{mn} is the mode shape function of order mn of the free-edged panel without any lumped element.

Figure 3 shows the variation of the natural frequencies with the stiffness of the mounts, k_s, denoted by the nondimensionalized parameter, $\kappa = k_s \times L_x^3/(EI)$, ranging from $\kappa = 10^{-4}$ to $\kappa = 10^6$. The dotted line and the dot-dashed line indicate the variation of the mass-spring resonance and the approximated natural frequencies of the two rotational modes, which are calculated assuming the locations of the mounts to be very near the corners. They are $\sqrt{4k_s/M}$ and $\sqrt{12k_s/M}$, respectively, when the total mass of the panel is M. The slight difference in the calculated natural frequencies

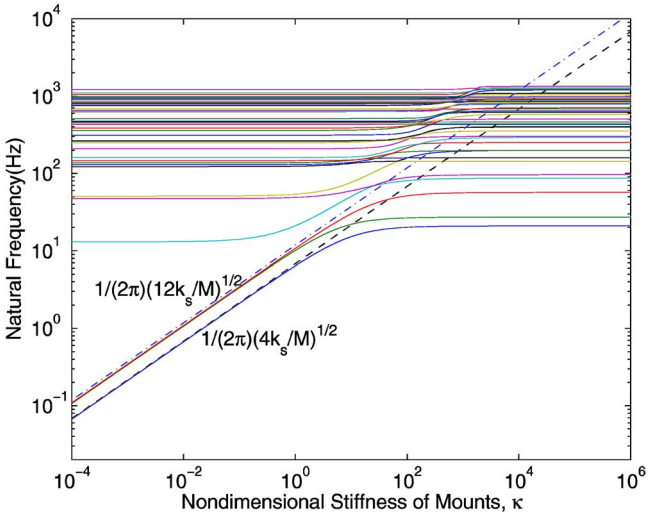


FIG. 3. (Color online) Variation of the natural frequencies with the stiffness of the mounts at the corners. The dotted line indicates the variation of the spring-mass resonance, $\sqrt{4k_s/M}$, and the dot-dashed line indicates the approximated natural frequencies of the two rotational modes, $\sqrt{12k_s/M}$.

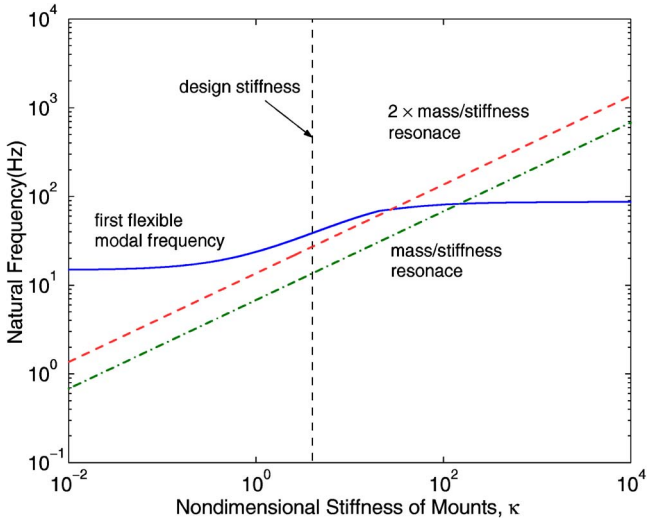
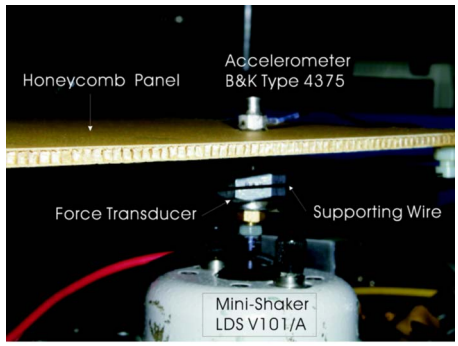
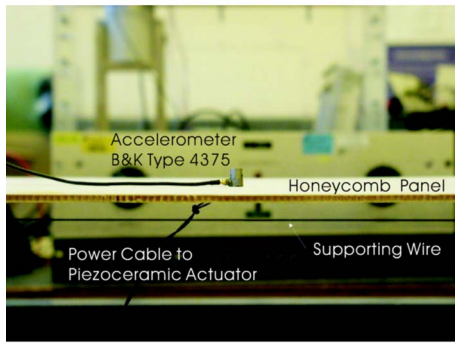


FIG. 4. (Color online) Design of resilient mounts for aircraft trim panels. The stiffness of the mounts selected for this study is $\kappa = 4$.

of the rotational modes is due to the slight deviation of the mounts' locations from the corners. It can be seen from Fig. 3 that the behaviors of the lower modes couple more easily to the mount stiffness, so that they move to those of a panel simply supported at the corners at low stiffness of the mounts less than 100($\kappa < 100$). The first three modes behave as if the panel is a rigid body when $\kappa < 1.0$ while the panel behaves with flexible modes when $\kappa > 1.0$. The natural frequencies of the same modes, which have large displacements at the corners, can cross over those of others. For example, the fourth



(a)



(b)

FIG. 5. (Color online) Measurement of the response functions of the honeycomb panel supported by wire. (a) Point mobility function measurement using a minishaker. (b) Response function measurement using a piezoceramic actuator.

mode at 13 Hz, being the first flexible mode, when the stiffness of mounts is lower than 0.1, moves to the fifth mode at 96 Hz when the stiffness is greater than 100 crossing at $\kappa = 20$.

B. Resilient mounts

The stiffness of the mounts seriously affects the stability and the control performance of the active control system when using piezoceramic actuators (Hong and Elliott, 2005). In practice, however, the mounts are typically designed to passively reduce the noise and vibration. This is achieved for aircraft trim panels by using mounts with such a stiffness that their mass-spring resonance is somewhat less than half of the first flexible panel resonance (Stothers, 2004). Figure 4 shows the design parameters of resilient mounts for aircraft trim panels. The dot-dashed line denotes the variation of the mass-spring resonance frequency and the dashed line twice this. The mount stiffness for this study is selected to be about $\kappa=4$, as indicated by the vertical dashed line in Fig. 4.

C. Local coupling and its effect

In this section the local coupling between the sensor and the actuator in the honeycomb panel is measured experimentally and its effect on the dynamics of the honeycomb panel is investigated. The accelerometer response is measured when a force is applied to the honeycomb panel first and the response to a piezoceramic actuator is then measured. An accelerometer (B&K 4375) is attached on the other side of honeycomb at the location of the actuator as shown in Fig. 5. The responses for both actuator-sensor pairs are also predicted using the CLPT model with the local coupling behavior. The local coupling behavior is modeled by a single degree of freedom mass-spring-damper system on the panel whose parameters are the mass of the accelerometer, the local stiffness, and the damping coefficient estimated from the transmissibility measurement between skins as described in Hong (2005).

Figure 6 shows the comparison of the measured frequency response with that calculated for the force excitation. The thick solid line is for the calculated result while the dashed line is the measured results, put together from measurements at different bandwidths. A reasonable agreement can be found at all frequencies. Some differences in the resonance frequencies are observed in the frequency range between 100 and 2000 Hz, where not only the bending waves but also the shear waves exist due to the effect of the shear deformation. A peak is also observed at about 4 kHz due to the local dynamics, which is accurately predicted by the model. No distinct modal behavior exists in the experimental results above this frequency because the modal density due to the shear waves becomes large enough to smooth the response due to the high modal overlapping in the real honeycomb. For the analytical model, however, the modal overlap is not high enough to entirely smooth the spectrum.

Figure 7 shows the response of the honeycomb panel excited by a piezoceramic actuator measured with an accelerometer on the opposite skin. The difference at frequencies less than 20 Hz is caused by the low signal-to-noise ratio in the experiment, due to the limited authority of the small piezoceramic actuator. Apart from that, the main characteristics are similar at each frequency range to that mentioned above. The local behavior for the piezoceramic actuator is not so distinct, however, since the local dynamic mechanism due to the piezoceramic actuation is highly damped.

IV. DIRECT VELOCITY FEEDBACK CONTROL WITH A PIEZOCERAMIC ACTUATOR

A. Plant response and stability

The effect of the local coupling between a piezoceramic actuator and a closely spaced accelerometer is to roll off the response between the actuator and sensor at higher frequency. A feedback control system should thus be able to operate with a higher gain margin than if a solid structure, e.g., an aluminum panel, was used (Hong, 2005). It is also clear that CLPT theory with the local coupling model is suit-

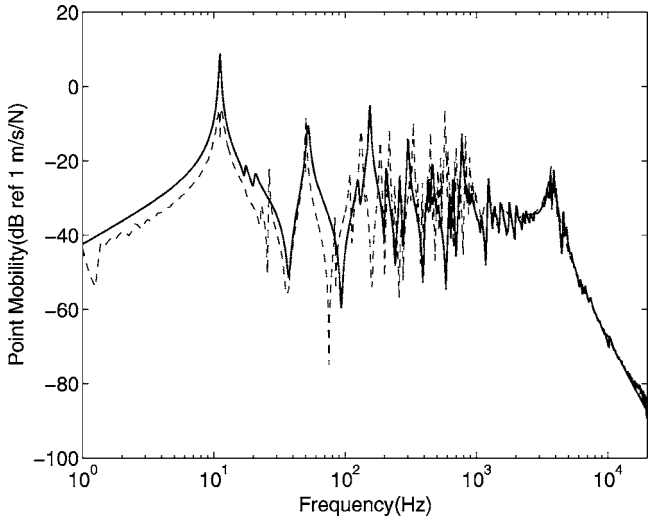


FIG. 6. Comparison of the measured point response function (dashed line) with the calculated (solid line) for the force actuator.

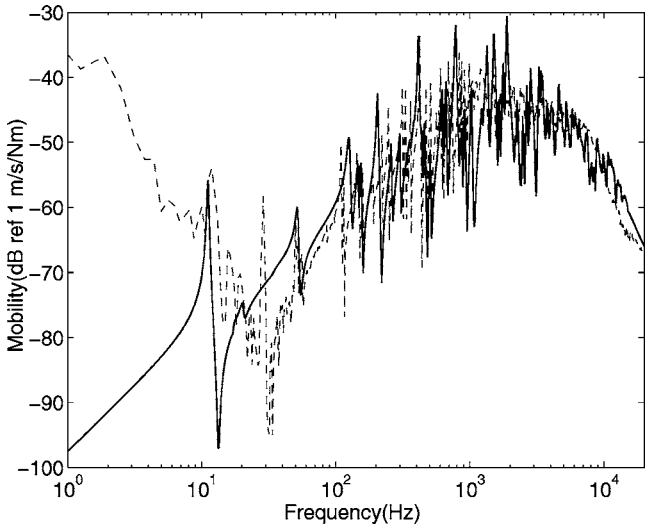


FIG. 7. Comparison of the measured response function (dashed line) with the calculated (solid line) for the piezoceramic actuator with an accelerometer on the opposite side of the honeycomb panel.

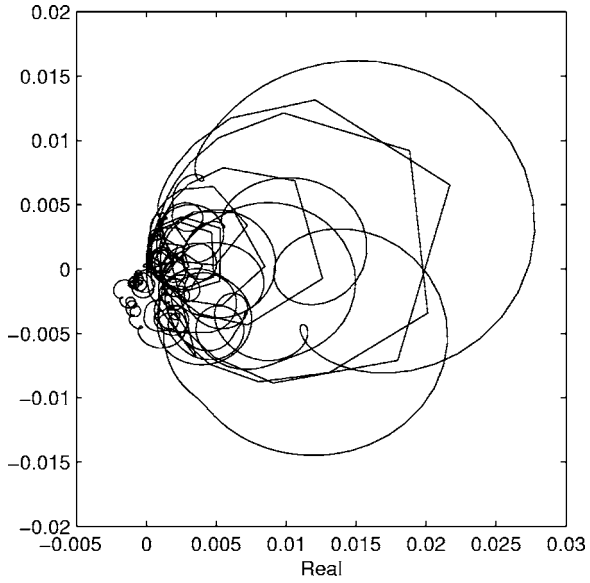


FIG. 8. Nyquist plot of the plant response up to 20 kHz when the piezoceramic actuator is placed at $(0.7L_x, 0.7L_y)$ (a) Calculated; (b) measured.

able for designing and characterizing the active control systems for active honeycomb trim panels since the local coupling cuts off the high-frequency response where the shear deformation effect and the rotary inertia effect are dominant.

Figure 8 shows the Nyquist plots of the plant response between the piezoceramic actuator and the velocity sensor calculated using the model and measured as shown in Fig. 7. The phase is shifted slowly as the frequency increases, approaching 180° at high frequencies. A large maximum gain of about 1000 is thus obtained. The loop on the left-hand side of the origin could lead to instability, and is due to two mechanisms. One is the the effect of the local dynamics. The local dynamics give an additional phase shift while it makes the magnitude of the response decrease at the frequency higher than $\sqrt{2}$ times the coupling resonance frequency. The other is due to the noncollocation of the piezoceramic actuator and the sensor, so that the size of the actuator is the key parameter to characterize it. The delay between the moment generated along the edge of the piezoceramic actuator and the sensor on the opposite skin of the panel at the center of the piezoceramic actuator leads to slowly changing phase as frequency increases.

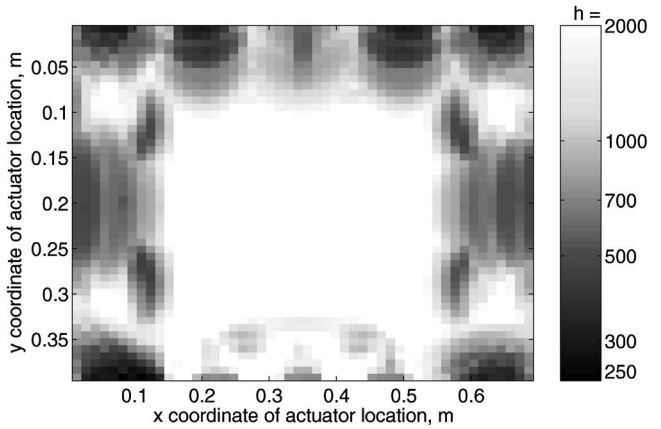


FIG. 9. Predicted variation of the maximum gain with the location of actuator.

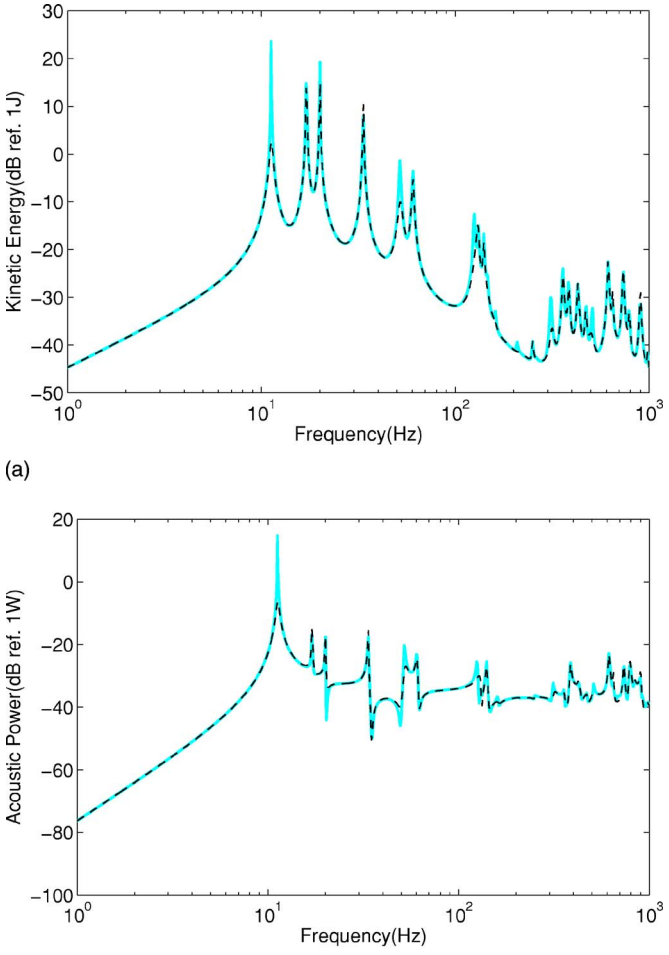
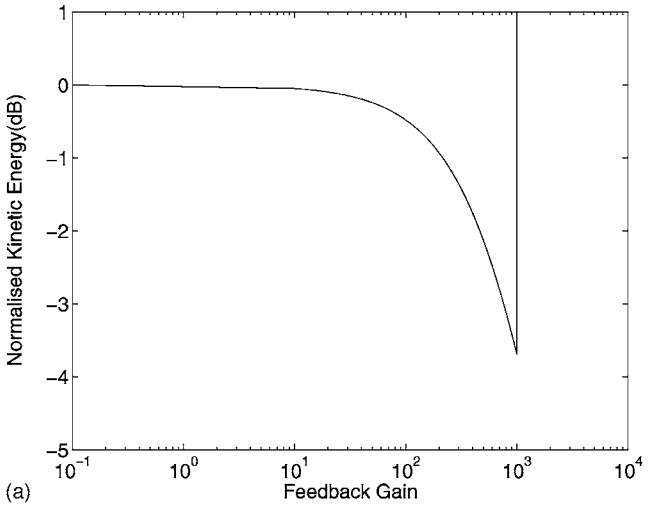


FIG. 10. (Color online) Total kinetic energy (a) and total acoustic power (b) of a resiliently mounted (κ=4) honeycomb panel disturbed by a concentrated force at (0.05, 0.05) with no control (solid line) and with a feedback gain of 1000 (dashed line), when subjected to the single-channel direct feedback control with sensor/piezoceramic actuator at $(0.7L_x, 0.7L_y)$.



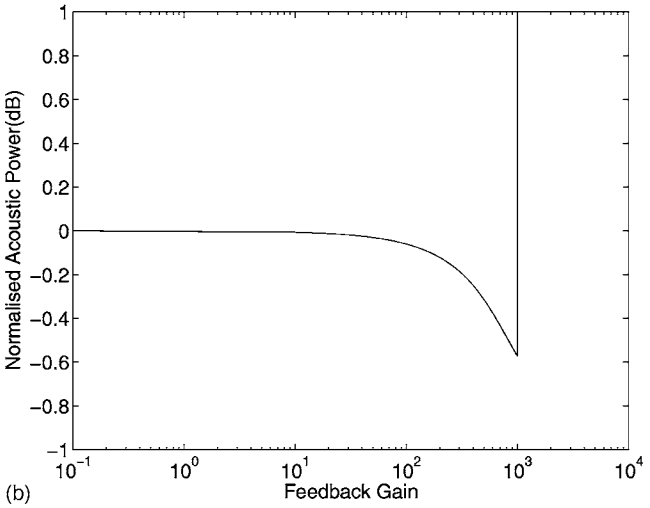
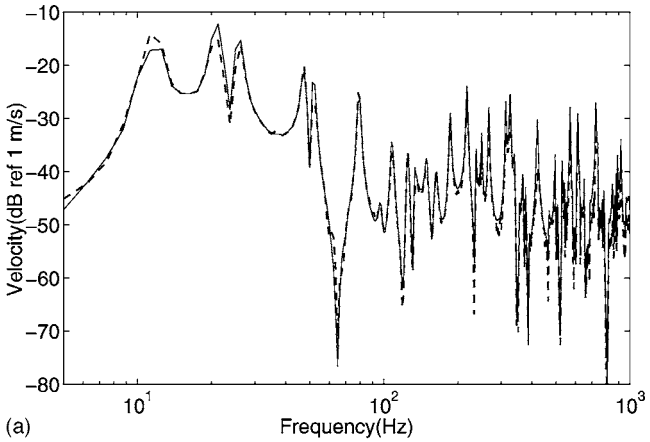


FIG. 11. Normalized control performance of a resiliently mounted (κ=4) honeycomb panel disturbed by a concentrated force at (0.05, 0.05), and when subjected to the single-channel direct feedback control with sensor/piezoceramic actuator at $(0.7L_x, 0.7L_y)$. The normalization is carried out up to 1 kHz. The vertical dashed line denotes the maximum gain. (a) normalized total kinetic energy; (b) normalized total acoustic power.

The details of the Nyquist plots and hence the maximum feedback gain vary with actuator/sensor location. Figure 9 shows the variation of the maximum gain with the location of actuator calculated from the theoretical model. The locations where the maximum gain is greater than 2000 in this case are represented by white. The smallest maximum gain in this diagram is about 250. It can be seen that the actuator should be located at least 25% of the plate width from the edges to obtain a large gain margin. The distribution of maximum gain is not symmetric on the panel because of the lumped mass and stiffness of the primary and secondary sources.

B. Control performance

The predicted performance of the feedback control system with a single-channel sensor/piezoceramic actuator system can be estimated in terms of either the total kinetic energy or the total radiated acoustic power given by Eqs. (A35) and (A36).



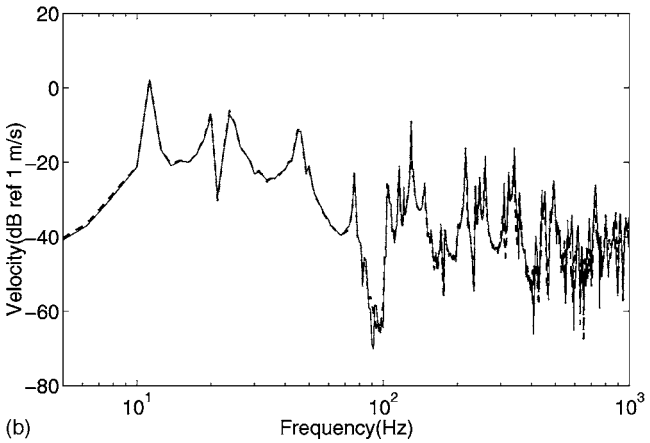


FIG. 12. Measured velocity levels up to 1 kHz (a) at the piezoceramic actuator location $(0.7L_x, 0.7L_y)$ and (b) at the point $(0.6L_x, 0.6L_y)$, 8 cm apart from the actuator location, of a resiliently mounted (κ=4) honeycomb panel disturbed by a concentrated force at (0.05, 0.05) (solid line), and when subjected to the single channel direct feedback control with the gains of about 200 (dashed line), using the piezoceramic actuator.

Figure 10 shows the predicted total kinetic energy and the radiated acoustic power of a resiliently mounted (κ=4) honeycomb panel disturbed by a concentrated force at (0.05,0.05), respectively, and subjected to the single-channel direct velocity feedback control with sensor/piezoceramic actuator at $(0.7L_x, 0.7L_y)$ with a feedback gain of 1000. Reductions in the predicted kinetic energy and acoustic power of about 20 dB can be achieved at the first modal frequency using the feedback control. Since the actuator weakly couples into the subsequent rigid body modes when the panelis supported by soft mounts, κ=4 in this case, the reductions at those frequencies are small. To understand how the piezoceramic actuator couples into those modes more intuitively, we may consider a completely free-edged panel. The modal force of the moments generated by the piezoceramic actuator at the first four modes are all zero.

Figure 11 shows the predicted normalized kinetic energy and acoustic power of the resiliently mounted honeycomb panel integrated up to 1 kHz, calculated from Eqs. (A39) and (A40). Since the control system is only conditionally stable with a maximum gain of 1000 in this case, the maximum performance is limited. The largest predicted reduction in the normalized kinetic energy and acoustic power are 4 and 0.3 dB, respectively. These relatively small reductions are due to the localization of the reduction around the actuator-

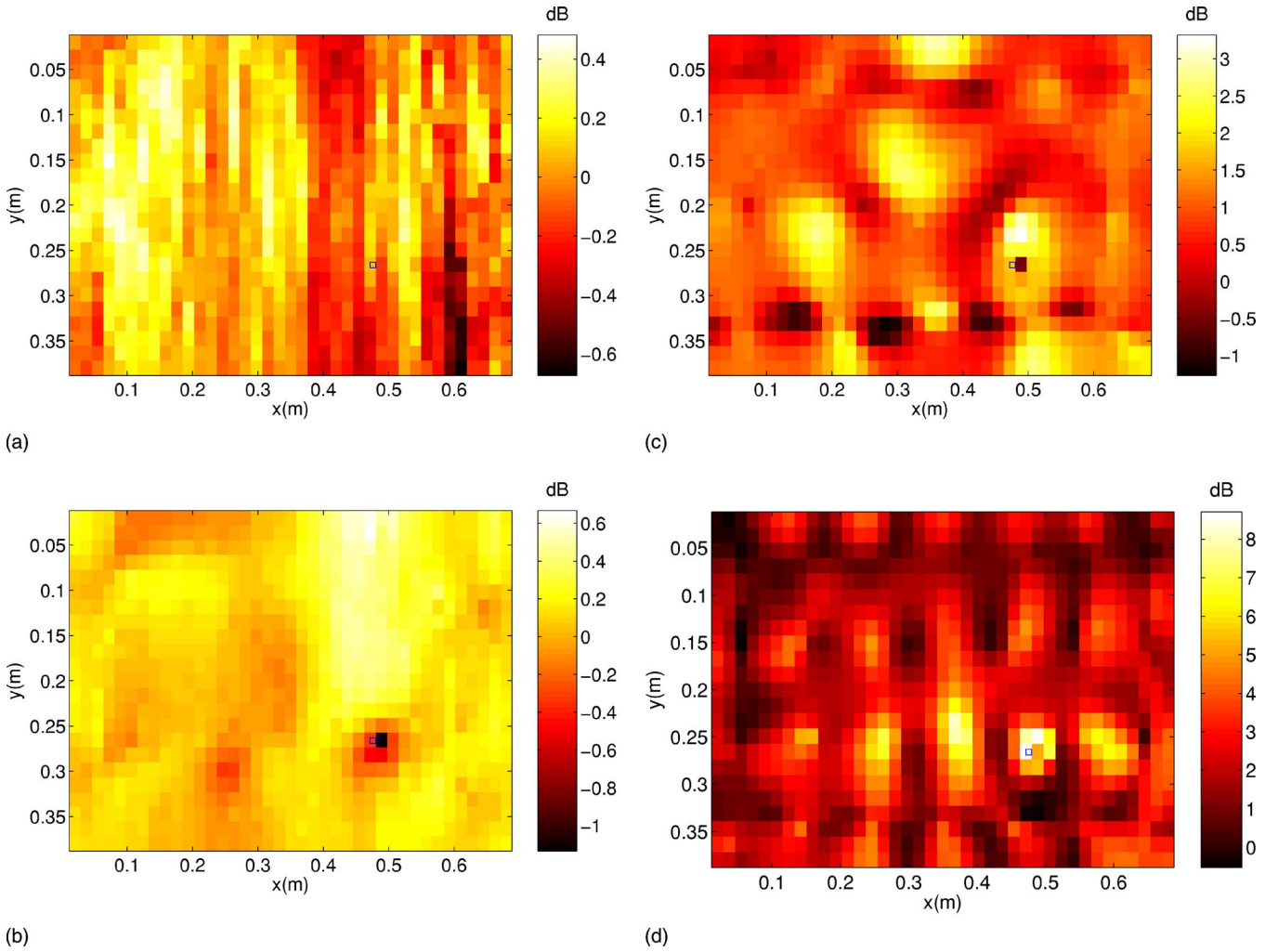


FIG. 13. (Color online) Measured reduction of the total velocity level distributions integrated at different bandwidths (a) up to 25 Hz; (b) 100 to 300 Hz; (c) 300 to 500 Hz; and (d) 500 to 800 Hz, of the resiliently mounted (κ=4) honeycomb panel disturbed by a concentrated force at (0.05, 0.05), and when subjected to the single-channel direct feedback control with feedback gain of 225, using the piezoceramic actuator at $(0.7L_x, 0.7L_y)$. The square mark on the shading denotes the location of the piezoceramic actuator.

sensor pair. To show this, the velocity levels at the piezoceramic actuator location $(0.7L_x, 0.7L_y)$ and at the point $(0.6L_x, 0.6L_y)$, 8 cm away from the the actuator location, are measured in the experimental arrangement shown in Fig. 1. Figure 12 shows the measured velocity spectrum before control and after the single-channel direct feedback control with the gain of 225 using the piezoceramic actuator. It can be seen that the velocity level at a point 8 cm away from the actuator location is hardly reduced, except for a few resonances from 300 to700 Hz, while a reduction in the velocity level can be obtained with a maximum of about 10 dB at the actuator location.

The velocity distributions over the panel with and without control were also measured using a laser scanning vibrometer. Figure 13 shows the distributions of the measured levels of velocity reduction integrated over different bandwidths (a) up to 25 Hz; (b) from 100 to 300 Hz; (c) from 300 to 500 Hz; and (d) from 500 to 800 Hz, when the panel is subjected to the single-channel direct feedback control with feedback gain of 225, using the piezoceramic actuator at $(0.7L_x, 0.7L_y)$. It can be seen that the reduction by the control system is significantly more localized as the frequency increases, while the global reduction can only be achieved by the control system at low frequencies.

V. MULTICHANNEL LOCAL VELOCITY FEEDBACK CONTROLLER

The localization of the reduction achieved with a single-channel control system motivates the investigation of multichannel feedback controllers. In order to achieve the global reduction performance up to higher frequency of interest, a 12-channel direct local velocity feedback control system is investigated theoretically. The assumed locations of the actuator-sensor pairs are shown in Fig. 14. Each sensor-actuator pair is modeled with a single degree of freedom mass-damper-spring system on the honeycomb panel model obtained from CLPT to take into account the local coupling dynamics. The stability and the performance of the control system are evaluated using the state-space representation, based on the block diagram shown in Fig. 15. In this state-space representation, the stability of the closed-loop system can be assessed by calculating the eigenvalues of the system

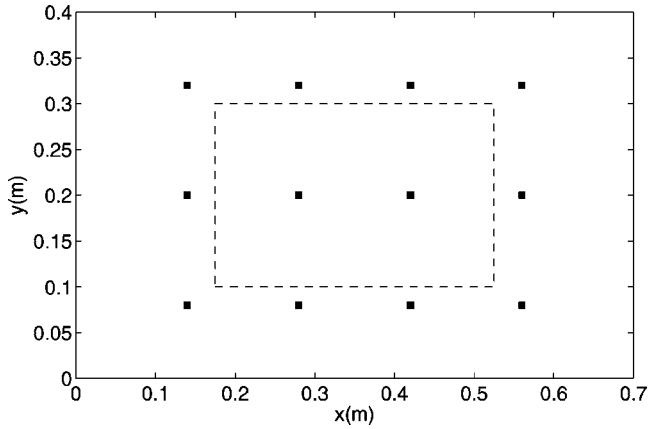


FIG. 14. Actuator-sensor pair locations of 12-channel direct velocity feedback control system. The dashed box is range of high stable feedback gain of individual sensor-actuator pair shown in Fig. 9.

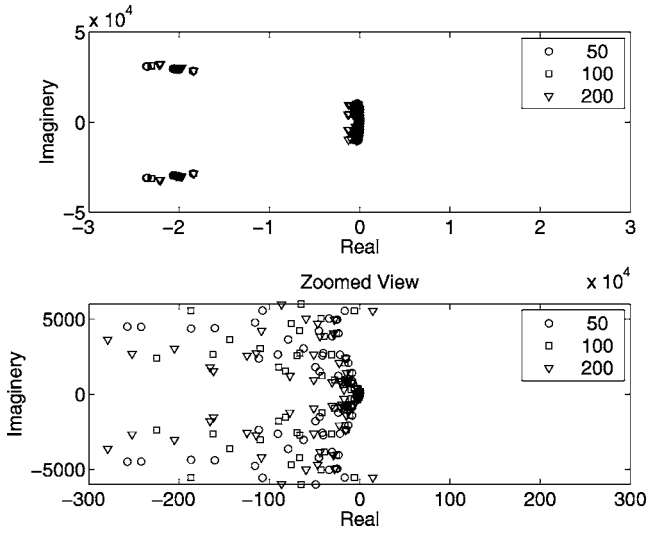


FIG. 16. Eigenvalues of the system matrix (poles) for the 12-channel local feedback control system with various feedback gains of 50(○), 100(▽), and 200(∗).

matrix (Fuller *et al.*, 1996). The imaginary part of the eigenvalue corresponds to the $\pm j\times$ natural frequency and the real part the decay rate of the modal participation factor of the response. The system is stable when all of the eigenvalues have a negative real part.

Figure 16 shows the eigenvalues of the system for various feedback gains ranging from 50 to 200. The upper part of the diagram shows all the poles, and the lower part shows a zoomed view to look at the behavior of the unstable poles. It can be seen that the natural frequencies are invariant with feedback gain at the low frequencies, while the decay rates of the lower order modes are affected. Slight change in natural frequencies is due to the new boundary effect of the sensor-actuator pair at a high feedback gain. It is found that the maximum feedback gain is about 120 in this case. The maximum gain is significantly lower than that of any one single-channel control system, which is about 1000. This is because of the interaction between the multiple actuators and sensors.

Figure 17 shows the predicted control performance of a resiliently mounted honeycomb panel disturbed by a concentrated force at (0.05,0.05), and when subjected to the 12-channel direct feedback control with sensor/piezoceramic actuators with gain of 120 in terms of (a) the total kinetic energy and (b) the total acoustic power. The resonance frequencies are decreased due to the mass loading from the sensor-actuator, and so the modal density at the frequencies between 100 Hz and 1 kHz is much higher that that of the panel with single channel control, as shown in Fig. 10. It can be seen that reductions in the kinetic energy and the acoustic power are much improved using the multichannel control, particularly at high frequencies, while they are almost the same at low frequencies. Figure 18 shows the calculated reduction levels of the velocity distribution integrated from 500 to 800 Hz of controlled panel with feedback gain of 120 using the sensor-actuator pairs positioned at the square marks on the shading. The global reduction in vibration is achieved by connecting the localized reduction of each single sensor-actuator pair. Figure 19 shows (a) the normalized total kinetic energy and (b) the normalized total acoustic power of the controlled panel, comparing with those of the single-channel feedback control. The normalization is carried out up to 1 kHz. The maximum reduction in the normalized kinetic energy is increased by only 0.5 dB while that in the normalized acoustic power is much increased by 3.5 dB. The increase in the acoustic power reduction is due to the globalization of the vibration reduction at high frequencies. It should be noted that, although the reduction in the kinetic energy is not very improved, the maximum reduction can be achieved at lower feedback gain. The control effort for each single channel is much reduced when using multichannel control to obtain the same reduction performance.

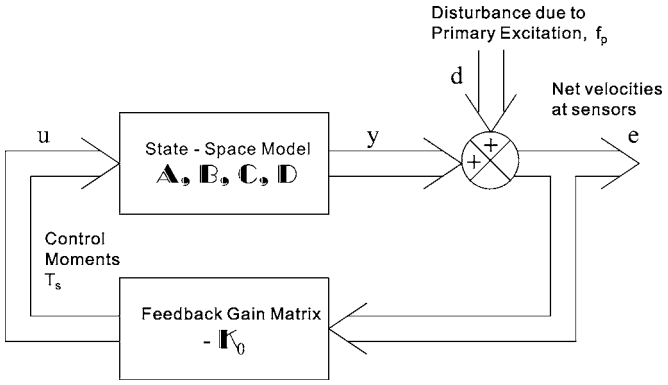


FIG. 15. Block diagram of the multichannel output feedback control system.

VI. CONCLUSION

This study has investigated the feedback control of vibration and sound radiation from honeycomb trim panels using piezoceramic actuators and closely spaced accelerometers. To account for the practical situation of aircraft trim panels, the honeycomb panel is resiliently mounted.

The local coupling of the sensor and actuator in the honeycomb panel was observed experimentally, and modeled using a single degree of freedom system, as developed in Hong (2005). The effect of the local coupling was to roll off the response between the piezoceramic actuator and sensor at high frequencies, so that a feedback control system can have

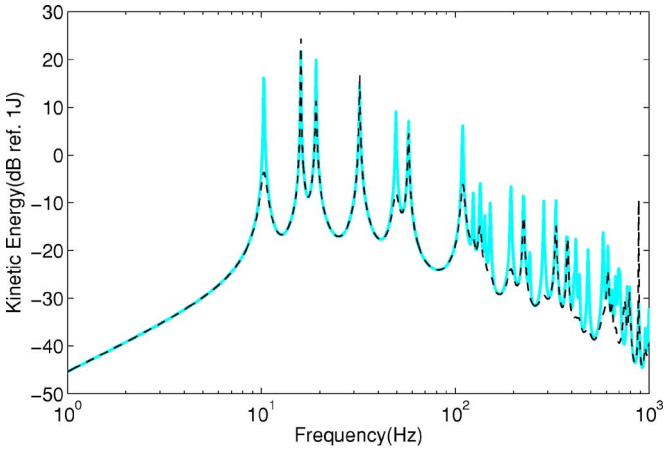


FIG. 17. (Color online) Total kinetic energy (a) and total acoustic power (b) of a resiliently mounted (κ=4) honeycomb panel disturbed by a concentrated force at (0.05, 0.05) with no control (solid line) and with a feedback gain of 120 (dashed line), when subjected to the 12-channel direct feedback control with sensor/piezoceramic actuators.

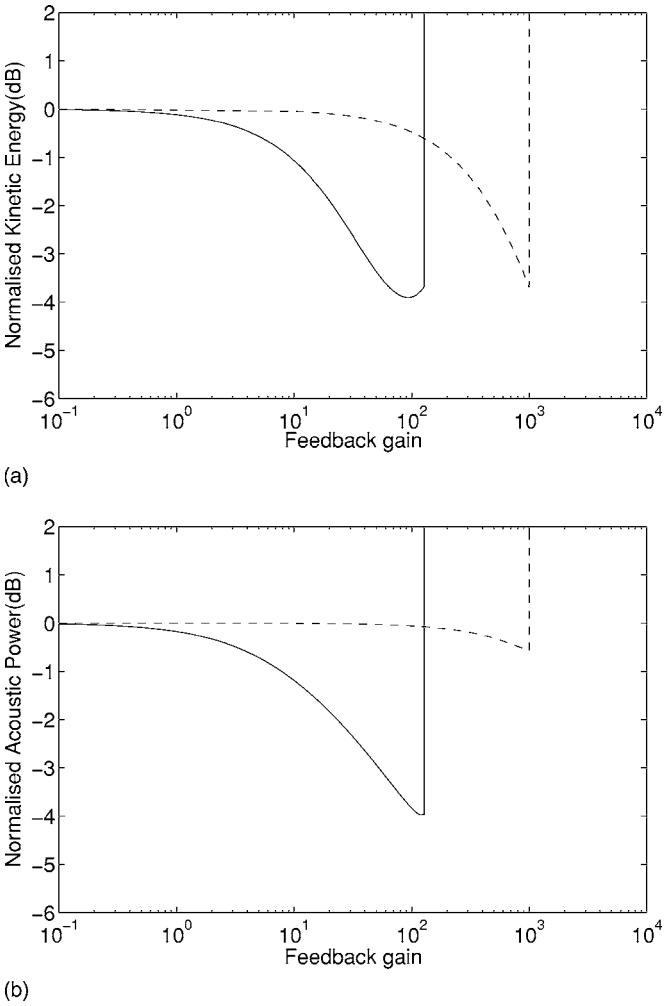


FIG. 19. Normalized control performance of a resiliently mounted honeycomb panel disturbed by a concentrated force at (0.05, 0.05), and when subjected to the 12-channel direct feedback control (solid line) comparing with those of the single-channel feedback control (dashed line). The normalization is carried out up to 1 kHz. (a) Normalized total kinetic energy; (b) normalized total acoustic power.

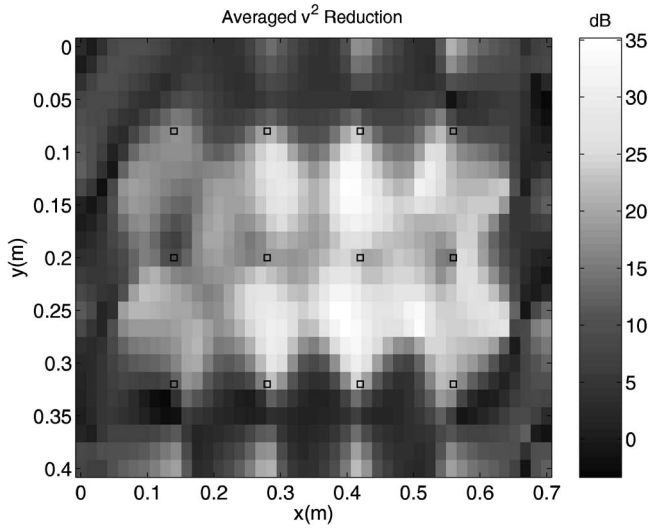


FIG. 18. Calculated total velocity level distribution integrated from 500 to 800 Hz of a resiliently mounted honeycomb panel disturbed by a concentrated force at (0.05, 0.05), and when subjected to the 12-channel direct feedback control with feedback gain of 120. The square mark on the shading denotes the location of the piezoceramic actuators.

high gain margins. Relatively poor global performance was achieved, however, because of localization of reduction around the position of the actuator/sensor at high frequencies. This localization prompts a theoretical investigation of a decentralized active control system using multiple actuator and sensor pairs. Globalized reduction was achieved using a 12-channel direct velocity feedback control. However, due to decreased gain margin, the multichannel system does not yield a significant improvement in the global performance. Although the reduction in the kinetic energy is not very improved, the maximum reduction can be achieved at lower feedback gain. Further work is required before the applicability of this work to real aircraft trim panels can be assessed, particularly with regard to the effects of curvature in real trim panels.

APPENDIX: RESPONSE OF A RESILIENTLY MOUNTED HONEYCOMB PANEL

The panel is considered to be resiliently mounted at the corners. The honeycomb panel with actuator-sensor pair po-

sitioned on the opposite skins can be effectively modeled using the simple panel as shown in Fig. 2. The equation of motion of the panel is given by Eq. (1).

When assuming the external force, F_p, is applied at a point, (x_p, y_p),

$$F_p(x,y,t) = F_p \delta_p(x - x_p, y - y_p), \tag{A1}$$

where $\delta(\ )$ is the delta function. The interaction forces in Eq. (4) are denoted as

$$F_s(x,y,t) = \sum_{i=0}^{N_s} \{k_s^{(i)} w + c_s^{(i)} \dot{w}\} \delta(x - x_s^{(i)}, y - y_s^{(i)}), \tag{A2}$$

$$F_m(x,y,t) = \sum_{i=0}^{N_a} m_l^{(i)} \ddot{w} \delta(x - x_a^{(i)}, y - y_a^{(i)}), \tag{A3}$$

$$F_a(x,y,t) = \sum_{i=0}^{N_a} [k_u^{(i)}\{y_m^{(i)} - w\} + c_u^{(i)}\{\dot{y}_m^{(i)} - \dot{w}\}] \times \delta(x - x_a^{(i)}, y - y_a^{(i)}), \tag{A4}$$

where N_s and N_a are the number of the mounts and the number of the lumped masses and the single degree of freedom systems on the panel, and $y_m(x,y,t)$ is the displacement of the mass of the mth single degree of freedom system, which is governed by

$$m_u^{(i)} \ddot{y}_m^{(i)} + c_u^{(i)} \dot{y}_m^{(i)} + k_u^{(i)} y_m^{(i)} = \{c_u^{(i)} \dot{w} + k_u^{(i)} w\} \times \delta(x - x_a^{(i)}, y - y_a^{(i)}) \quad i = 1,2,3,\ldots,N_a. \tag{A5}$$

The forced response can be expressed as a superposition of normal modes

$$w(x,y,t) = \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} a_{mn}(t) \Theta_{mn}(x,y), \tag{A6}$$

where m and n are the modal indexes, $\Theta_{mn}(x,y)$ is the mode shape function of free-edged panels. The natural frequencies and the mode shape functions can be estimated by the method suggested by Warburton (Fahy and Walker, 2004; Leissa, 1969; Warburton, 1951). Assuming the response of the mass of the sensors to be harmonic, the coupled system governed by the matrix equation

$$[-\omega^2 \mathbf{M}_{gg} + j\omega \mathbf{D}_{gg} + \mathbf{K}_{gg}]\mathbf{w}_g = \mathbf{f}_g, \tag{A7}$$

where

$$\mathbf{M}_{gg} = \left[\begin{array}{c|c} \mathbf{M}_{aa} & \mathbf{M}_{as} \\ \hline \mathbf{M}_{sa} & \mathbf{M}_{ss} \end{array}\right], \tag{A8}$$

$$\mathbf{D}_{gg} = \left[\begin{array}{c|c} \mathbf{D}_{aa} & \mathbf{D}_{as} \\ \hline \mathbf{D}_{sa} & \mathbf{D}_{ss} \end{array}\right], \tag{A9}$$

$$\mathbf{K}_{gg} = \left[\begin{array}{c|c} \mathbf{K}_{aa} & \mathbf{K}_{as} \\ \hline \mathbf{K}_{sa} & \mathbf{K}_{ss} \end{array}\right]. \tag{A10}$$

Defining the modal vector as

$$\mathbf{\Theta}(x,y) = \left\{\begin{array}{c} \Theta_{11}(x,y) \\ \Theta_{21}(x,y) \\ \Theta_{31}(x,y) \\ \vdots \\ \Theta_{12}(x,y) \\ \Theta_{22}(x,y) \\ \Theta_{32}(x,y) \\ \vdots \\ \Theta_{1k}(x,y) \\ \Theta_{2k}(x,y) \\ \Theta_{3k}(x,y) \\ \vdots \\ \Theta_{MN}(x,y) \end{array}\right\} = \left\{\begin{array}{c} \phi_1(x)\psi_1(y) \\ \phi_2(x)\psi_1(y) \\ \phi_3(x)\psi_1(y) \\ \vdots \\ \phi_1(x)\psi_2(y) \\ \phi_2(x)\psi_2(y) \\ \phi_3(x)\psi_2(y) \\ \vdots \\ \phi_1(x)\psi_k(y) \\ \phi_2(x)\psi_k(y) \\ \phi_3(x)\psi_k(y) \\ \vdots \\ \phi_M(x)\psi_N(y) \end{array}\right\}, \tag{A11}$$

the submatrices in Eqs. (A8)–(A10) can be evaluated as

$$\mathbf{M}_{aa} = \rho h_e L_x L_y \mathbf{I} + \sum_{i=1}^{N_a} m_l^{(i)} \{\mathbf{\Theta}(x_a^{(i)}, y_a^{(i)})\}\{\mathbf{\Theta}(x_a^{(i)}, y_a^{(i)})\}^T, \tag{A12}$$

$$\mathbf{D}_{aa} = \rho h_e L_x L_y \operatorname{diag}(2\zeta_{mn}\omega_{mn}) + \sum_{i=1}^{N_a} c_u^{(i)} \{\mathbf{\Theta}(x_a^{(i)}, y_a^{(i)})\} \times \{\mathbf{\Theta}(x_a^{(i)}, y_a^{(i)})\}^T + \sum_{i=1}^{N_s} c_s^{(i)} \{\mathbf{\Theta}(x_s^{(i)}, y_s^{(i)})\}\{\mathbf{\Theta}(x_s^{(i)}, y_s^{(i)})\}^T, \tag{A13}$$

$$\mathbf{K}_{aa} = \rho h_e L_x L_y \operatorname{diag}(\omega_{mn}^2) + \sum_{i=1}^{N_a} k_u^{(i)} \{\mathbf{\Theta}(x_a^{(i)}, y_a^{(i)})\} \times \{\mathbf{\Theta}(x_a^{(i)}, y_a^{(i)})\}^T + \sum_{i=1}^{N_s} k_s^{(i)} \{\mathbf{\Theta}(x_s^{(i)}, y_s^{(i)})\} \times \{\mathbf{\Theta}(x_s^{(i)}, y_s^{(i)})\}^T, \tag{A14}$$

$$\mathbf{M}_{as} = \mathbf{0} \tag{A15}$$

$$\mathbf{D}_{as} = [c_u^{(1)} \mathbf{\Theta}(x_a^{(1)}, y_a^{(1)}) \cdots c_u^{(N_a)} \mathbf{\Theta}(x_a^{(N_a)}, y_a^{(N_a)})], \tag{A16}$$

$$\mathbf{K}_{as} = [k_u^{(1)} \mathbf{\Theta}(x_a^{(1)}, y_a^{(1)}) \cdots k_u^{(N_a)} \mathbf{\Theta}(x_a^{(N_a)}, y_a^{(N_a)})], \tag{A17}$$

$$\mathbf{M}_{ss} = \begin{bmatrix} m_u^{(1)} & 0 & 0 & 0 & 0 & 0 & 0 & \cdots \\ 0 & m_u^{(2)} & 0 & 0 & 0 & 0 & 0 & \cdots \\ & & & \ddots & 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & 0 & m_u^{(k)} & 0 & 0 & \cdots \\ & & & & \vdots & & & \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & m_u^{(N_a)} \end{bmatrix}, \tag{A18}$$

$$\mathbf{D}_{ss} = \begin{bmatrix} c_u^{(1)} & 0 & 0 & 0 & 0 & 0 & 0 & \cdots \\ 0 & c_u^{(2)} & 0 & 0 & 0 & 0 & 0 & \cdots \\ & & & \ddots & 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & 0 & c_u^{(k)} & 0 & 0 & \cdots \\ & & & \vdots & & & & \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & c_u^{(N_a)} \end{bmatrix}, \tag{A19}$$

$$\mathbf{K}_{ss} = \begin{bmatrix} k_u^{(1)} & 0 & 0 & 0 & 0 & 0 & 0 & \cdots \\ 0 & k_u^{(2)} & 0 & 0 & 0 & 0 & 0 & \cdots \\ & & & \ddots & 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & 0 & k_u^{(k)} & 0 & 0 & \cdots \\ & & & \vdots & & & & \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & k_u^{(N_a)} \end{bmatrix}, \tag{A20}$$

where the subscript *aa* denotes the free-edged panel with lumped masses and stiffness, *ss* denotes the single degree of freedom mass-spring-damper-damper system, and *as* and *sa* denote their coupling.

$$\mathbf{w}_g = \begin{Bmatrix} \mathbf{a} \\ \mathbf{y}_m \end{Bmatrix} \tag{A21}$$

and

$$\mathbf{f}_g = \begin{Bmatrix} F_{pq} + T_{pq} \\ \mathbf{0}_{N_a} \end{Bmatrix}, \tag{A22}$$

where $\mathbf{a}$ is the modal amplitudes of the panel, $\mathbf{y}_m = \{Y_m^{(1)}, Y_m^{(2)}, \ldots, Y_m^{(N_a)}\}^T$, which is the displacements of the masses of single degree of freedom (SDOF) systems on the panel, and $\mathbf{0}_{N_a}$ is a zero vector of size $N_a \times 1$. Note that $\mathbf{M}_{sa} = \mathbf{M}_{as}^T$, $\mathbf{D}_{sa} = \mathbf{D}_{as}^T$, and $\mathbf{K}_{sa} = \mathbf{K}_{as}^T$. F_{pq} and T_{pq} are the modal force terms of externally applied forces and moments, respectively, which can be expressed as

$$F_{pq} = \int_A F(x,y)\Theta_{pq}(x,y)dA$$

$$T_{pq} = -\int_A \left(\frac{\partial T(x,y)}{\partial x} + \frac{\partial T(x,y)}{\partial y} \right) \Theta_{pq}(x,y)dA. \tag{A23}$$

The forced response including the actuator and sensor dynamics can be obtained from Eq. (A7). The modal amplitude and the displacement of the mass of the lumped SDOF system can be extracted from $\mathbf{w}_g$ as

$$\mathbf{a} = \mathbf{E}_a \mathbf{w}_g, \tag{A24}$$

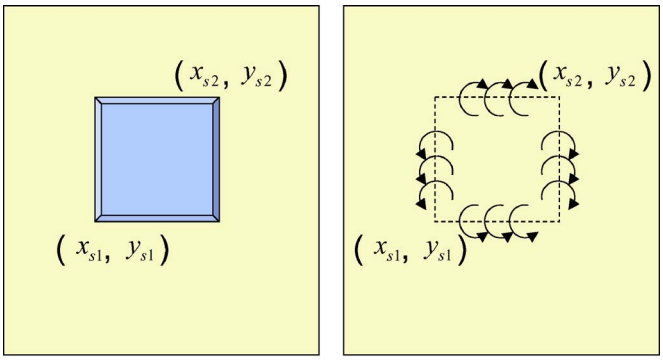
$$\mathbf{y}_m = \mathbf{E}_y \mathbf{w}_g, \tag{A25}$$

where

$$\mathbf{w}_g = [-\omega^2 \mathbf{M}_{gg} + j\omega \mathbf{D}_{gg} + \mathbf{K}_{gg}]^{-1} \mathbf{f}_g, \tag{A26}$$

$$\mathbf{E}_a = \left[\begin{array}{c|c} \mathbf{I}_1 & \mathbf{0}_1 \\ \hline \mathbf{0}_1^T & \mathbf{0}_2 \end{array} \right], \tag{A27}$$

and



(a) Piezoceramic actuator (b) Equivalent Moments

FIG. 20. (Color online) Moments generated by a piezoceramic actuator.

$$\mathbf{E}_y = \left[\begin{array}{c|c} \mathbf{0}_3 & \mathbf{0}_1 \\ \hline \mathbf{0}_1^T & \mathbf{I}_2 \end{array} \right]. \tag{A28}$$

$\mathbf{I}_1$ and $\mathbf{I}_2$ in Eqs. (A27) and (A28) are the unit matrices of size of $MN \times MN$ and $N_a \times N_a$, respectively, and $\mathbf{0}_1$, $\mathbf{0}_2$, and $\mathbf{0}_3$ are zero matrices of size of $MN \times N_a$, $N_a \times N_a$, and $MN \times MN$, respectively. Thus, the displacement at a point (x_r, y_r) of the panel can be expressed as

$$w(x_r, y_r, \omega) = \mathbf{\Theta}^T(x_r, y_r)\mathbf{E}_a \mathbf{w}_g(\omega). \tag{A29}$$

When the panel is subjected to a point force, F_p at (x_p, y_p), the modal force can be written as $F_{mn} = F_p \mathbf{\Theta}(x_p, y_p)$ from Eq. (A23). So, the mobility function can be expressed as

$$\frac{\dot{w}(x_r, y_r, \omega)}{F_p} = j\omega \mathbf{\Theta}^T(x_r, y_r)\mathbf{E}_a \mathbf{w}_g(\omega)\mathbf{\Theta}(x_p, y_p). \tag{A30}$$

On the other hand, when the panel is subjected to a piezoceramic actuator, the moments are generated at the edges of the piezoceramic, $x = x_{s1}, x_{s2}$, and $y = y_{s1}, y_{s2}$, as shown in Fig. 20. The moments can be written in the form

$$T_{xx}(x,y) = T_s \delta(x - x_{s2})\{U(y - y_{s1}) - U(y - y_{s2})\} - T_s \delta(x - x_{s1})\{U(y - y_{s1}) - U(y - y_{s2})\}$$

$$T_{yy}(x,y) = T_s \delta(y - y_{s1})\{U(x - x_{s1}) - U(x - x_{s2})\} - T_s \delta(y - y_{s2})\{U(x - x_{s1}) - U(x - x_{s2})\}, \tag{A31}$$

where $\delta(\,)$ is the delta function and $U(\,)$ is the step function. Substituting Eq. (A31) into Eq. (A23) leads to the modal force of the moments written as

$$T_{mn} = T_s \int_{x_{s1}}^{x_{s2}} \left\{ \frac{\partial \Theta_{mn}(x, y_{s1})}{\partial y} - \frac{\partial \Theta_{mn}(x, y_{s2})}{\partial y} \right\} dx + T_s \int_{y_{s1}}^{y_{s2}} \left\{ \frac{\partial \Theta_{mn}(x_{s1}, y)}{\partial x} - \frac{\partial \Theta_{mn}(x_{s2}, y)}{\partial x} \right\} dy. \tag{A32}$$

Using Eq. (A11), Eq. (A32) can be expressed as

$$T_{mn} = T_s \left\{ \frac{\partial \psi_n(y_{s1})}{\partial y} - \frac{\partial \psi_n(y_{s2})}{\partial y} \right\} \int_{x_{s1}}^{x_{s2}} \phi_m(x) dx + T_s \left\{ \frac{\partial \phi_m(x_{s1})}{\partial x} - \frac{\partial \phi_m(x_{s2})}{\partial x} \right\} \int_{y_{s1}}^{y_{s2}} \psi_n(y) dy \equiv T_s \breve{\mathbf{\Theta}}(x_s, y_s). \quad \text{(A33)}$$

Letting $T_{mn} \equiv T_s \breve{\mathbf{\Theta}}(x_s, y_s)$ as shown in Eq. (A33), the mobility function can be expressed as

$$\frac{\dot{w}(x_r, y_r, \omega)}{T_s} = j\omega \mathbf{\Theta}^T(x_r, y_r) \mathbf{E}_a \mathbf{w}_g(\omega) \breve{\mathbf{\Theta}}(x_s, y_s). \quad \text{(A34)}$$

The time-averaged total kinetic energy (KE) of the panel in terms of the modal velocity amplitude, **a**, is given by

$$\mathrm{KE}(\omega) = \frac{M_p}{2} \mathbf{a}^H \mathbf{a}, \quad \text{(A35)}$$

where M_p is the total mass of the panel and the superscript H indicates the Hermitian transpose.

Assuming that the panel is in an infinite rigid baffle, the sound power radiated from the beam can be calculated by

$$W = \mathbf{a}^H \mathbf{M} \mathbf{a}, \quad \text{(A36)}$$

where **M** is the *modal* radiation resistance matrix given by

$$\mathbf{M} = \mathbf{\Theta}^H \mathbf{R} \mathbf{\Theta}. \quad \text{(A37)}$$

R in Eq. (A37) is the radiation resistance matrix for an array of I elemental radiators expressed in the form

$$\mathbf{R} = \frac{\omega^2 \rho_0 S^2}{4\pi c_0} \begin{bmatrix} 1 & \frac{\sin(kr_{12})}{kr_{12}} & \cdots\cdots & \frac{\sin(kr_{1I})}{kr_{1I}} \\ \frac{\sin(kr_{21})}{kr_{21}} & 1 & \cdots\cdots & \vdots \\ \vdots & & \cdots\cdots & \vdots \\ \frac{\sin(kr_{I1})}{kr_{I1}} & & & 1 \end{bmatrix}, \quad \text{(A38)}$$

where ρ_0 and c_0 are the density and the speed of the sound of the acoustic medium, respectively, k is the acoustic wave number, and r_{ij} is the distance from element i to element j. Since $r_{ij}=r_{ji}$, the matrix is symmetric.

The normalized average kinetic energy and acoustic power, used for estimation of the control performance, can be represented by

$$\overline{\mathrm{KE}}(h) = 10 \log_{10} \frac{\int_{f_1}^{f_2} \mathrm{KE}(f,h) df}{\int_{f_1}^{f_2} \mathrm{KE}_p(f) df}, \quad \mathrm{dB} \quad \text{(A39)}$$

and

$$\bar{W}(h) = 10 \log_{10} \frac{\int_{f_1}^{f_2} W(f,h) df}{\int_{f_1}^{f_2} W_p(f) df}, \quad \mathrm{dB} \quad \text{(A40)}$$

where $\overline{\mathrm{KE}}$ and $\bar{W}$ represent the normalized kinetic energy and acoustic power, respectively. Also, KE and W represent the total kinetic energy and acoustic power under the feedback control, respectively, and KE_p and W_p are those without control; f_1 and f_2 are the lower and the upper frequency of the frequency range of interest.

Barton, C. K., and Mixson, J. S. (**1981**). "Noise transmission and control for a light twin-engine aircraft," J. Aircr. **18**, 570–575.

Bhat, W. V. (**1999**). "Flight test measurement of exterior turbulent boundary layer pressure fluctuations on Boeing model 737 airplane," J. Sound Vib. **14**, 439–457.

Fahy, F., and Walker, J. (**2004**). *Advanced Applications in Acoustics, Noise and Vibration* (Spon, London).

Fuller, C. R., Elliott, S. J., and Nelson, P. A. (**1996**). *Active Control of Vibration* (Academic, London).

Gardonio, P. (**2002**). "Review of active technique for aerospace vibro-acoustic control," J. Aircr. **39**, 206–214.

Gardonio, P., and Elliott, S. J. (**1999**). "Active control of structure-borne and airborne sound transmission through double panel," J. Aircr. **36**, 1023–1032.

Hong, C. (**2005**). "Active control of resiliently mounted flexible structures," Ph.D. thesis, Institute of Sound and Vibration, University of Southampton, UK.

Hong, C., and Elliott, S. J. (**2005**). "Active control of resiliently-mounted beams with a moment pair," Smart Mater. Struct. **14**, 727–738.

Leissa, A. W. (**1969**). *Theory of Plates* (NASA).

Malker, H. (**1999**). "Active damping of impacts in panels," Master's thesis, Institute of Sound and Vibration, University of Southampton, UK.

Maury, C., Gardonio, P., and Elliott, S. J. (**2002**). "Model for active control of flow-induced noise transmitted through double partitions," AIAA J. **40**, 1113–1121.

Mixson, J. S., and Powell, C. A. (**1984**). "Review of recent research on interior noise of propeller aircraft," AIAA/NASA 9th Aeroacoustics Conference, Williamsburg, VA.

Mixson, J. S., and Wilby, J. F. (**1995**). *Interior Noise, Aeroacoustics of Flight Vehicles, Theory and Practice*, edited by H. H. Hubbard (NASA Langley Research Center).

Mixson, J. S., Barton, C. K., and Vaicatis, R. (**1978**). "Investigation of interior noise in a twin-engine light aircraft," J. Aircr. **15**, 227–233.

Nelson, P. A., and Elliott, S. J. (**1992**). *Active Control of Sound* (Academic, New York).

Nilsson, E., and Nilsson, A. C. (**2002**). "Prediction and measurement of some dynamic properties of sandwich structures with honeycomb and foam cores," J. Sound Vib. **251**, 409–430.

Pope, L. D., Wilby, E. G., and Wilby, J. F. (**1987a**). "Propeller aircraft interior noise model. I. Theoretical model," J. Sound Vib. **118**, 449–467.

Pope, L. D., Willis, C. M., and Mayes, W. (**1987b**). "Propeller aircraft interior noise model. II. Scale-model and flight-test comparison," J. Sound Vib. **118**, 469–493.

Reddy, J. N. (**1996**). *Mechanics of Laminated Composite Plates* (CRC Press, Boca Raton, FL).

Saito, T., Parbery, R. D., Okuno, S., and Kawano, S. (**1997**). "Parameter identification for aluminum honeycomb sandwich panels based on orthotropic Timoshenko beam theory," J. Sound Vib. **208**(2), 271–287.

Stothers, I. (2004). Personal communications.

Sulc, J., Hofr, J., and Benda, L. (**1982**). "Exterior noise on the fuselage of light propeller driven aircraft in flight," J. Sound Vib. **84**, 105–120.

Unruh, J. F. (**1988**). "Structure-borne noise control for propeller aircraft," J. Aircr. **17**, 752–757.

Unruh, J. F. (**1989**). "Installation effects on propeller wake/vortex-induced structure-borne noise transmissions," J. Aircr. **27**, 444–448.

Vaicatis, R. (**1980**). "Noise transmissions into a light aircraft," J. Aircr. **17**, 81–86.

Warburton, G. B. (**1951**). "The vibration of rectangular plates," Proc. Inst. Mech. Eng. **168**, 371–384.

Acoustic energy relations in Mudejar-Gothic churches

Teófilo Zamarreño,[a] Sara Girón,[b] and Miguel Galindo[c]
Departamento de Física Aplicada II, Universidad de Sevilla, ETS de Arquitectura IUCC, Avda. Reina Mercedes 2, 41012-Sevilla, Spain

(Received 3 February 2006; revised 11 October 2006; accepted 14 October 2006)

Extensive objective energy-based parameters have been measured in 12 Mudejar-Gothic churches in the south of Spain. Measurements took place in unoccupied churches according to the ISO-3382 standard. Monoaural objective measures in the 125–4000 Hz frequency range and in their spatial distributions were obtained. Acoustic parameters: clarity C_{80}, definition D_{50}, sound strength G and center time T_S have been deduced using impulse response analysis through a maximum length sequence measurement system in each church. These parameters spectrally averaged according to the most extended criteria in auditoria in order to consider acoustic quality were studied as a function of source-receiver distance. The experimental results were compared with predictions given by classical and other existing theoretical models proposed for concert halls and churches. An analytical semi-empirical model based on the measured values of the C_{80} parameter is proposed in this work for these spaces. The good agreement between predicted values and experimental data for definition, sound strength, and center time in the churches analyzed shows that the model can be used for design predictions and other purposes with reasonable accuracy. [DOI: 10.1121/1.2390665]

PACS number(s): 43.55.Br, 43.55.Gx [NX] Pages: 234–250

I. INTRODUCTION

The study of the sound field in places of worship, in any type of religion and both for new buildings and for rehabilitation of ancient existing buildings (patrimony preservation), has recently aroused great interest within the general field of architectural acoustics as is evident from the research literature,[1–3] from some of the latest international congresses such as in Rome[4] and Seville (Spain),[5] which featured a specific session for worship building acoustics, and from other related projects.[6] On one hand, this interest is of a practical nature as a result of the growing demand for acoustic comfort in public places which are used for oral or musical liturgical purposes or for other cultural performances. On the other hand, the interest is more fundamental since the investigation into the acoustics of these complex spaces gives information about the general acoustic aspects of architectural heritage and aids the general comprehension of room acoustics.

This paper deals with the experimental results of certain monoaural energy-based acoustic parameters: clarity, definition, sound strength, and center time, all studied as a function of source-receiver distance, and proposes a model to interpret such experimental data. These parameters are considered very relevant for the evaluation of acoustic quality based on energy criteria. The parameters have been measured for octave bands between 125 and 4000 Hz and within their spatial distribution in each church. In the present analysis each parameter has been averaged spectrally in accordance with the most widely accepted way.

From the starting point of the deviation of the ideal pattern of a diffuse sound field in these places as shown by all the parameters analyzed and the comparison of the experimental data with the revised theory of Barron *et al.*[7] for concert halls, the semiempirical model deduced from the experimental sound pressure values, proposed by Sendra *et al.*[8] for these types of religious spaces and the modified theory of Cirillo *et al.*,[9] originally conceived for Romanesque churches and recently refined to include a wider type of churches,[10] have led the authors[11] to propose a relatively simple model based on measured C_{80} values. Hence, with volume, reverberation time, source-receiver distances and μ parameter as premises, this model enables the prediction of all the remaining measured acoustic parameters studied in churches of this character in the south of Spain. These churches constitute an important sample of the cultural inheritance of the city which had a strong influence on Latin American Baroque architecture which includes the Mudejar timber roofs in important churches in Columbia, Bolivia, El Salvador and Cuba.

II. MEASUREMENT TECHNIQUE

The procedures employed have been those established in the ISO-3382 standard and all measures have been carried out in unoccupied churches. Temperature and relative humidity have been measured with a precision electronic thermohygrometer and a barometer has determined the atmospheric pressure. The range of variation was 22.6–27.4 °C for the temperature, 35.7–65.7% for the relative humidity, and 101.7–102.5 kPa for the atmospheric pressure.

Monoaural impulse responses and other room responses to stationary signals have been measured to determine reverberation times (T) together with the energy-based parameters for each frequency band in all receiver positions: clarity (C_{80}), definition (D_{50}), sound strength (G) and center time

[a]Electronic mail: teofilo@us.es
[b]Electronic mail: sgiron@us.es
[c]Electronic mail: mgalindo@us.es

(T_S). From this point on, these parameters are spectrally averaged in each position as follows: clarity and center time as a direct average of 500, 1000, and 2000 Hz octave band values, definition as a weighted average in the same way as Marshall[12] proposed for C_{50}:

$$D_{50\text{av}} = 0.15D_{50}(500\ \text{Hz}) + 0.25D_{50}(1\ \text{kHz}) + 0.35D_{50}(2\ \text{kHz}) + 0.25D_{50}(4\ \text{kHz}) \quad (1)$$

and a mean of the corresponding mid-frequency 500 and 1000 Hz bands for sound strength[13] and reverberation time.

The impulse response has been obtained using maximum length sequence (MLS) signals. The analyzer used has been the Maximum Length Sequence System Analyzer, from DRA Laboratories (MLSSA) based on a full-length Industry Standard Architecture (ISA) card of data acquisition housed in a slot of a PC with the corresponding software which runs under MS-DOS operating system. For the measurement of the impulse response in order to obtain the decay curves and the energy parameters, the MLS spectrum must be conditioned to that of pink noise before being fed to the amplifier. This is done here by means of the CWF-1 filter from One-on-One Technical Products. In order to improve the signal-to-noise ratio at low frequencies, the signal captured by the microphone is inverse filtered before entering MLSSA for its analysis. In addition, to improve this signal-to-noise ratio, in each position, the impulse response is obtained by averaging eight MLS periods.

An omnidirectional source B&K 4296 is placed at the most usual point of location of the natural source: the altar at a height of 1.70 m from the floor. The microphone is located at the approximate height of the head of a seated person which is 1.20 m from the floor, in a predetermined number of positions distributed in the central nave and the lateral naves ranging from 12 reception points of Santa Catalina church to 23 of Santa Marina (see Fig. 1).

The microphone used is a 1/2 in. B&K 4190, and a B&K 2669 preamplifier and a bias-source B&K 2804, to supply the 200 V dc polarization voltage necessary.

In order to study sound pressure levels, two different techniques for the calculation of sound strength have been used. The first technique uses the emission of a stationary signal, later corrected with the level produced by the source, at 10 m distance under free field conditions. The other technique is derived from the impulse response generated from the MLS signals, by means of calculating the reference level from the direct sound of this impulse response and from its source-receiver distance. This distance is calculated from the initial flying time of the impulse responses. The results coming from these two techniques have been compared through lineal regression and a slope near to one has been obtained: $y=0.962x$; $R^2=0.824$. The coefficient of the determination is acceptable if one takes into account that the microphone positions can vary for the impulse response and stationary signal measurements. In this work the data of the stationary signal is presented as a more reliable and clearer method of determining the spectral variation when a calibrated source is available.

The stationary levels are produced by a B&K 4296 omnidirectional source which emits pink noise and these levels are measured using MLSSA configured in Scope mode. The power level of emission is adjusted to 111 dB. The microphone used is the same as described before, thereby allowing calibration in situ.

III. CHURCHES ANALYZED

Twelve churches of the same typology varying in volume, dimensions, inner endings, and furnishing were acoustically reported. All these churches were built in the Middle Ages and their architectural style was a unique Spanish artistic movement since it was influenced by both Islamic and Christian Gothic elements.

The Mudejar-Gothic churches in Seville are morphologically characterized by this stylistic dualism: a vaulted Gothic apse and a body of three naves with a wooden timber roof (collar beam in the main nave) of Moorish origin. Its brick walls are complemented with portals and a stone apse. The supports are also clearly Islamic, with quadrangular or sometimes octagonal pillars and with raised brick moldings as decoration. Pointed, round, or segmental arches rest on these supports.

Among other elements of particular interest, funeral chapels have been added successively to the side naves and, on some occasions, are housed in remaining sections of preexisting mosques. Funeral chapels do not exist in San Julián, San Esteban, and San Marcos churches.

According to the historian Angulo,[14] the first Mudejar church built in Seville was Santa Marina; this church established the so-called *Seville parish type*. The 12 Catholic temples analyzed in this paper are all located in Seville's historical center.

Figure 1 shows the ground plan for each church under study with the source and receiver positions for measurements and the seating areas. These drawings are on the same scale and presented in order of decreasing volume. Table I summarizes some geometrical data of interest: volume V, length L, width W, average height H, total surface S_T, central nave ground surface S_C, lateral nave ground surface S_L, and total ground surface S_G, for the 12 churches. The congregational seating area consists of wooden pews with sometimes an additional group of seats and a textile carpet may be present at the entrance of the central nave. A complete description of the furnishings and other acoustic information on these temples has been published previously.[15]

IV. THEORETICAL MODELS

A. Introduction

The acoustic parameter prediction in churches is generally not an easy task. Simulation techniques may provide acceptable values of acoustic parameters but an elaborate three-dimensional model is an essential prerequisite. This complexity may discourage professionals from considering acoustic requirements for restoration or new designs.

Reverberation time is calculated through different well-known formulas. However, in religious buildings the calculation is more complicated since the different partial volumes

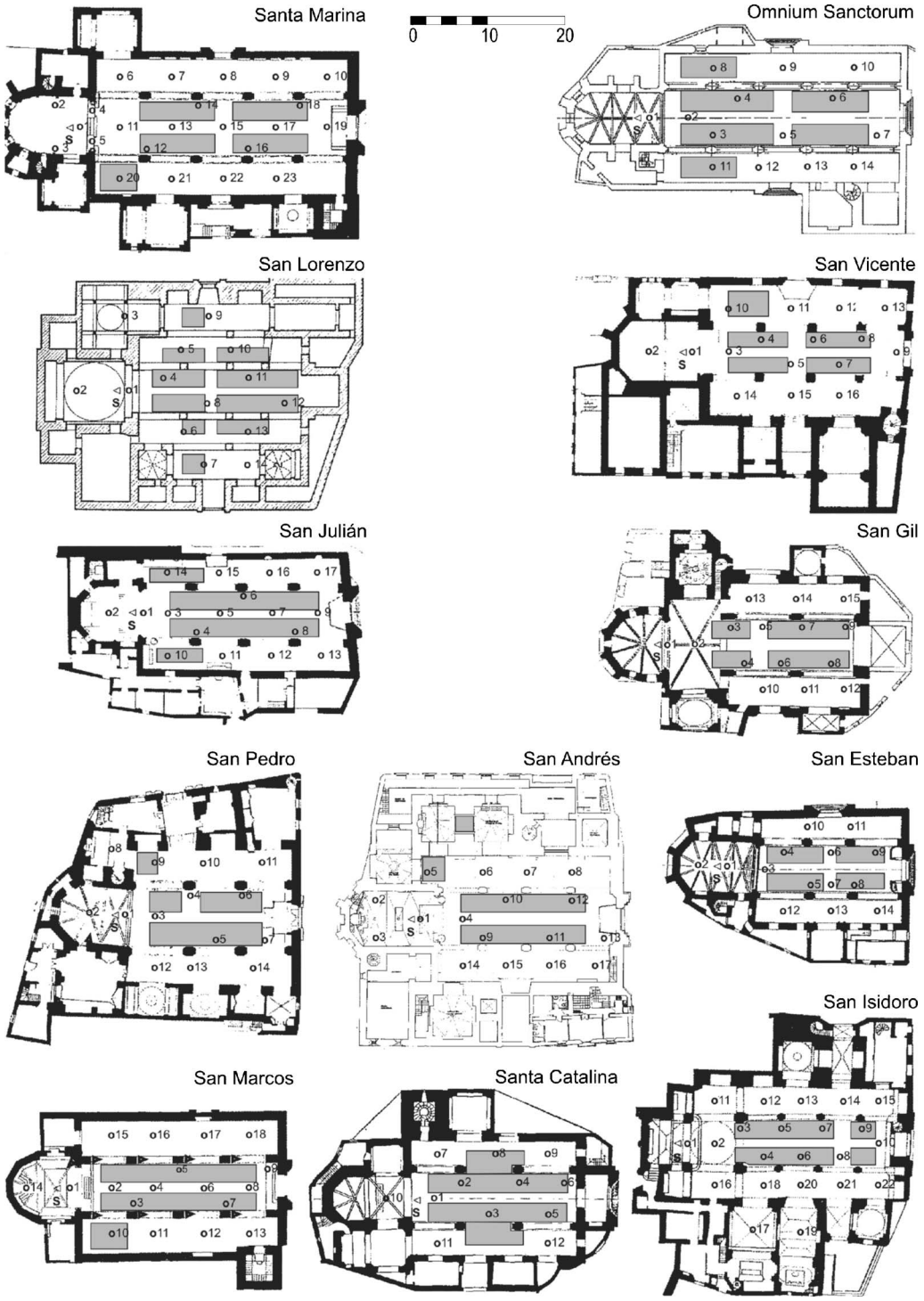


FIG. 1. Ground plan with the source (S) and receiver positions (o) for measurements and pew zone (shaded area) for the 12 churches (same scale for all churches).

which configure them, (chapels, vaults, lateral naves, choir seating, and other architectural elements), may be only partially coupled to the main volume and consequently the classic formulas do not properly account for the energy balance and reverberation times. Knudsen and Harris[16] proposed a simple empirical method to adapt Sabine's equation to these enclosed rooms. In that work, an apparent absorption for the coupled surface in relation to the main volume is assigned. The effect of coupled spaces in churches has been well covered by Carvalho[17] and Magrini and Ricciardi.[18]

If the geometry is not too complex, then the effect of coupling may be negligible in a first approximation or be dealt with by Carvalho's procedure, or by simplifications such as Kundsen's method, which allows reverberation times to be calculated. Predictions of other acoustic parameters are notably more difficult since they are more sensitive to position than reverberation time.

It has been verified[19,20] that strong correlations between certain acoustic parameters exist, and that, even though they are statistically significant, there is no point in calculating

TABLE I. Geometrical data in Mudejar-Gothic churches.

Church	Reception	V (m^3)	L (m)	W (m)	H (m)	S_T (m^2)	S_C+S$_L$ (m^2)	V/S_G (m^3/m^2)
Sta. Marina	23	10,708	34	19	15	4517	250+279	20.24
O. Sanctorum	14	8180	29	16	16	3760	194+219	19.81
S. Lorenzo	14	7040	24	24	16	3346	132+267	17.64
S. Vicente	16	6915	26	18	11	3656	144+180	21.34
S. Julián	17	6226	27	15	13	3321	187+150	18.47
S. Gil	15	6200	22	15	14	3249	170+148	19.50
S. Pedro	14	6180	20	17	16	3035	123+154	22.05
S. Andrés	17	5955	23	15	11	3380	132+154	20.82
S. Esteban	14	4746	20	14.5	14	2691	141+134	17.26
S. Marcos	18	4623	26	17	10	3041	144+226	12.49
Sta. Catalina	12	4362	22	15	12	2474	121+102	19.56
S. Isidoro	22	3947	26	14.5	11	2547	133+144	14.25

some parameters as a function of others, since the last ones must be measured. Nevertheless, correlations between acoustic parameters and geometric characteristics, such as source-receiver distance, are useful and welcome. A theoretical model which implements these characteristics by accounting for the variation of the reflected acoustic field with source-receiver distance was proposed by Barron and Lee.[7] This model has been widely used to predict G and C_{80} averaged values in concert halls and auditoria. However, it is not applicable in places of worship as has been highlighted by several authors.[8,9,21] The reason is that in the presence of surfaces which are more diffusing in these spaces, the model tends to overestimate the early energy; the presence of certain architectural elements peculiar to churches, such as decorations and lateral chapels, affect the lateral energy in the first reflections as well, since sound is reflected according to complex paths which prevent them from returning to the main volume. Otherwise the desproportinate dimensions (length much greater than width and height) affect the sequence of arrival of reflected energy that follows immediately after direct sound.

B. Theoretical model revision

According to the ideal statistical theory of sound propagation in rooms, if the absorption is uniformly distributed and the sound field is diffuse, then the instantaneous normalized reflected energy density $\epsilon'(t')$, when taking the integrated energy density of the direct sound at 10 m E_{D10} as reference, is given by[22]

$$\epsilon'(t') = \frac{\epsilon(t)}{E_{D10}} = \frac{13.82 \cdot 31{,}200}{V} e^{-\frac{13.82t'}{T}}(s^{-1}) \quad 0 \leqslant t' < \infty, \tag{2}$$

where $t'=0$ coincides with the emission of sound from the source, V is the volume, T the reverberation time and $\epsilon'(t')$ becomes dimensionless after integration. Implicit in this derivation is that the onset time of the reflected sound occurs simultaneously at all points throught the space at the moment the direct sound is emitted.

All the acoustical parameters based on the energy may be calculated using Eq. (2), and will be expressed as a function of V, T, and r, the source-receiver distance. Direct sound (d) and total reflected energy (ℓ_T) are given by

$$d = \frac{100}{r^2}, \tag{3}$$

$$\ell_T = \int_0^\infty \epsilon'(\theta) d\theta = \frac{31{,}200T}{V}. \tag{4}$$

Hence the sound strength can be expressed as

$$G(r) = L_P - L_{P_{10}} = 10 \log(d + \ell_T) = 10 \log\left(\frac{100}{r^2} + \frac{31{,}200T}{V}\right)(\text{dB}), \tag{5}$$

where $G(r)$ in Eq. (5) is distance dependent due to the direct sound energy but not to the total reflected energy and hence, beyond the reverberant radius, the sound strength is uniform and takes the value $10 \log(31{,}200T/V)$ dB. The same happens with clarity, definition, and center time which tend towards $10 \log(e^{1.11/T}-1)$ dB, $(1-e^{-0.69/T})$ 100%, and $T/13.82$ s, respectively. However, measurement experimental data show this to be an oversimplification.

Barron and Lee[7] developed a revised theory to take into account that reflected levels significantly decrease with increasing distance from the source in concert halls. The authors remarked that in those rooms the situation differed from diffuse requirements, mainly due to the concentration of absorption in the audience area. This model can be derived by reasoning that as reflected sound cannot arrive earlier than direct sound, the arrival of the direct sound can also be taken as the onset time of the reflected sound at that location. This can be explicitly expressed by writing $t'=t_D+t$ in Eq. (2), where $t_D=r/c$ is the flight time of direct sound from the source to the position at a distance r. Using this revised integration time gives a revised value for the instantaneous normalized reflected energy density which is now distance dependent:

$$\epsilon'_B(t,r) = \frac{\epsilon(t,r)}{E_{D10}} = \frac{13.82 \cdot 31{,}200}{V} e^{-\frac{0.04r}{T}} e^{-\frac{13.82t}{T}} (s^{-1})$$
$$0 \leqslant t < \infty, \tag{6}$$

where $t=0$ coincides with the arrival of direct sound at r distance. In other words, it supposes that, at location r, a pure exponential decay starts from the arrival of direct sound at that location and that reflected energy[23] is reduced at this point by a factor $e^{-\frac{0.04r}{T}}$ with respect to the classic expression of reflected sound energy.

In order to predict total sound pressure level and other indices, the sound energy is divided into three components: the direct sound (d) expressed in Eq. (3), the early reflected energy (e) (from 0 to τ ms), and the late reflected energy (ℓ) (from τ ms to infinity), with $\tau=80$ ms for clarity index and $\tau=50$ ms for definition index. The corresponding reflected energies in this model become:

$$e_B = \frac{31{,}200T}{V} e^{-\frac{0.04r}{T}} (1 - e^{-\frac{13.82\tau}{T}}), \tag{7}$$

$$\ell_B = \frac{31{,}200T}{V} e^{-\frac{0.04r}{T}} e^{-\frac{13.82\tau}{T}} \tag{8}$$

and consequently sound strength is given by

$$G_B(r) = 10 \log(d + e_B + \ell_B)$$
$$= 10 \log\left(\frac{100}{r^2} + \frac{31{,}200T}{V} e^{-\frac{0.04r}{T}} \right) (\text{dB}). \tag{9}$$

In the same manner, clarity and definition indices can be expressed as a function of source-receiver distance by

$$C_{80B}(r) = 10 \log\left(\frac{d + e_B}{\ell_B} \right)_{\tau=0.08}$$
$$= 10 \log\left(\frac{V e^{\frac{0.04r+1.11}{T}}}{312Tr^2} + e^{\frac{1.11}{T}} - 1 \right) (\text{dB}), \tag{10}$$

$$D_{50B}(r) = \left(\frac{d + e_B}{d + e_B + \ell_B} \right)_{\tau=0.05}$$
$$\times 100 = \left(1 - \frac{\frac{31{,}200T}{V} e^{-\frac{0.04r+0.69}{T}}}{\frac{100}{r^2} + \frac{31{,}200T}{V} e^{-\frac{0.04r}{T}}} \right) \times 100 \ (\%). \tag{11}$$

Likewise center time becomes

$$T_{\text{SB}}(r) = \frac{\int_0^\infty t \epsilon'_B(t,r) dt}{d + e_B + \ell_B} = \frac{\frac{31{,}200T^2}{13.82V} e^{-\frac{0.04r}{T}}}{\frac{100}{r^2} + \frac{31{,}200T}{V} e^{-\frac{0.04r}{T}}} \ (s). \tag{12}$$

The aforementioned authors[7] have verified the predictions of revised theory in many concert halls and auditoria mainly in order to establish G and C_{80} mean values.

In the same line, Vorländer[24] proposed for reverberant test rooms that the onset of reflected sound should start at the arrival of the first reflection and so, by substituting r in Barron's expression with the mean free path, the normalized reflected energy density is given by

$$\epsilon'_V(t) = \frac{\epsilon(t)}{E_{D10}} = \frac{13.82 \cdot 31{,}200}{V} e^{-\frac{A}{S}} e^{-\frac{13.82t}{T}} (s^{-1}), \tag{13}$$

where A is the absorption and S the surface of the enclosure.

Sendra, Zamarreño, and Navarro showed that reflected sound levels in Mudejar-Gothic churches generally fell below those predicted by Barron's revised theory. In their study, an adjustment parameter β, for each octave band, is experimentally determined through nonlinear regression in order to maximize correlation between predicted and measured sound level L_p data. Average values of β are proposed for general use in this specific type of church. According to their hypothesis, the normalized reflected energy density can be expressed for those spaces as

$$\epsilon'_\beta(t,r) = \frac{\epsilon(t,r)}{E_{D10}} = \frac{13.82 \cdot 31{,}200}{V} e^{-\frac{\beta r}{T}} e^{-\frac{13.82t}{T}} (s^{-1})$$
$$0 \leqslant t < \infty. \tag{14}$$

Again $t=0$ coincides with the arrival of direct sound at r distance. Equation (14) supposes that direct sound can be considered as the onset of reflected energy at location r, the same as in Barron's model, but the reflected energy is reduced by a factor $e^{-\frac{\beta r}{T}}$, where the β coefficient is determined experimentally as explained above. According to the paper by Sendra, Zamarreño, and Navarro,[8] β coefficients oscillate between 0.06 and 0.12 (sm^{-1}) values which supposes larger factors of attenuations than in Barron's model. This loss of early energy is associated with the presence of certain architectural elements characteristic to churches, and therefore sound strength becomes

$$G_\beta(r) = 10 \log(d + e_\beta + \ell_\beta)$$
$$= 10 \log\left(\frac{100}{r^2} + \frac{31{,}200T}{V} e^{-\frac{\beta}{T} r} \right) (\text{dB}). \tag{15}$$

Equation (15) formally takes an expression similar to the corresponding Barron's Eq. (9). In fact, the expressions for the early and late reflected energies, clarity, definition and center time would be similar to the set of Eqs. (8)–(12), respectively, if 0.04 were replaced by β. Nevertheless, this model based on experimental G values failed to predict clarity and definition versus source-receiver distances in these churches as Galindo, Zamarreño, and Girón[25] pointed out. In their work they remarked that clarity and definition experimental values showed a larger decrease with respect to source-receiver distance than that obtained through the β model expected values, and that in fact the predictions from Barron's model and β adjustment coincided for those two indices. By means of the simple calculation below, those features can be shown explicitly. The differences between predicted Barron's values G_B and the β adjustment G_β for sound strength at great distances where direct sound can be negligible are given by

$$G_B - G_\beta \approx 10\log\left(\frac{e_B + \ell_B}{e_\beta + \ell_\beta}\right) = 10\log\left(e^{\frac{(\beta-0.04)r}{T}}\right)$$
$$= 4.3\left(\frac{\beta - 0.04}{T}\right)r. \quad (16)$$

Quantifying Eq. (16) for Santa Marina church for instance, $T=2.85$ s at the 2000 Hz band, $\beta=0.08$, at the rear part $r=35$ m, $G_B-G_\beta=2.1$ dB, which is a significant quantity. On the other hand, by taking into account clarity differences from the two models at the same source-receiver distance, a zero value is obtained:

$$C_{80B} - C_{80\beta} \approx 10\log\left(\frac{e_B \times \ell_\beta}{e_\beta \times \ell_B}\right) = 10\log(1) = 0. \quad (17)$$

The same conclusion can be reached for the definition index. As both the revised theory and the β model introduce attenuations of the reflected energy which affects early and late energy equally, they are unable to predict energy indices based on early-to-late or early-to-total ratios.

A more general model to be applied to rooms with asymmetrical absorption was proposed by Arau.[26] The model is based in revised theory but the author divided the extinction decay curve into three different parts with different slopes, the initial early sound, the early sound, and the late sound, respectively. The rhythm of the two first is governed by the early decay time (EDT) and by the mean reverberation time. The principal disadvantages of this model are its complexity and the fact that it is largely based on EDT values which are very sensitive to location and hence difficult to predict.

Recently Cirillo and Martellota[9] derived a modified theory and an alternative adjustment to the energy density for a type of church. Their solution assumes that the linear level decay starts after a certain delay time t_R after the arrival of the direct sound, which is considered $t=0$ in this work, and which is proportional to the source-receiver distance, $t_R=\rho r$. The constant of proportionality ρ is determined experimentally by establishing a similitude of churches according to room characteristics. They derived a correction on this basis and accounted for the specific early reflection pattern as proposed by Vorländer[24] such that the instantaneous normalized energy density can be written as

$$\epsilon'_C(t,r) = \frac{\epsilon(t,r)}{E_{D10}} = \begin{cases} \left(\frac{\epsilon'_B(t_R) - \gamma d}{t_R}\right)t + \gamma d \ (s^{-1}) & 0 \leqslant t \leqslant t_R, \\ \frac{13.82 \cdot 31{,}200}{V} e^{-\frac{(0.04+13.82\rho)r}{T}} e^{-\frac{13.82t}{T}}(s^{-1}) & t_R \leqslant t < \infty, \end{cases} \quad (18)$$

where the factor γ weights the magnitude of the early reflections against the direct sound d and depends on the mean absorption coefficient, the mean scattering coefficient, and the mean free path of the church. Although the model was initially tested for Apulian Romanesque churches,[9] ultimately they have refined it to provide a more reliable and physically acceptable hypothesis and have extended their prediction to other types of churches.[10] They have established some sort of typological behavior and the comparison of their predictions with the experimental values for sound strength, clarity, and center time is reasonably accurate. However, the model does not retain the simplicity in the calculations of the other statistical formulas.

C. Proposed model

Gade[27] carried out a statistical analysis of room acoustical and architectural data from 32 concert halls in order to look for relationships between spatially averaged acoustic parameters and various design variables. The author reached the conclusion that volume and reverberation time are the main factors governing the behavior of the acoustic parameters which measure level, reverberance, and clarity. However, geometrical factors have an additional influence on clarity and, particularly on spaciousness and on the conditions for musicians on the orchestra platform. This author suggests that the results point towards the possibility of obtaining empirical formulas by means of regression that would predict the values of the acoustic parameters in halls, if their reverberation time and geometry were known.

In this context, following Gade's ideas and taking advantage of the relevant contribution by Barron, it would be interesting to model the observed additional attenuation in the acoustic energy parameters as a function of source-receiver distance. The proposed model tries to retain the simplicity of Barron's revised theory and tune it according to experimental data. In this way it will be possible to implement typological models in terms of architectural characteristics that could be applied in an immediate way to make design decisions.

With the purpose of comparing the extinction curves at different distances in these spaces, these curves have been superimposed at the 1000 Hz octave band for two churches studied, that of the largest volume and one of the most reverberant Santa Marina church, and another smaller and less reverberant Santa Catalina church. In Fig. 2(a) the decay trace for different reception points belonging to Santa Marina central nave and Fig. 2(b) for different reception points belonging to its lateral nave are shown. Likewise, Figs. 2(c) and 2(d) show the equivalent results for Santa Catalina church. In each case the curve corresponding to the nearest point in front of the source located on the central nave has been chosen as a reference, and hence vertical scale is relative. However, the temporary axis is absolute and has its origin at the instant the source starts emitting.

Although all curves decay linearly and with the same attenuation rhythm, remarkable differences appear in the initial parts for both churches. On trying to model this behavior with the mentioned β adjustment model proposed by Sendra Eqs. (14) and (15), the early and late energies are reduced by the same attenuation factor which is adjusted through nonlinear regression from the measured global levels. However, trying to predict the behavior of the energy parameters with this method, produced values which were very near those derived from Barron's pattern and far from experimental data

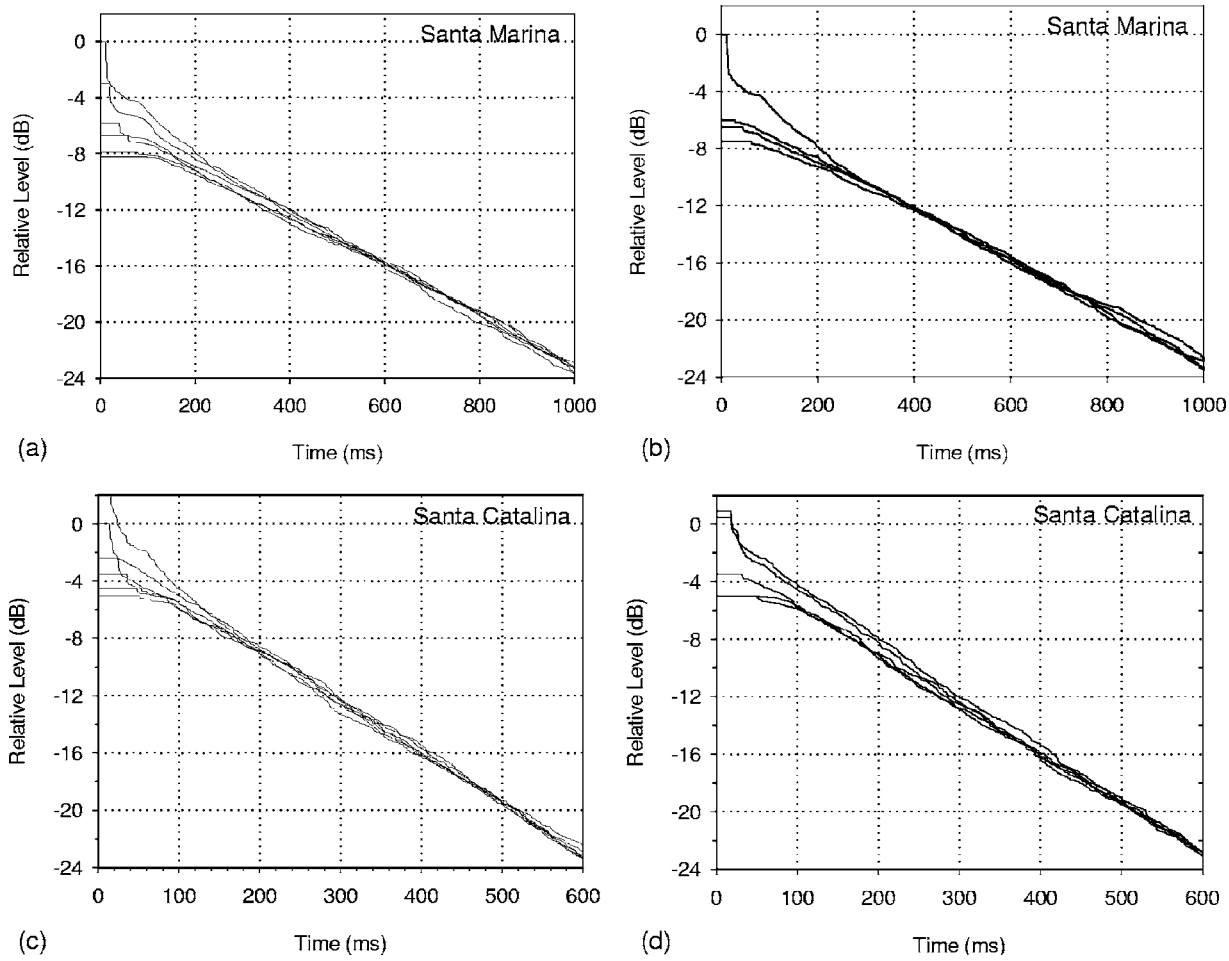


FIG. 2. Extinction decay traces at several location points at the 1000 Hz band in Santa Marina church, (a) central nave, (b) lateral nave; and in Santa Catalina church (c) central nave, (d) lateral nave. In this graph, time $t=0$ coincides with the time of emission of the source, and the vertical scale is relative: 0 dB corresponds to the stationary level of the nearest point to the source in the central nave.

as has already been mentioned. Other approaches need to be considered that, without losing the simplicity of the pattern, allow the prediction of the values of all energy parameters appropriately.

To reach this objective, the first part of the energy-time curves measured in churches has been analyzed. In Figs. 3(a)–3(d), these curves appear octave filtered at 1 kHz, for positions 11, 13, 15, and 17 respectively, located in the central nave of Santa Marina church (see Fig. 1). In each case, the first cursor (dotted vertical line) shows the arrival time of the direct sound and the second cursor (continuous vertical line) is placed 80 ms later. This is the usual time limit interval to calculate early and late energy indices. It can be observed how the pattern of first reflections shows complex behavior that can also be observed by ray-tracing simulation using Catt Acoustic (see Fig. 4, which shows complete and early simulated echograms for position 15 in Santa Marina church).

In order to retain the simplicity of the model, this behavior can be modeled by considering a linear level attenuation with the same rate as the classic theory for early energy and by assuming an additional attenuation. This supposes a discontinuity at $t=80$ ms that will be discussed in the next section and is shown qualitatively in Fig. 3(d) by means of the superimposed straight lines.

The best option for the evaluation of this additional attenuation appears by adjusting the early energy expression such as is written in Eq. (19) (considering $\tau=0.08$ s)

$$e_\mu = \frac{31{,}200T}{V} e^{-\frac{\mu r}{T}} \left(1 - e^{-\frac{13.82\tau}{T}}\right) \tag{19}$$

to the spatial distribution of the experimental values of C_{80} instead of those of G as in Sendra, Zamarreño, and Navarro;[8] and by supposing that the late energy behaves in the same way as in Barron's model $\ell_\mu = \ell_B$; with $\tau = 0.08$, Eq. (8). Henceforth this will be known as the μ model.[11]

An evaluation has been carried out of the differences in the total energy, or in other words of $G_\mu(r)$, by considering $\tau=50$ ms and $\tau=80$ ms limits as the temporal interval to separate early and late energy. As a function of distance in Fig. 5(a), early, late, and total temporal energy values for Santa Marina church are shown, taking $\tau=50$ ms and $\tau = 80$ ms as separation limits in Eqs. (19) and (8), respec-

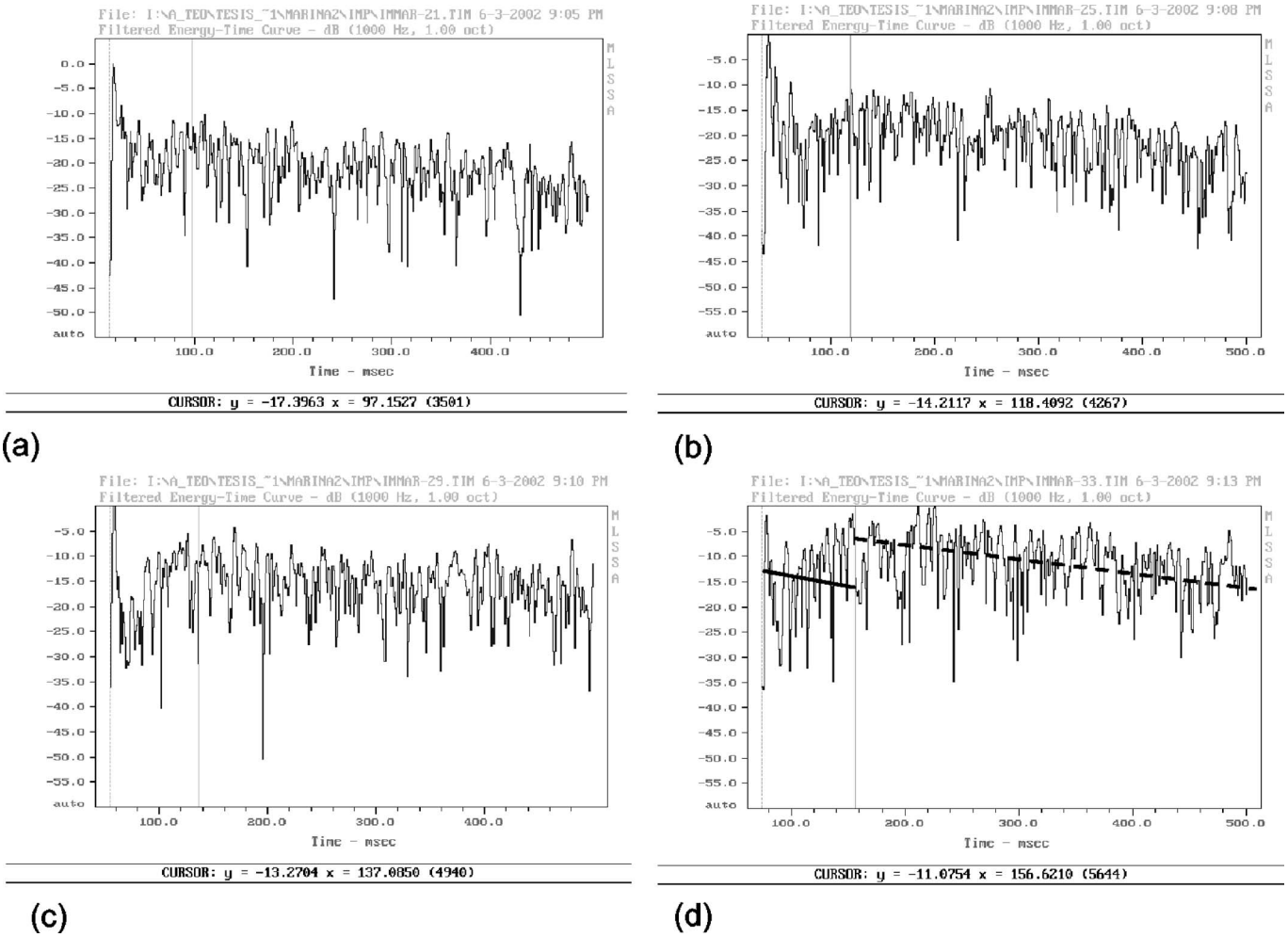


FIG. 3. Energy-time curves measured in Santa Marina church at positions 11 (a), 13 (b), 15 (c) and 17 (d) in the central nave. The 80 ms interval after the arrival of direct sound is located between vertical cursors. In (d) the qualitative behavior of the proposed model is superimposed.

tively. $G_\mu(r)$ differences (Fig. 5(b)), when choosing one or the other separation limit, are in the order of 0.4 dB at the farthest measurement points of this church, approaching the experimental values better, those of $G_{\mu 80}(r)$. In consequence, the $\tau=80$ ms limit following Barron's criterion has been used in the following sections.

The values of μ coefficient, for each church, obtained through nonlinear regression process in order to maximize the correlation between predicted and experimental values of C_{80}, are shown in Table II. As before, their volumes V and measured reverberation times T are exposed as the whole set of data required to implement this model, and in the last two lines the mean value $\bar{\mu}$ and its standard deviation σ_μ appear. This mean value can be used to describe the behavior of this specific church typology in order to estimate any parameter. On proposing this $\bar{\mu}$ mean value it is verified that the set of values (except San Julián church) are included in the $\bar{\mu}\pm 1.5\sigma_\mu$ interval.

The expression of clarity $C_{80\mu}$ (remembering this is the parameter that has been used in the regression) becomes

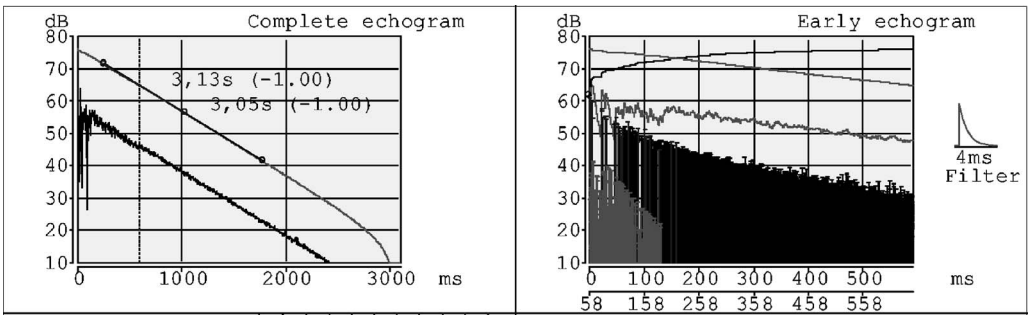


FIG. 4. Simulated echograms in Santa Marina church at position 15 in the central nave.

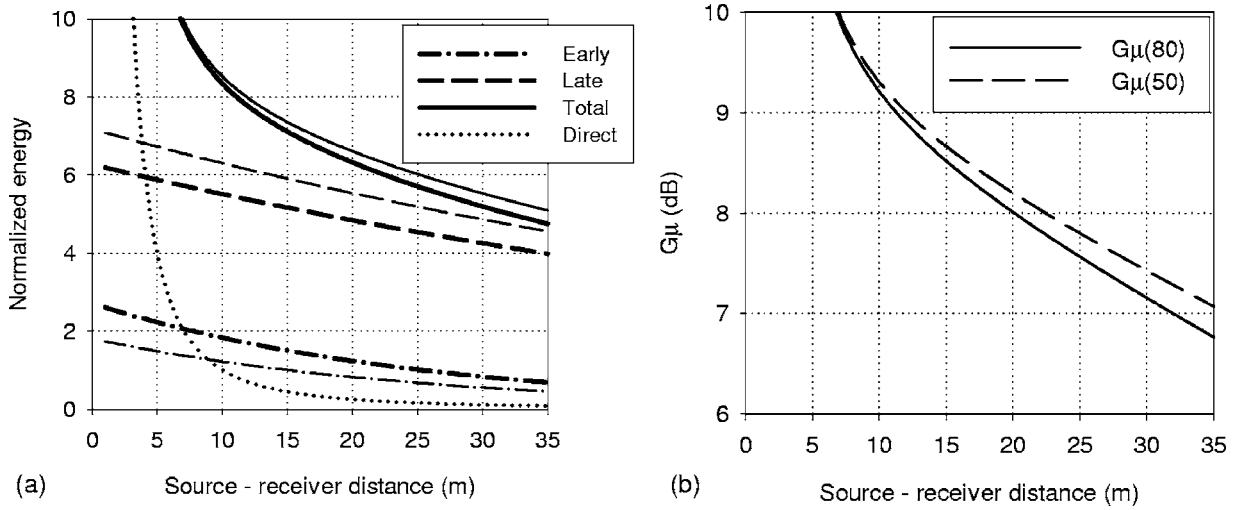


FIG. 5. Theoretical behavior of, (a) normalized energy components (fine line for $\tau=50$ ms), (heavy line for $\tau=80$ ms) and (b) their corresponding G_μ vs source-receiver distance in the proposed model, for Santa Marina church.

$$C_{80\mu}(r) = 10\log\left(\frac{d+e_\mu}{\ell_B}\right)_{\tau=0.08} = 10\log\left\{\frac{\frac{100}{r^2}+\frac{31{,}200T}{V}e^{-\frac{\mu r}{T}}\left(1-e^{-\frac{1.11}{T}}\right)}{\frac{31{,}200T}{V}e^{-\frac{1.11+0.04r}{T}}}\right\}\ (\text{dB}). \tag{20}$$

The definition will be expressed in a similar way, bearing in mind that the separation limit is now $\tau=50$ ms for e_μ in its nominator and, $\tau=80$ ms for e_μ in its denominator:

$$D_{50\mu}(r) = \frac{(d+e_\mu)_{\tau=0.05}}{(d+e_\mu+\ell_B)_{\tau=0.08}}\times 100 = \left\{\frac{\frac{100}{r^2}+\frac{31{,}200T}{V}e^{-\frac{\mu r}{T}}\left(1-e^{-\frac{0.69}{T}}\right)}{\frac{100}{r^2}+\frac{31{,}200T}{V}\left[e^{-\frac{\mu r}{T}}\left(1-e^{-\frac{-1.11}{T}}\right)+e^{-\frac{1.11+0.04r}{T}}\right]}\right\} \times 100\ (\%). \tag{21}$$

TABLE II. Input parameters to implement the proposed model

Church	V (m^3)	T (s)	μ (s/m)
Sta. Marina	10,708	3.08	0.1210
O. Sanctorum	8180	2.71	0.1400
S. Lorenzo	7040	2.28	0.1392
S. Vicente	6915	2.47	0.1571
S. Julián	6226	2.27	0.0901
S.Gil	6200	2.27	0.1200
S. Pedro	6180	2.04	0.1303
S. Andrés	5955	1.99	0.1520
S. Esteban	4746	1.87	0.1277
S. Marcos	4623	3.58	0.1604
Sta. Catalina	4362	1.60	0.1072
S. Isidoro	3947	2.21	0.1324
		Mean value $\bar{\mu}$	0.1314
		Standard deviation σ_μ	0.0205

After substituting d, e_μ, and ℓ_μ with $\tau=80$ ms in Eqs. (20) and (21), sound strength is written as

$$G_\mu(r) = 10\log(d+e_\mu+\ell_B) = 10\log\left\{\frac{100}{r^2}+\frac{31{,}200T}{V}\times\left[e^{-\frac{\mu r}{T}}+e^{-\frac{1.11}{T}}\left(e^{-\frac{0.04r}{T}}-e^{-\frac{\mu r}{T}}\right)\right]\right\}\ (\text{dB}), \tag{22}$$

which in the case $\mu=0.04$ sm^{-1}, Eq. (22) would coincide with Barron's proposal and so corresponds to the whole model. As shown in Table II, μ values are significantly larger than 0.04 in all these Mudejar-Gothic churches.

Only partial C_{80} estimated values (for 1 kHz octave band) versus source-receiver distances are available from some papers for different churches.[9,10] From these values their corresponding μ coefficients have been estimated. These corresponding attenuation factors μ for the early energy are volume dependent (Fig. 6(a)). This dependence increases when μ is plotted on a graph versus the geometric ratio V/L, as Fig. 6(b) shows. The V/L ratio can be taken as a measure of the area of transversal section and so the higher the values of V/L are, the more attenuation for early energy is expected. Some acoustic characteristics have an additional incidence over the attenuation of the early energy: the presence of scattering surfaces such as altarpieces, altars and pillars, and the timber roof characteristic to these Mudejar-Gothic churches; in this case lateral naves and chapels act as coupled rooms which disminish the reflected energy from the laterals to the main volume. Although more precise measurements are required, it is possible to establish that the μ value for each Mudejar-Gothic church can be substituted by their mean value in order to characterize this ecclesiastical typology. Deeper analyses into this topic are accomplished in Sec. V.

An expression for the instantaneous normalized reflected energy density $\epsilon'(t,r)$, deducible from e_μ and $\ell_\mu=\ell_B$, Eqs. (19) and (8), respectively, would be

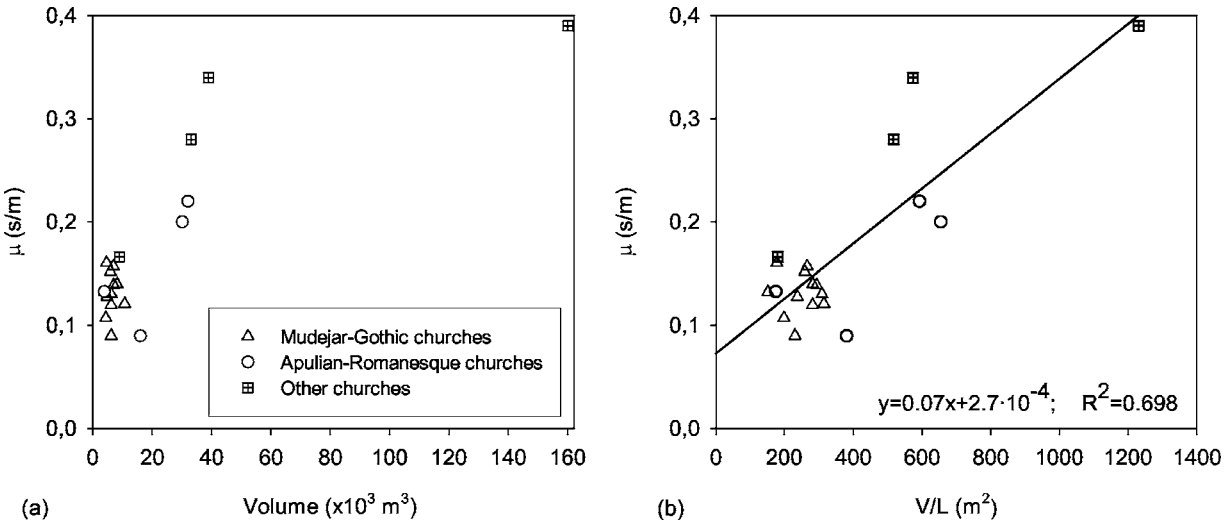


FIG. 6. Values of μ for Mudejar-Gothic and other churches vs volume (a) and vs volume-length ratio (b).

$$\epsilon'_{\mu}(t,r) = \frac{\epsilon(t,r)}{E_{D10}} = \begin{cases} \dfrac{13.82 \cdot 31{,}200}{V} e^{-\frac{\mu T}{T}} e^{-\frac{13.82t}{T}} \ (s^{-1}) & 0 \leqslant t \leqslant 80 \text{ ms}, \\ \dfrac{13.82 \cdot 31{,}200}{V} e^{-\frac{0.04r}{T}} e^{-\frac{13.82t}{T}} \ (s^{-1}) & 80 \text{ ms} < t < \infty, \end{cases} \tag{23}$$

where $t=0$ coincides in this expression with the arrival of direct sound at r location. This reflected energy proposal supposes that the decay traces can be modeled in two steps: the first beginning at the arrival of direct sound at that location where stationary reflected energy is reduced by a factor $e^{-\frac{\mu r}{T}}$ with respect to the classic expression, which is distance dependent and finishes at 80 ms; the second step from 80 ms on corresponds to that decay considered in Barron's hypothesis. Figure 7 shows the magnitude of the step at 80 ms, as a function of source-receiver distance for Santa Marina and Santa Catalina churches. Santa Catalina church is the most profusely decorated, and in this way increases the scattering properties of its surfaces and shows major attenuations for all distances.

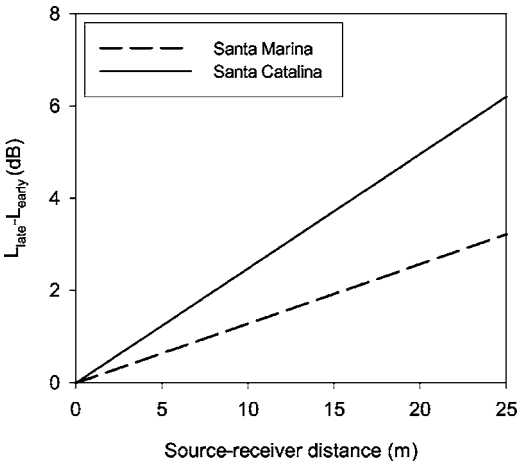


FIG. 7. Differences between early and late energies at $t=80$ ms for the μ model as a function of source-receiver distance for two significant churches.

In this way center time can be deduced by means of

$$T_{S\mu}(r) = \frac{\int_0^{0.08} t\epsilon'_{\mu}(t,r)dt + \int_{0.08}^{\infty} t\epsilon'_{\mu}(t,r)dt}{d + e_{\mu} + \ell_B} \ (s), \tag{24}$$

which, according to Eq. (23), becomes an immediate integral. Although this proposed normalized reflected energy density presents a discontinuity at 80 ms, it should be observed that the extinction process is physically discrete and hence this approximation to reality cannot be considered as inadequate at all, particularly when taking into account the very good agreement between the experimental results and the theoretical values for center time (see Figs. 8(d) and 11(d), and Tables III and IV), whose calculations have been carried out with this proposed discontinued normalized reflected energy density.

V. DISCUSSION OF THE MODEL

The procedure established here has been applied to the set of churches under study in order to check its reliability. Figure 8 shows the predictions with this μ adjustment and the experimental values for the four energy-based parameters analyzed: clarity, definition, sound strength, and center time versus source-receiver distance in Santa Marina church.[11] Likewise, Figs. 9–11 show the same theoretical predictions and the experimental results for San Julián, San Esteban, and Santa Catalina churches, respectively. Every single graph shows a significant deviation from Barron's predictions in all parameters studied which has been taken as a reference model that is able to predict the attenuations of early-to-late energy parameters with source-receiver distances.

Both models are based on modeling first reflections before direct sound reaches the reception position. In this sense, the sequence and magnitude of early reflections are

TABLE III. Differences between measured values, calculated values according to Barron's model, and the proposed model which have been spatially averaged for the four energy-based parameters.

	C_{80} (dB)		D_{50} (%)		G (dB)		T_s (ms)	
Church	Barron	Proposed	Barron	Proposed	Barron	Proposed	Barron	Proposed
Sta. Marina	2.7	0.6	6.7	0.1	0.8	0.1	46	−14
O. Sanctorum	2.1	−0.5	6.8	−2.0	1.3	0.5	37	2
S. Lorenzo	2.4	0.0	7.5	−1.5	1.4	0.5	32	3
S. Vicente	2.6	−0.4	9.9	−0.6	1.2	0.3	42	2
S. Julián	1.9	0.3	3.7	−1.9	1.5	0.7	18	6
S. Gil	2.1	−0.3	4.6	−4.2	2.1	1.2	28	8
S. Pedro	1.9	−0.3	7.7	−1.2	0.0	−0.9	22	11
S. Andrés	3.4	0.3	12.6	0.6	−0.1	−1.2	35	9
S. Esteban	2.4	−0.5	9.8	−1.5	0.9	−0.5	27	14
S. Marcos	3.0	0.2	7.4	0.5	1.5	0.7	37	−1
Sta. Catalina	1.2	−0.6	6.1	−2.2	0.7	−0.2	14	14
S. Isidoro	2.6	−0.2	10.7	0.1	2.0	0.9	39	3
Average	2.4	−0.1	7.8	−1.2	1.1	0.2	31	5

strongly influenced by the geometric and acoustic characteristics near the source and near the different reception positions, especially the scattering properties of these surfaces as has been suggested by Chiles[23] and Cirillo.[10]

In this case, the source position is in the main altar where the usual presence of elaborately decorated altarpieces and furniture produce scattered reflected energy near the source which is seldom found in concert halls. In a similar way, near each receiver position we can find pillars, lateral altars and chapels, wooden pews, a diffusing timber roof (the main characteristic of these churches), which all have a scattering effect on the early reflected energy. In this manner the patterns of first reflections are different in churches to other buildings. In order to retain the simplicity of Barron's model, additional attenuation has been considered for the early reflected energy as shown in Eq. (19). It has also been checked

TABLE IV. Mean values of the absolute value differences between those calculated from Barron's model and from the proposed model and the measured values in each position (first row), and their corresponding standard deviation (second row) for each church.

	C_{80} (dB)		D_{50} (%)		G (dB)		T_s (ms)	
Church	Barron	Proposed	Barron	Proposed	Barron	Proposed	Barron	Proposed
Sta. Marina	1.4	0.9	7.6	3.6	0.9	0.4	47	20
	0.9	0.8	3.8	3.4	0.5	0.3	24	12
O. Sanctorum	2.2	1.2	9.4	5.3	1.5	0.9	40	20
	1.9	1.0	5.5	4.9	0.9	0.7	30	14
S. Lorenzo	2.4	0.5	8.8	3.1	2.0	1.2	35	10
	1.3	0.3	4.6	2.9	0.8	0.8	16	8
S. Vicente	2.8	1.2	11.7	4.5	1.4	0.7	46	16
	1.0	0.9	3.1	4.6	0.5	0.6	13	13
S. Julián	2.2	1.2	9.4	6.1	1.5	0.7	26	12
	1.5	0.6	5.7	6.0	0.6	0.4	14	11
S. Gil	2.2	0.9	7.3	5.3	2.2	1.5	32	14
	1.5	0.6	4.3	4.8	1.1	0.6	21	10
S. Pedro	2.3	1.5	11.7	8.5	0.6	0.9	32	21
	1.3	1.5	6.3	8.4	0.6	0.8	15	27
S. Andrés	3.4	1.8	12.6	6.4	0.6	1.3	35	24
	2.6	1.8	8.5	5.0	0.5	0.6	24	11
S. Esteban	2.5	0.7	10.2	3.3	1.3	0.8	30	14
	1.2	0.4	6.2	2.2	0.7	0.7	15	9
S. Marcos	3.2	1.4	7.6	3.3	1.7	1.0	44	25
	1.5	0.9	4.0	1.9	0.6	0.6	23	15
Sta. Catalina	2.4	1.2	13.9	7.4	1.5	1.0	28	15
	1.1	1.1	4.9	6.6	0.7	0.7	12	13
S. Isidoro	2.7	0.7	10.9	3.7	2.2	1.3	41	13
	1.5	0.6	5.8	2.3	1.0	0.6	24	12
Average	2.5	1.1	10.1	5.0	1.5	1.0	36	17
	1.4	0.9	5.2	4.4	0.7	0.6	19	13

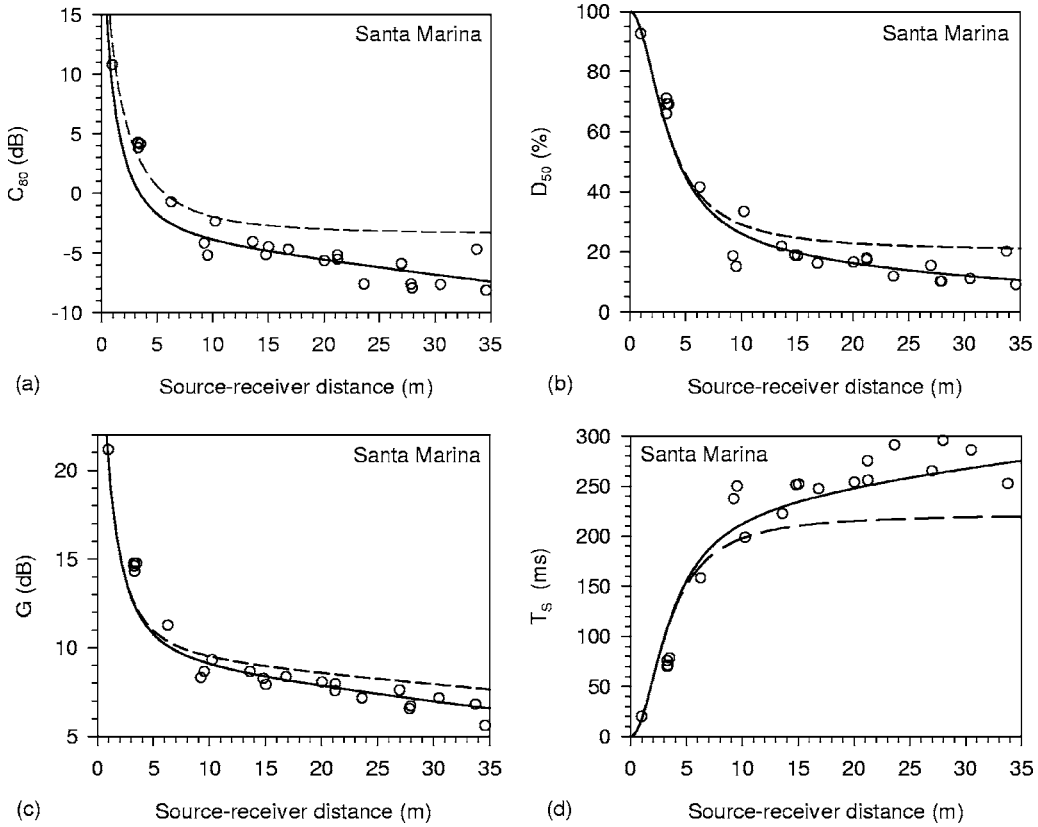


FIG. 8. Santa Marina church, measured values (o) and calculated values from Barron's model (dashed line) and from the proposed model (continuous line), vs source-receiver distance: (a) C_{80}, used to obtain μ through nonlinear regression, (b) D_{50}, (c) G, and (d) T_S.

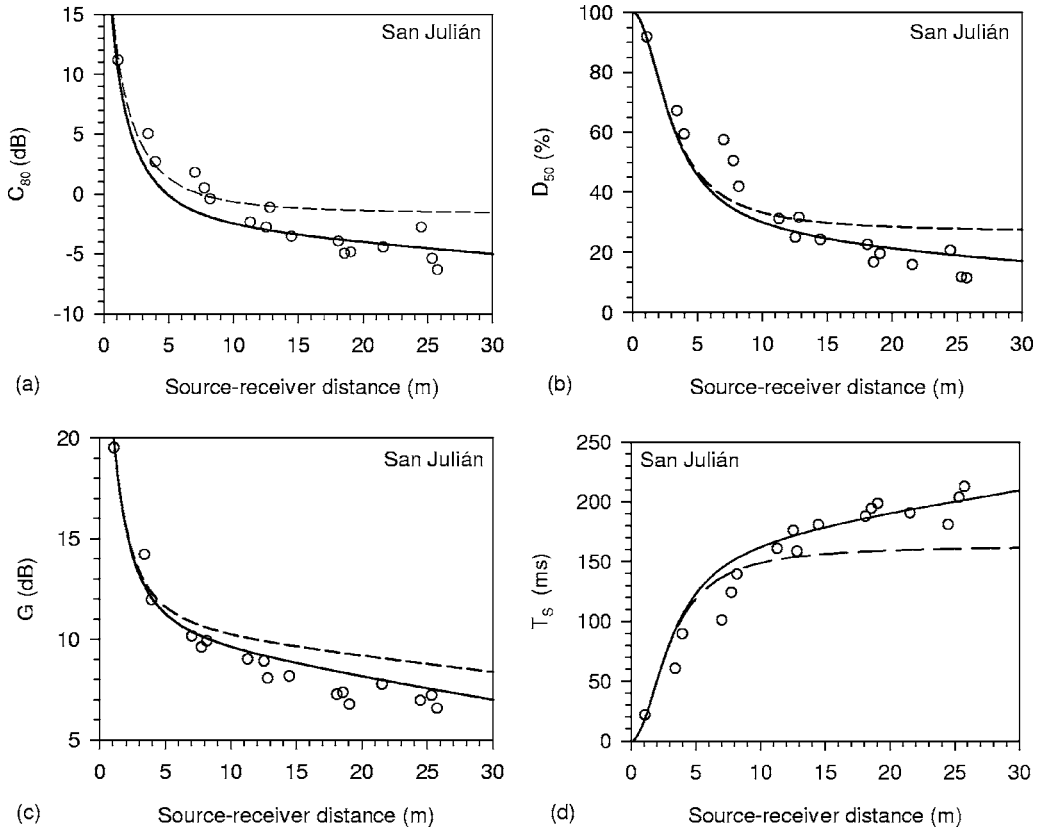


FIG. 9. San Julián church, measured values (o) and calculated values from Barron's model (dashed line) and from the proposed model (continuous line), vs source-receiver distance: (a) C_{80}, used to obtain μ through nonlinear regression, (b) D_{50}, (c) G, and (d) T_S.

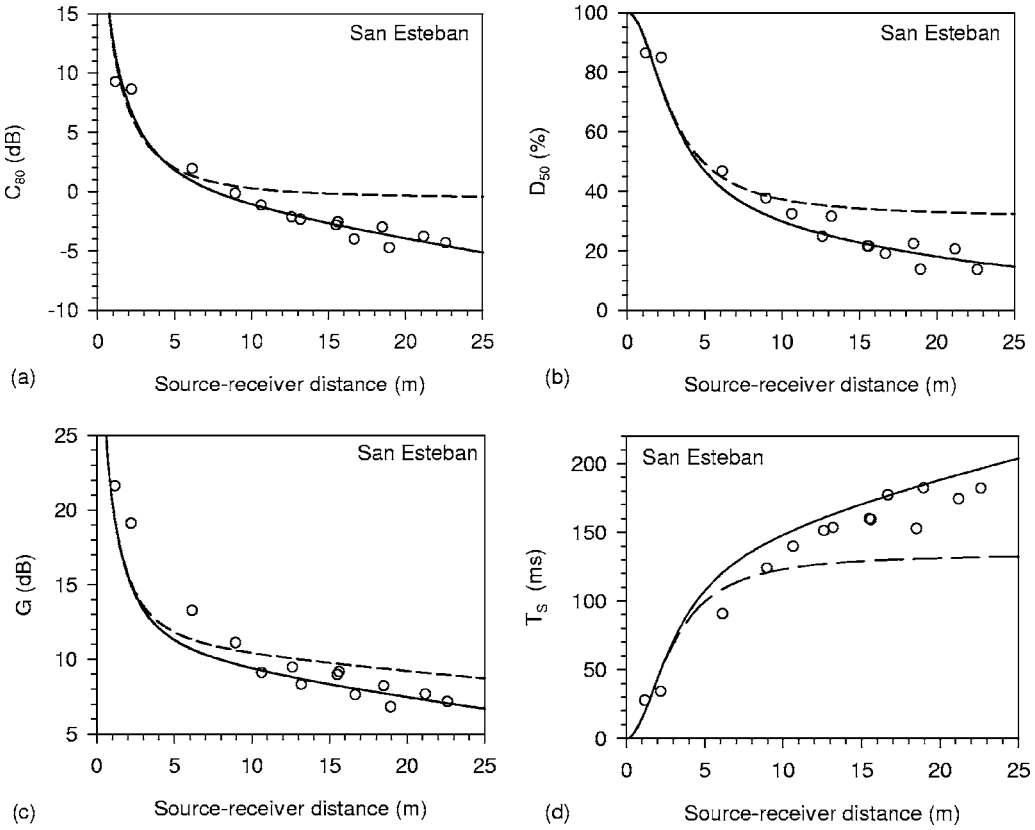


FIG. 10. San Esteban church, measured values (o) and calculated values from Barron's model (dashed line) and from the proposed model (continuous line), vs source-receiver distance: (a) C_{80}, used to obtain μ through nonlinear regression, (b) D_{50}, (c) G, and (d) T_S.

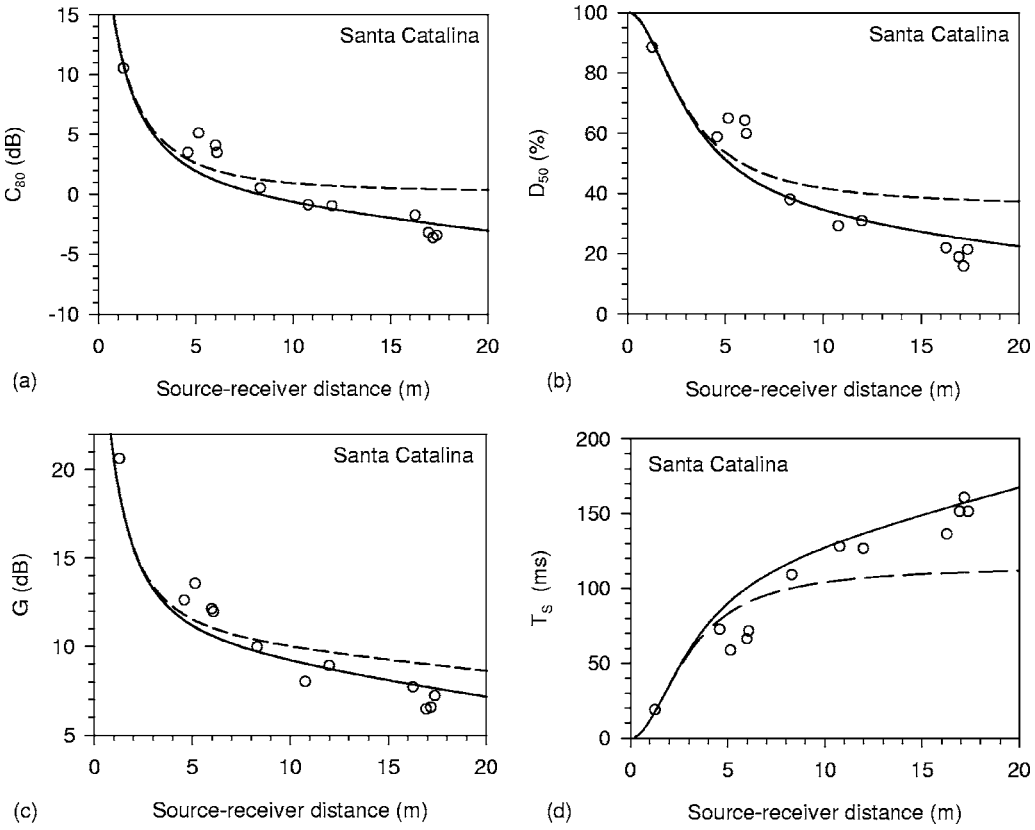


FIG. 11. Santa Catalina church measured values (o) and calculated values from Barron's model (dashed line) and from the proposed model (continuous line), vs source-receiver distance: (a) C_{80}, used to obtain μ through nonlinear regression, (b) D_{50}, (c) G, and (d) T_S.

in some cases that, in accordance with Carvalho's[17] paper, the apse (where the source is located) behaves as a coupled space.

From the set of Figs. 8–11, it may be highlighted that, due to the coupling effect, there are differences between predictions and measurements in points belonging to the presbyteries at short distances from the source. In the farthest points belonging to the lateral naves some discrepancies also exist due to the scattering effect near the reception positions.

The results of the calculation both through the proposed model and Barron's model have been compared with the measured data in each church. Table III shows the differences between measured and calculated values according to Barron's model and the proposed model which have been spatially averaged for the four energy-based parameters. Since C_{80} values have been used to obtain the μ coefficients and thereby the proposed model is tuned to experimental data, the differences between measured and calculated spatially averaged values are slight for the proposed model. This observation is taken into account below and hence Barron's model is considered as a reference and not as a comparison. Clarity values vary from 3.4 dB for San Andrés church to 1.2 dB for Santa Catalina church in Barron's model (average, differences of 2.4 dB for the set of churches). For the proposed model this interval is reduced to 0.6 dB for Santa Marina church to −0.6 dB for Santa Catalina church with average differences of −0.1 dB for all churches. Obviously Barron's model tends to overestimate C_{80} values.

The differences between measured and calculated spatially averaged values for D_{50} oscillates from 12.6% for San Andrés church to 3.7% for San Julián church in Barron's model and between 0.6% for San Andrés church and −4.2% for San Gil church in the proposed model. For all churches the averages of these differences are 7.8% for Barron's model and −1.2% for the proposed model. In this case Barron's model also overestimates D_{50} values, whereas the proposed model fluctuates between overestimating in some cases and underestimating in others.

For G parameter the range of variation oscillates between 2.1 dB for San Gil church and −0.1 dB for San Andrés church with an average value for all churches of 1.1 dB for Barron's model, whereas for the proposed model the oscillation is between 1.2 dB for San Gil church and −1.2 dB for San Andrés church and an average value of 0.2 dB. Although in this parameter Barron's model may suppose a better approach in some churches (San Andrés) average values lean clearly towards the proposed model.

The improvement in prediction for T_S values is remarkable in favor of the proposed model. The differences from spatially averaged values from Barron's model vary between 46 ms for Santa Marina church and 14 ms for Santa Catalina church. In the proposed model this interval is reduced to 14 ms for Santa Catalina church and −14 ms for Santa Marina church with an average difference of 31 ms for Barron's model against 5 ms in the proposed model.

Continuing with this analysis, Table IV shows the average absolute values (first row for each church), and their respective standard deviation (second row for each church) of the differences between predicted parameter values and measured ones for each position. In this way it is possible to assess the improvement that the proposed model supposes on predicting the variation of the different parameters with the source-receiver distance. For these temples mean values of those differences in C_{80} are significantly lower in the proposed model, as can be seen in Table IV. Moreover, its dispersions are notably lower (except San Pedro church). For the whole set of churches, mean value of the differences disminishes from 2.5 dB obtained in Barron's model to 1.1 dB in the proposed model and the mean value of the standard deviation is reduced from 1.4 to 0.9 dB.

Something similar happens for the D_{50} parameter: in each church average differences are significatively lower in the proposed model than in Barron's model, and in the whole set mean value is reduced from 10.1% for Barron's model to 5.0% in the new proposal. Dispersions of these differences from the mean value in each church are higher in the proposed model in four churches: San Vicente, San Gil, San Pedro, and Santa Catalina. For the set as a whole, the mean value of the dispersions decreases from 5.2% for Barron's case to 4.4% in the proposed model.

The same comments can be applied to the prediction of the proposed model in the spatial distribution of G values for all churches, except San Pedro and San Andrés. For this parameter, the accuracy of the model proposed is less notable due to the lower relative weight of the early energy. The differences in the dispersions described by standard deviation are similar in both models. For the whole set of churches the average of the differences moves from 1.5 dB (Barron), to 1.0 dB (proposed model). The average value of standard deviation is similar (0.7 and 0.6 dB, respectively).

Finally, it should be mentioned that the greatest improvement appears in the T_S values. The differences are reduced by half in each church for all churches and the average difference is reduced from 36 ms (Barron) to 17 ms (proposed model). The standard deviation is also reduced from 19 to 13 ms.

Table III and Table IV data confirm that a considerable improvement is provided by the proposed model for this kind of space as compared with Barron's revised theory, in clarity, definition, and center time and a more modest improvement for G.

The analysis is extended in Figs. 12(a) and 12(b) by means of a comparison between the different temporal energy components calculated through: Barron's model (B); Cirrillo's model (C), estimating from their suggestions[10] a $\rho=0.0058$ sm^{-1} value and a dispersion coefficient $s=0.20$; and the proposed model (P); all versus source-receiver distance for both Santa Marina and San Marcos churches. In these figures, the direct sound, which is common to the three models, their respective early reflected energy (received until 80 ms after the arrival of direct sound), their late reflected energy (from 80 ms to infinity), and total energy, respectively, are shown.

Several aspects of the graphs should be emphasized. In the early reflected energy, Barron's values present a great deviation from the two other models; on comparing Cirillo's and the proposed μ adjustment, the former gives greater values in the vicinities of the source and greater attenuations at

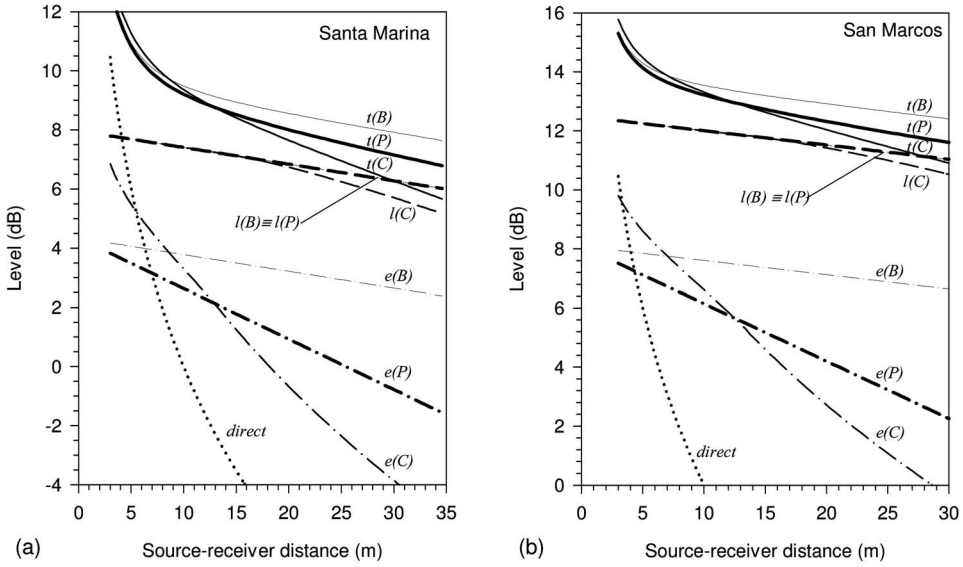


FIG. 12. Dependence for the different temporal energy components (e, early; ℓ, late; t, total), as a function of source-receiver distance, calculated from Barron's model (B), Cirillo's model (C), and the proposed in this work (P), (a) for Santa Marina church, and (b) for San Marcos church.

longer distances than the proposed μ adjustment. For the late reflected energy the three models practically coincide as expected and, for the total energy, Barron's value shows a remarkable overestimation while a notable coincidence for the other two models may be highlighted. It is only at great distances that Cirillo's model presents some additional attenuation (about 0.6 dB at 25 m from the source).

Finally, this method is applied to some churches analyzed by Cirillo and Martellota, although only partial data (1 kHz octave band) of C_{80} are used to obtain the corresponding values of μ parameters as described in Fig. 6. The results for St. Nicholas Basilica in Bari and Lucera Cathedral are shown in Fig. 13, where the experimental data and the predicted values from Cirillo's model and μ model are plotted as a function of source-receiver distance. These partial results suggest the possibility of a generalization of this methodology so that it can be applied to a wide range of architectural typologies whereby a range of values of μ capable of predicting the acoustic energy parameters from basic geometric and acoustic data of the space are provided.

From the presented results it is possible to propose a typological model to evaluate and predict acoustic energy parameters for these religious spaces. The proposal is to use the $\bar{\mu}$ value and its geometric (V) and acoustic (T) characteristics for each church. To this end, Fig. 14 shows the T_S function calculated from this typological model (through Eq. (24) with $\bar{\mu}$ value), those values from their corresponding μ in Eq. (24), and the experimental results versus source-receiver distance, for four churches: San Julián which is in the lower limit of the μ range, Santa Marina and San Esteban which are in the middle range, and San Marcos in the upper limit. Only slight differences appear between the two theoretical lines. The differences found at the farthest point from the source are, 23 ms (11%) at 30 m for San Julián church, 4.4 ms (1.6%) at 35 m for Santa Marina church, 1.9 ms (0.9%) at 25 m for San Esteban church, and 9.6 ms (3%) at 30 m for San Marcos church, respectively.

Consequently, with the calculation of μ coefficients developed in these churches, a relatively simple model based on C_{80} experimental data may be obtained which provides an

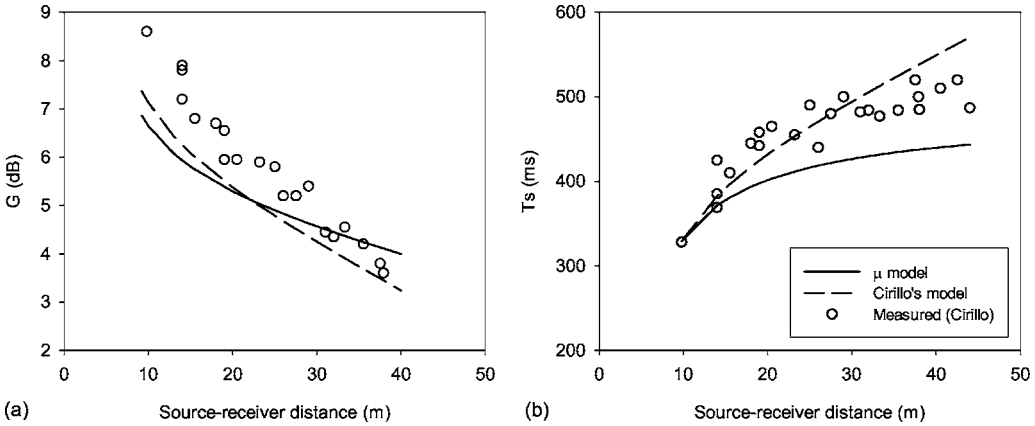


FIG. 13. Measured and predicted values from Cirillo's model and the proposed μ model of G for St. Nicholas Basilica in Bari (a) and of T_S for Lucera Cathedral (b).

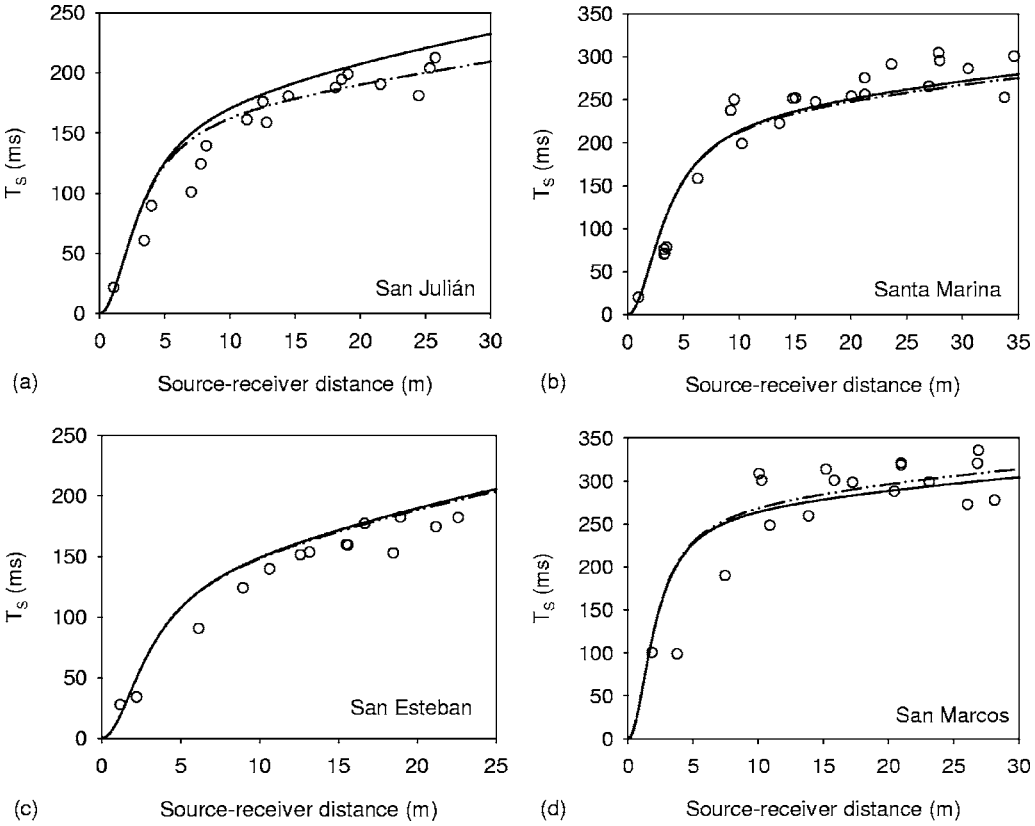


FIG. 14. Comparison of theoretical T_S function calculated from $\bar{\mu}$ (continuous line) and from its μ (dotted line) value, and the experimental results (o) vs source-receiver distance, (a) for San Julián church, (b) for Santa Marina church, (c) for San Esteban church and (d) for San Marcos church.

acceptable prediction of the values of the most important parameters in order to evaluate acoustic quality of rooms based on sound energy criteria. By rigidly maintaining the upper limit of the early energy interval at 80 ms for all reception points, Eq. (23) differs from Cirillo's model which establishes a variable time for each position in which the normalized energy density changes its behavior from linear to exponential. Hence certain difficulties arise for evaluating the early energy for clarity calculations according to the positions of 80 ms time with respect t_R, and although Cirillo's proposal is a very coherent and elegant theoretical model, the procedure developed here avoids these complications and also yields similarly good results in its predictions.

VI. CONCLUSIONS

In the 12 Mudejar-Gothic churches under study an analysis has been undertaken of the most relevant energy-based monoaural parameters: clarity, definition, sound strength, and center time, as a function of source-receiver distance. These parameters have been averaged spectrally in the most widely accepted way suggested for concert hall and auditoria in order to describe sound quality. The limitation of the previous adjustment model adopted by Sendra, Zamarreño, and Navarro[8] for this type of religious building is that it is only capable of justifying the behavior of the total sound pressure level with source-receiver distance. This limitation has been overcome by using the procedure developed in this work. The proposed methodology involved calculating the early energy by adjusting the μ coefficient introduced in the distance-dependent exponential through nonlinear regression in order to maximize the correlation between the predicted and the experimental values of clarity in these spaces, and by maintaining the late energy expression suggested by Barron's model. The input variable μ, which takes into account the conformation of the boundary surfaces, interior decoration and furniture in those spaces, together with the input variables V, T, and r, enables all the remaining acoustic parameters (definition, sound strength, and center time) to be deduced in these worship enclosures with acceptable accuracy.

The accuracy of the methodology has been checked by comparing spatially averaged measured values for each parameter with those calculated using Barron's model and the proposed model. Likewise, a comparison between mean values of absolute differences for each position in each church has been analyzed. In both cases the results confirm the prevalence of the proposed model for these religious spaces.

From the set of calculated μ coefficients, a typological model is proposed using the $\bar{\mu}$ value and the specific geometrical (V) and acoustic (T) data for each church. This typological model has been verified extensively through the T_S parameter.

The similarity of the predictions for the total early and total temporal components of the energy with Cirillo's model and the one proposed here is remarkable, and although Cirillo's study is a notable theoretical contribution to the insight of sound field in churches, more simple relations are deduced with the typological model proposed in this work.

Finally, it is necessary to mention that the methodology accomplished here may be extended to include other types of closed rooms, thereby indicating a greater prediction potential. This would require more intensive in situ measurement to establish the μ values for different typologies. In order to verify the feasibility of this extension, the μ model methodology has been applied to the churches analyzed by Cirillo and Martellota[10] and, despite only partial data having been used, the results nevertheless point towards the validity of this potential.

ACKNOWLEDGMENTS

The authors would like to thank the priests and church management for allowing the measurements to be carried out. This work has been partially supported by the Spanish MCYT Project No. BIA2003-09306-CO4-02.

[1]A. P. O. Carvalho, "Influence of architectural features and styles on various acoustical measures in churches," Ph.D. dissertation, University of Florida (1994).
[2]J. J. Sendra, T. Zamarreño, and J. Navarro, "Acoustics in churches," *in Computational Acoustics in Architecture*, edited by J. Sendra (Witpress, Southampton, 1999), pp. 133–165.
[3]A. A. Adel, "Measurement of acoustical characteristics of mosques in Saudi Arabia," J. Acoust. Soc. Am. **113**, 1505–1517 (2003).
[4]*Proceedings of the 17th International Congress on Acoustics*, CD-ROM Vol. **III**. Architectural Acoustics: Acoustics of Worship Buildings (14 communications) (ISBN 88-88387-02-1) Rome, Italy, 2–7 September (2001).
[5]*Proceedings of the Forum Acusticum*, CD-ROM: ARC Architectural Acoustics: RBA-05: Worship Building Acoustics (16 communications) (ISBN 84-87985-07-6) Seville, Spain, 16–20 September (2002).
[6]Z. Karabiber, "A new approach to an ancient subject: CAHRISMA project," on CD-ROM: 4–7 July, *Seventh International Congress on Sound and Vibration*, Garmisch-Partenkirchen, Germany (2000).
[7]M. Barron and L. J. Lee, "Energy relations in concert auditoriums I," J. Acoust. Soc. Am. **84**, 618–628 (1998).
[8]J. J. Sendra, T. Zamarreño, and J. Navarro, "An analytical model for evaluating the sound field in Gothic-Mudejar churches," in *Computational Acoustics and its Environmental Applications II*, edited by C. A. Brevia, J. Kenny, and R. D. Ciskowski (Computational Mechanics, Southampton, 1997), pp. 139–148.
[9]E. Cirillo and F. Martellota, "An improved model to predict energy-based acoustic parameters in Apulian-Romanesque churches," Appl. Acoust. **64**, 1–23 (2003).
[10]E. Cirillo and F. Martellota, "Sound propagation and energy relations in churches," J. Acoust. Soc. Am. **118**(1), 232–248 (2005).
[11]M. Galindo, "La acústica en espacios religiosos católicos: Iglesias Gótico-Mudéjares," ("Acoustics in Catholic worship spaces: Mudejar-Gothic churches"), Ph.D. thesis, University of Seville, Spain (2003).
[12]L. G. Marshall, "An acoustic measurement program for evaluating auditoriums based on the early/late sound energy ratio," J. Acoust. Soc. Am. **96**, 2251–2261 (1994).
[13]W. Ahnert and H. P. Tennhardt "Acoustics for Auditoriums and Concert Halls," in *Handbook for Sound Engineers*, edited by G. M. Ballou (Elsevier, New York, 2005) pp. 109–155.
[14]I. D. Angulo, "Arquitectura mudéjar sevillana de los siglos XIII, XIV, y XV," ("Sevillian mudejar architecture in the XIIIth, XIVth and XVth centuries"), Ayuntamiento de Sevilla (1983).
[15]M. Galindo, T. Zamarreño, and S. Girón, "Acoustic analysis in Mudejar-Gothic churches: Experimental results," J. Acoust. Soc. Am. **117**(5), 2873–2888 (2005).
[16]V. O. Knudsen and C. M. Harris, *Acoustical Design in Architecture*, 5th ed. (Acoustical Society of America, New York, 1998).
[17]A. P. O. Carvalho, "The use of the Sabine and Eyring reverberation time equations to churches," J. Acoust. Soc. Am. **97**, 3319 (1995).
[18]A. Magrini and P. Ricciardi, "Coupling effects in Christian churches: Preliminary analysis based on a simple theoretical model and some experimental results," on CD-ROM: Seville (Spain), 16–20 September, *Proceedings of the Forum Acusticum* (ISBN 84-87985-07-6) (2002).
[19]A. P. O. Carvalho, "Relations between rapid speech transmission index (RASTI) and other acoustical and architectural measures in churches," Appl. Acoust. **58**, 33–49 (1998).
[20]E. Cirillo and F. Martellota, "Acoustics of Apulian-Romanesque churches: Correlations between architectural and acoustical parameters," Build. Acoust. **10**(1), 55–76 (2003).
[21]A. Magrini and P. Ricciardi, "Churches as auditoria: Analysis of acoustical parameters for a better understanding of sound quality," Build. Acoust. **10**, 135–158 (2003).
[22]L. Cremer, H. A. Müller, and T. J. Schultz, *Principles and Applications of Room Acoustics* (Applied Science, London, 1982), Vol. 1.
[23]S. Chiles and M. Barron, "Sound level distribution and scatter in proportionate spaces," J. Acoust. Soc. Am. **116**(3), 1585–1595 (2004).
[24]M. Vorländer, "Revised relation between the sound power and the average sound pressure level in rooms and consequences for acoustic measurements," Acustica **81**, 332–343 (1995).
[25]M. Galindo, T. Zamarreño, and S. Girón, "Clarity and definition in Mudejar-Gothic churches," Build. Acoust. **6**, 1–16 (1999).
[26]H. Arau, "General theory of the energy relations in halls with asymmetrical absorption," Build. Acoust. **5**(3), 163–183 (1998).
[27]A. C. Gade, "The influence of architectural design on the acoustics of concert halls," Appl. Acoust. **31**, 207–214 (1990).

Optimum speech level to minimize listening difficulty in public spaces

Masaaki Kobayashi[a] and Masayuki Morimoto
Environmental Acoustics Laboratory, Faculty of Engineering, Kobe University, Rokko, Nada, Kobe 657-8501 Japan

Hiroshi Sato
Institute for Human Science and Biomedical Engineering, National Institute of Advanced Industrial Science and Technology, 1-1-1 Higashi, Tsukuba, Ibaraki 305-8566, Japan

Hayato Sato
Environmental Acoustics Laboratory, Faculty of Engineering, Kobe University, Rokko, Nada, Kobe 657-8501 Japan

(Received 26 May 2006; revised 27 September 2006; accepted 5 October 2006)

For ideal speech communication in public spaces, it is important to determine the optimum speech level for various background noise levels. However, speech intelligibility scores, which is conventionally used as the subjective listening test to measure the quality of speech communication, is near perfect in most everyday situations. For this reason, it is proposed to determine optimum speech levels for speech communication in public spaces by using listening difficulty ratings. Two kinds of listening test were carried out in this work. The results of the tests and our previous work [M. Morimoto, H. Sato, and M. Kobayashi, J. Acoust. Soc. Am. **116**, 1607–1613 (2004)] are jointly discussed for suggesting the relation between the optimum speech level and background noise level. The results demonstrate that: (1) optimum speech level is constant when background noise level is lower than 40 dBA, (2) optimum speech level appears to be the level, which maintains around 15 dBA of SN ratio when the background noise level is more than 40 dBA, and (3) listening difficulty increases as speech level increases under the condition where SN ratio is good enough to keep intelligibility near perfect.
[DOI: 10.1121/1.2382499]

PACS number(s): 43.55.Hy, 43.71.Gv [NX] Pages: 251–256

I. INTRODUCTION

Speech communication quality in public spaces is influenced by several factors such as speech level, background noise level, sound reflections, and so on. Many previous studies have focused on the relation between the speech communication quality (as measured by speech intelligibility scores) and speech to noise ratios (SN ratios), even though speech level is an important factor in itself. No previous study has demonstrated the relation between the speech communication quality and speech level for situations with identical SN ratios.

Bradley *et al.*[1] used the modified rhyme test (MRT) with a constant speech level of 55 dBA and with noise levels varied from 45 to 60 dBA. Sato *et al.*[2] also used the MRT with a constant noise level and speech levels varying from 39 to 66 dBA. Although they could show the SN ratio where speech intelligibility scores reached ceiling, they did not determine the optimum speech level for high speech communication quality, since only SN ratio was used as a parameter in those tests.

Pearsons *et al.*[3] reported speech levels of real voices in various everyday situations, but did not refer to speech level from the standpoint of the speech communication quality. Studebaker *et al.*[4] demonstrated that word intelligibility is near perfect for speech levels varying from 64 to 99 dB in quiet. However, it is easy to imagine that a speech level of 99 dB is too loud to be acceptable in normal speech communication. There must be an optimum or preferred speech level lower than 99 dB. ISO9921:2003 (Ref. 5) mentions that the quality of speech communication is affected by not only the speech intelligibility, but also the vocal effort. However, ISO9921:2003 (Ref. 5) mentions only the allowance of the vocal effort in various cases but does not mention the relation between speech communication quality and speech level, which is one of the factors of the vocal effort. Thus, it is important to determine optimum speech levels for various background noise levels to realize the good speech communication quality in public spaces.

Another possible reason why no previous study has demonstrated the relation between speech level and speech communication quality is that there was no appropriate subjective measure for evaluating speech communication quality in public spaces with moderate to excellent SN conditions. Speech intelligibility scores using words or sentences are the conventional subjective measures for evaluating speech communication quality. However, speech intelligibility is almost

[a]Presently at Technical Research Institute, Toda Corporation, 315 Kaname, Tsukuba, Ibaraki 300-2622, Japan. Electronic mail: masaaki.kobayashi@toda.co.jp

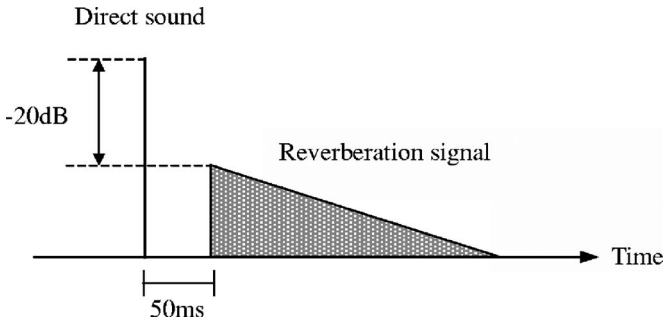


FIG. 1. Schematic diagram of impulse response of the test sound field.

always near perfect, even if the speech level is too loud as mentioned before. That is, intelligibility tests cannot evaluate speech communication quality in public spaces with higher SN ratios where intelligibility scores are all close to 100%.

Morimoto *et al.*[6] proposed the use of "listening difficulty" as a new subjective measure for the evaluation of speech quality in public spaces and demonstrated that it can evaluate speech communication quality more accurately and sensitively than word intelligibility tests for sound fields with higher SN ratios. Sato *et al.*[3] indicated that word intelligibility scores do not fall below 90%, whereas listening difficulty ratings vary from 0 to 100% for SN ratios from 0 to 15 dB, which is typical of those conditions found in most everyday situations.[3]

In this paper, the following two listening tests were carried out in order to determine the relation between the optimum speech level and the background noise level by using listening difficulty ratings. In the first test, only speech level was changed without noise. In the second test, both speech and background noise levels were changed for constant SN ratio. The results of two listening tests and our previous work[6] which background noise level was changed while speech level was constant are jointly discussed for suggesting optimum speech level.

II. LISTENING TEST I: EFFECTS OF SPEECH LEVEL ON SPEECH COMMUNICATION QUALITY IN QUIET

A. Test sound fields and test words

Figure 1 shows schematically the impulse response of the sound fields used in this test. The sound fields were composed of a direct sound followed by a reverberant energy, which was electrically generated by sound effect processor (YAMAHA SPX-900) and was digitized in the PC. The reverberant decay started 50 ms after the direct sound. As Bradley *et al.*[7] reported, increased early reflection energy arriving at the listener within the first 50 ms after the direct sound has the same effect on speech intelligibility scores as an equal increase in the direct sound energy. In this study, simulated sound fields are used including only the direct sound component as useful sound energy and reverberant signal as detrimental sound energy. The ratio of the initial amplitude of the reverberant decay to the direct sound amplitude was −20 dB which was least reverberant condition in previous work.[6] As a parameter, the speech level of the test signal had the following values: 25, 30, 35, 40, 45, 50, 55, 60, 65, and 70 dBA. The speech level was adjusted for each word in terms of the A-weighted peak level (slow response) at the position corresponding to the center of subject's head, when only the direct sound was presented. The variance of the level was within ±0.3 dBA. The reverberation time was set at 0.5 s, which is assumed to be measured in residential house, classrooms, small meeting rooms and so on as in most real-life situations. The frequency characteristics of the reverberation time in octave bands was flat within ±0.1 s for the center frequencies from 63 to 8 kHz. The reverberation decay increased the speech level by 4.6 dBA at the listening position. The ambient noise level at the listening position was 17.8 dBA.

The word lists developed by Sakamoto *et al.*[8] were used in the listening tests. Each word list included 50 phonetically balanced Japanese words with similar word familiarity. "Word familiarity" is a subjective variable, which indicates how familiar native speakers are with the word.[9] Each word had four syllables. The words were spoken by a female talker and recorded in an anechoic room. Four word lists with the highest word familiarity were used in the listening test. Two hundred words (50 words×4 lists) were used in total.

All of sound fields were prepared as impulse responses. Test signals were prepared by convoluting anechoically recorded words with impulse responses in PC.

B. Subjects

Nine university students (5 males and 4 females, 21–24 years old) and one adult (male, 31 years old) participated in the test as subjects. All subjects had normal hearing sensitivity.

C. Procedure

In the test, 20 of the 200 words were presented to each subject for each of the ten different speech levels. Thus, each subject listened to a set of 200 test signals in total (20 words ×10 speech levels). Moreover, each word was presented to a subject only once. Two hundred test signals were arranged in a random order and divided into four units of 50 words. The four units were presented with a break between each unit. Different sets consisting of 200 test signals were presented to ten subjects so that all 2000 test signals (200 words×10 speech levels) were presented during the whole of the listening test. The interval between test signals was 6 s. The total length of each unit consisting of 50 words was 6 min.

The test signals were presented to a subject from a loudspeaker in an anechoic room. The listening positions were located 2 m from the loudspeaker. The frequency response of the loudspeaker was flat within ±5 dB from 100 to 10 kHz at the listening position.

Each subject was asked to write down each test word as they heard in Japanese and simultaneously to rate the listening difficulty into one of four categories shown in Table I. Word intelligibility and listening difficulty were obtained from the subjects' responses. The word intelligibility score is the percentage of the test words heard correctly. The listening difficulty rating is the percentage of the sum of the dif-

TABLE I. Four categories for listening difficulty.

1	Not difficult
2	A little difficult
3	Fairly difficult
4	Extremely difficult

ficulty responses that were "2," "3," or "4" (i.e., all responses except "1," where "1" corresponds to not difficult). The object of measuring listening difficulty is classifying conditions as either to "no-difficulty" or "difficulty." However, it was difficult for listeners to identify the boundary between "no-difficulty" and "difficulty" and to answer in terms of one of the two categories. The solution employed in the listening test was to ask subjects to answer in terms of one of four categories shown in Table I and then to calculate the difficulty ratings with the boundary between "no-difficulty" or "difficulty" set in analysis which is explained above.

D. Results and discussion

Figure 2 shows the results of the first listening test. Panel (a) is the result of the word intelligibility tests and panel (b) is that of the listening difficulty ratings. The abscissa indicates the speech level and the ordinate indicates the score as a percentage.

First, let us discuss the relation between scores of word intelligibility and listening difficulty. The intelligibility [panel (a)] was more than 90% for all speech levels except 25 dBA. Even for 25 dBA of speech level, it is about 80%. Thus, the intelligibility scores were not much affected by speech level. On the other hand, the listening difficulty ratings [panel (b)] were clearly affected by speech level. The scores varied by 70% (i.e., from 28.5% to 98%). These results reconfirm that listening difficulty ratings can evaluate speech communication quality more accurately and sensitively than word intelligibility scores for sound fields with higher SN ratios as previously demonstrated.[6]

Here, let us discuss the optimum speech level in terms of listening difficulty. The minimum listening difficulty is given at speech level of 50–55 dBA. When the speech level is more than 55 dBA, the listening difficulty increases as the speech level increases and reaches 80.5% at 70 dBA even though the intelligibility is near perfect. On the other hand, when the speech level is less than 50 dBA, listening difficulty increases as speech level decreases and approaches 98.0% at 25 dBA of speech level. Listening difficulty decreases as intelligibility score increases in these conditions.

As a result, the optimum speech level is 50–55 dBA in quiet, corresponding to the listening difficulty being minimized.

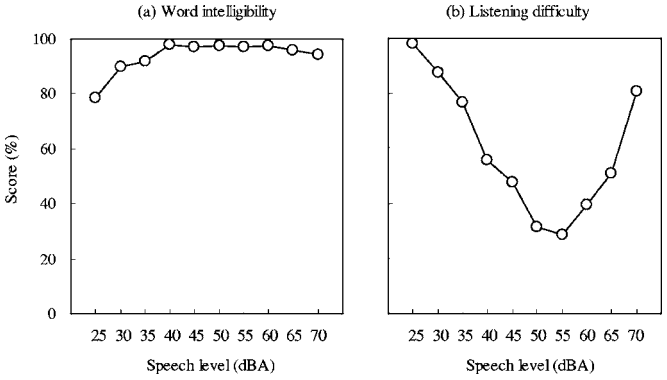


FIG. 2. Word intelligibility scores and listening difficulty ratings as a function of the speech level. The reverberation time of the sound fields are constant at 0.5 s. The panel (a) shows the word intelligibility and the panel (b) shows the listening difficulty.

III. LISTENING TEST II: EFFECTS OF SPEECH LEVEL ON SPEECH COMMUNICATION QUALITY IN SOUND FIELDS WITH BACKGROUND NOISE

A. Test sound fields and test words

In the second test, background noise was added to the sound fields used in the first test that was shown in Fig. 1. The parameters that were varied were the speech level and the SN ratio. The speech levels of the test signals were set at 50, 55, 60, 65, and 70 dBA, slow, peak response for the average of all test words at the listening position when only the direct sound was presented. The variance of the level among test words was within ±0.2 dBA. The SN ratio was set at +15 dBA or +30 dBA. Hence the total number of sound fields used in the listening test was ten (5 speech level ×2 SN ratios). The background noise was a steady-state noise and had the same power spectrum shape as that found by Hoth[10] for each octave bands from 125 to 8 kHz.

Four word lists (200 words) with high familiarity in Japanese were used. None of these lists were used in the first test.

B. Subjects

Ten university students (6 males and 4 females, 21–24 years old) with normal hearing participated to the test as subjects. Five male and four female subjects of them were the same subjects used in the first test.

C. Procedure

The procedure of experiment and timing of presentation for speech and noise are exactly the same as previous work[6] to compare results of this study with previous work in the discussion. In this test, 20 words were presented to each subject for each of 10 different sound fields. Thus, each subject listened to a set of 200 test signals in total (20 words ×10 sound fields). Moreover, each word was presented to a subject only once. Two hundred test signals were arranged in a random order and divided into four units of 50 words. The four units were presented separately. Different sets consisting of 200 test signals were presented to 10 subjects so that all 2000 test signals (200 words×10 sound fields) were presented during the whole of the listening test. The duration of each word was around 700 ms. The duration of the background noise for each test signal was around 7 s depending on the duration of test signal. The background noise was presented in advance of the test signal by 135 ms and was ended after each test signal finished. The interval between

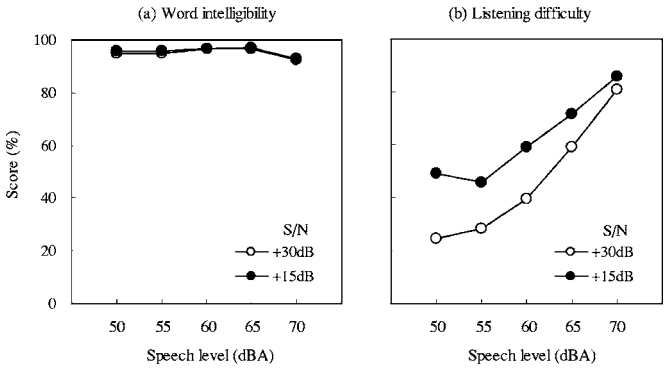


FIG. 3. Word intelligibility scores and listening difficulty ratings as a function of the speech level for each SN ratio. The reverberation time of the sound fields are constant at 0.5 s. The panel (a) shows the word intelligibility and the panel (b) shows the listening difficulty. The open and filled symbols indicate the scores for the SN ratio=+30 dBA and +15 dBA, respectively.

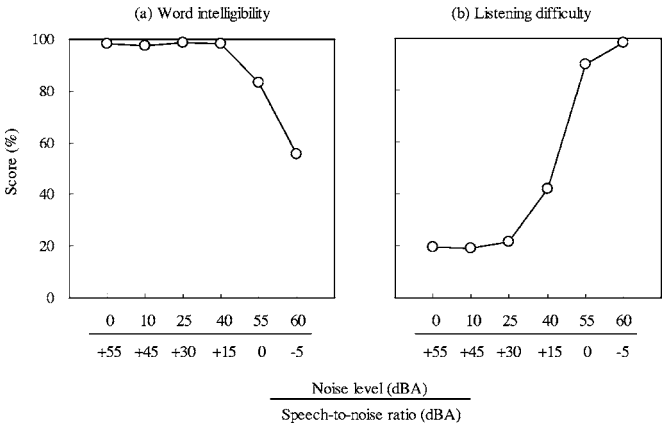


FIG. 4. Word intelligibility scores and listening difficulty ratings as a function of the background noise level and SN ratio. The reverberation time of the sound fields are constant at 0.5 s. Panel (a) shows the word intelligibility and panel (b) shows the listening difficulty [data from Morimoto *et al.* (Ref. 6)].

presentations of background noise was 6 s. During the interval no noise was presented. The total length of each unit consisting of 50 words was 10 min.

The test signals were presented to two subjects at a time from a loudspeaker in an anechoic room. The listening positions were located 2 m from the loudspeaker and at ±30° from the central axis of the loudspeaker. The frequency response of the loudspeaker was flat within ±5 dB from 100 to 10 k Hz at both listening positions.

Each subject was asked to respond by writing down each test word as they listened and to rate the listening difficulty into one of the four categories shown in Table I.

D. Results and discussion

Figure 3 shows the results of the second listening test. Panel (a) is the result of the word intelligibility test and panel (b) is that of the listening difficulty ratings. The abscissa indicates the speech level and the ordinate indicates the score as a percentage.

First, let us discuss the relation between scores of word intelligibility and listening difficulty. The intelligibility [panel (a)] was more than 90% for all test sound fields. On the other hand, the listening difficulty [panel (b)] was clearly affected by both speech level and SN ratio. The scores varied by about 60% for 30 dBA SN ratio and by 40% for 15 dBA SN ratio with changing speech level. This means that it is possible to discuss the optimum speech level in sound fields with background noise based on the listening difficulty ratings, as well as in quiet conditions.

The minimum listening difficulty is given at speech level of 50–55 dBA for each SN ratio. The minimum values were 45% and 25% for the SN ratio of 15 dBA and 30 dBA, respectively.

Listening difficulty increases as the speech level increases for both SN ratios. However, the difference between the listening difficulty ratings for SN of 15 dBA and 30 dBA, at the same speech level, decreases as the speech level increases. The difference is smallest at speech level of 70 dBA, where the listening difficulty is about 80% for both cases. This means that the speech level affects the listening difficulty more strongly than the difference in SN ratios where speech level is more than 55 dBA.

The possible reason why the difficulty ratings increase as the speech level increases for speech levels more than 55 dBA for both SN ratios may be that listeners feel speech levels more than 55 dBA are too loud even in sound fields with background noise as well as in quiet. However, if the speech level were less than 50 dBA, the listening difficulty rating would be unlikely to decrease, because it would be more difficult for listeners to recognize the words.

As a result, the optimum speech level in sound fields with background noise is probably 50–55 dBA. By most accounts, the optimum speech level in sound fields with background noise is higher than that in quiet. However, these speech levels are identical with that in quiet obtained in the first listening test. Furthermore, note that the listening difficulty at the optimum speech level for SN ratio of 30 dBA is 25% which is the same as that in quiet and it is 45% for SN ratio of 15 dBA which has been found to correspond to approximately 100% intelligibility as indicated by Bradley *et al.*[1] and others.

IV. FURTHER DISCUSSION ON OPTIMUM SPEECH LEVEL FOR SPEECH COMMUNICATION

One of the two tests performed in the previous work[6] has shown the effects of SN ratio on listening difficulty and word intelligibility. In the test, the speech level was kept constant at 55 dBA and the background noise level set at 0 (no noise was presented from the loudspeaker), 10, 25, 40, 55, and 60 dBA was added. The background noise level was controlled by the attenuator, which linearity was confirmed, and 40 dBA and higher of background noise level were actually measured by a sound level meter. Hence, background noise levels described above correspond to the SN ratios of infinity (no additional noise), 45, 30, 15, 0, and −5 dBA, respectively. The reverberation time was set at 0.5 and 2.0 s.

Figure 4 shows the results of the previous test.[6] Panel (a) is the result of word intelligibility and panel (b) is that of

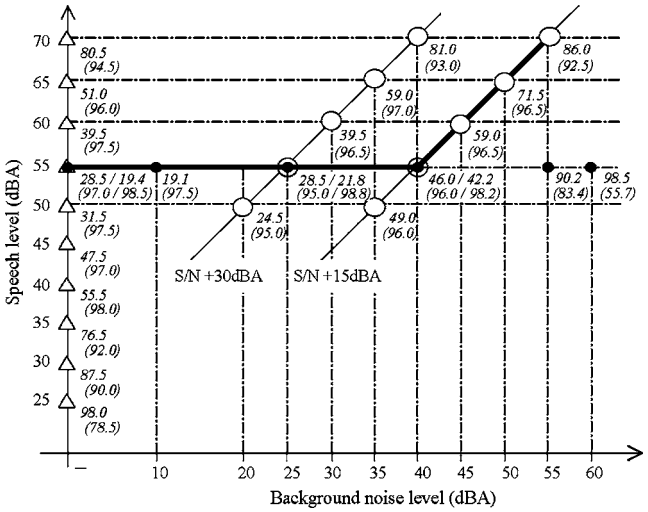


FIG. 5. Optimum speech levels as a function of speech and background noise level. The reverberation time of the sound fields are constant at 0.5 s. The open triangles and circles indicate the sound fields used in the first and second test, respectively. The filled circles indicate the sound fields used in the previous test (Ref 6). The numbers and numbers under the symbols representing the sound fields indicate listening difficulty ratings and intelligibility scores, respectively.

listening difficulty. The abscissa indicates the background noise level (SN ratio) and the ordinate indicates the score as a percentage.

From the results in Fig. 4, listening difficulty ratings can evaluate speech communication quality more sensitively and accurately than intelligibility scores when the background noise is lower than 40 dBA (i.e., SN ratio $>$ 15 dBA). Conversely, the intelligibility scores can evaluate speech communication quality more sensitively than listening difficulty when background noise is higher than 40 dBA. Namely, the results of the previous test suggest that listening difficulty ratings can evaluate speech communication quality more sensitively than the intelligibility scores for sound fields with higher SN ratios, and vice versa.

Let us now discuss the optimum speech level for speech communication quality based on listening difficulty ratings. Figure 5 shows listening difficulty ratings as functions of the speech and background noise levels. The abscissa indicates the background noise level and the ordinate indicates the speech level. The open triangles and open circles indicate the sound fields used in the first and second tests, respectively. The filled circles indicate the sound fields used in the previous test.[6] The numbers and numbers in brackets under the symbols representing the sound fields indicate the scores of listening difficulty and word intelligibility, respectively. Here, both results from two different tests are shown for the cases, background noise level of 0, 25, and 40 dBA and speech level of 55 dBA.

From these results, the intelligibility score is more than 95% when the SN ratio is more than +15 dBA and the speech level is between 40 and 65 dBA. Knudsen[11] reported that for the best conditions the percentage articulation was 96% rather than 100% because the syllable itself possesses some obscurity. Therefore, it is difficult to differentiate among these conditions by the intelligibility scores. On the other hand, the listening difficulty rating is clearly affected by both speech and background noise levels in these conditions and the optimum speech level can be evaluated sensitively by using the listening difficulty ratings.

First, let us discuss the case of sound fields with noise, which is less than 25 dBA. When there is no background noise, the minimum listening difficulty is given at speech level of 50–55 dBA as shown in the first test. When speech level is 55 dBA, there is little difference between listening difficulty at 0 (no background noise) and 25 dBA (SN ratio $=+30$ dBA) for both background noise as shown in Fig. 5. Therefore, the optimum speech level is constant at 50–55 dBA when the background noise level is 25 dBA or less.

Second, let us consider the sound fields with noise levels between 25 and 40 dBA. When the background noise level is 35 dBA, listening difficulty is 49% and 59% for speech level of 50 dBA and 65 dBA, respectively. When the speech level is 55 dBA, listening difficulty is about 25% and 45% for background noise level of 25 dBA and 40 dBA, respectively, so that it would be said that listening difficulty is between 25% and 45% at 35 dBA noise level. Therefore, listening difficulty for 55 dBA of speech level would be smaller than that for 50 dBA and that for 65 dBA of speech level when the background noise level is 35 dBA. Furthermore, as shown in Fig. 5, listening difficulty ratings for 15 dBA of SN ratio are lower than that for 30 dBA of SN ratio at the same background noise level. For example, listening difficulty ratings for 55 dBA of speech level is lower than that for 70 dBA of speech level when background noise level is 40 dBA. Thus, the optimum speech level is 50–55 dBA when background noise is 40 dBA or less.

Next, let us consider the sound fields with noise levels more than 40 dBA. When the background noise level exceeds 40 dBA, 55 dBA and higher speech level is required to reach +15 dBA SN ratio. In fact, when the background noise level is 55 dBA, speech communication quality for +15 dBA of SN ratio (i.e., speech level=70 dBA) is higher than that for 0 dBA of the SN ratio (i.e., speech level=55 dBA) for both listening difficulty and intelligibility. Another point is that louder speech levels make it more difficult for listeners to listen to speech, which is found in comparison between speech level of 55 dBA and that of 70 dBA under 40 dBA of the background noise level. Therefore, the optimum speech level would be the level required not to maintain 30 dBA of the SN ratio but around 15 dBA of the SN ratio (i.e., 19.6 dBA of the SN ratio in the case of including the reverberation). This value is similar to the SN ratio for perfect intelligibility suggested by Bradley's work.[1]

However, listening difficulty increases as background noise increases even though speech level maintains optimum level and the value is 86.0% at 55 dBA noise level. Therefore, listeners always feel difficult to listen to speech under the sound fields with high background noise level and speech intelligibility is more suitable for evaluating the optimum speech level for these sound fields as previous test[6] indicated. Furthermore, for maintaining high speech intelligibility without degrading listening difficulty, it is necessary not to increase speech level, but to reduce the background noise level.

V. CONCLUSIONS

The present paper has demonstrated the relation between the optimum speech level and background noise level by using listening difficulty. The results of two listening tests in this work and our previous work[6] indicated the following results: (1) optimum speech level is constant when background noise level is lower than 40 dBA, (2) optimum speech level appears to be the level, which maintains around 15 dBA of SN ratio when the background noise level is more than 40 dBA, and (3) listening difficulty increases as speech level increases under the condition where SN ratio is good enough to keep intelligibility near perfect.

These results indicate that listeners could find louder speech levels above a certain level more difficult to listen to speech under background noise at any noise levels. Therefore, an ideal communication situation should avoid such high speech levels. That is, amplifying speech to achieve an adequate SN ratio will not ensure a high quality communication situation; people will still find it to be difficult to listen to the speech. The preferred solution would be to reduce ambient noise levels so that intelligibility can be maintained without degrading listening difficulty.

ACKNOWLEDGMENTS

This study was partly supported by a Grant-in-Aid for Scientific Research (No. 16260292) from the Ministry of Education, Culture, Sports, Science and Technology of Japan, Research and Development Grant Program of Japan Institute of Construction Engineering and by the Twenty-First Century Center of Excellence (COE) Program "Design Strategy towards Safety and Symbiosis of Urban Space" awarded to the Graduate School of Science and Technology, Kobe University. The Ministry of Education, Culture, Sports, Science and Technology of Japan sponsored the Program.

[1] J. S. Bradley, R. D. Reich, and S. G. Norcross, "On the combined effects of signal-to-noise ratio and room acoustics on speech intelligibility," J. Acoust. Soc. Am. **106**, 1820–1828 (1999).
[2] H. Sato, J. S. Bradley, and M. Morimoto, "Using listening difficulty ratings of conditions for speech communication in rooms," J. Acoust. Soc. Am. **117**, 1157–1168 (2005).
[3] K. S. Pearsons, R. L. Bennett, and S. Fidell, "Speech levels in various noise environments," Report No. EPA-600/1–77–025, Washington, D.C. (1977).
[4] G. A. Studebaker, R. L. Sherbecoe, D. M. McDaniel, and C. A. Gwaltney, "Monosyllabic word recognition at higher-than-normal speech and noise levels," J. Acoust. Soc. Am. **105**, 2431–2444 (1999).
[5] Ergonomics-Assessment of speech communication, ISO9921 (2003).
[6] M. Morimoto, H. Sato, and M. Kobayashi, "Listening difficulty as a subjective measure for evaluation of speech transmission performance in public spaces," J. Acoust. Soc. Am. **116**, 1607–1613 (2004).
[7] J. S. Bradley, H. Sato, and M. Picard, "The importance of early reflections for speech intelligibility in rooms," J. Acoust. Soc. Jpn. **113**, 3233–3244 (2003).
[8] S. Sakamoto, Y. Suzuki, S. Amano, K. Ozawa, T. Kondo, and T. Sone, "New lists for word intelligibility test based on familiarity and phonetic balance," J. Acoust. Soc. Jpn. **54**, 842–849 (1998).
[9] http://www.ais.riec.tohoku.ac.jp/lab/wordlist/index-j.html (in Japanese). Last accessed 11/10/06.
[10] D. F. Hoth, "Room noise spectra at subscribers' telephone locations," J. Acoust. Soc. Am. **12**, 499–504 (1941).
[11] V. O. Knudsen, *Architectural Acoustics* (Wiley, New York 1932).

Steady-spectrum contexts and perceptual compensation for reverberation in speech identification

Anthony J. Watkins[a] and Simon J. Makin
Department of Psychology, The University of Reading, Reading RG6 6AL, United Kingdom

(Received 26 March 2006; revised 11 October 2006; accepted 12 October 2006)

Perceptual compensation for reverberation was measured by embedding test words in contexts that were either spoken phrases or processed versions of this speech. The processing gave steady-spectrum contexts with no changes in the shape of the short-term spectral envelope over time, but with fluctuations in the temporal envelope. Test words were from a continuum between "sir" and "stir." When the amount of reverberation in test words was increased, to a level above the amount in the context, they sounded more like "sir." However, when the amount of reverberation in the context was also increased, to the level present in the test word, there was perceptual compensation in some conditions so that test words sounded more like "stir" again. Experiments here found compensation with speech contexts and with some steady-spectrum contexts, indicating that fluctuations in the context's temporal envelope can be sufficient for compensation. Other results suggest that the effectiveness of speech contexts is partly due to the narrow-band "frequency-channels" of the auditory periphery, where temporal-envelope fluctuations can be more pronounced than they are in the sound's broadband temporal envelope. Further results indicate that for compensation to influence speech, the context needs to be in a broad range of frequency channels. [DOI: 10.1121/1.2387134]

PACS number(s): 43.55.Hy, 43.71.Gv, 43.71.Es, 43.71.An [PEI] Pages: 257–266

I. INTRODUCTION

In perceptual compensation for reverberation, a room's acoustic properties appear to inform listeners' decisions about the characteristics of speech sounds heard in the room. The mechanism responsible seems to pick up information about reverberation from lengths of running speech that extend over several words, and to use this information in ameliorating effects that reverberation might otherwise have on individual syllables in the neighborhood (Watkins, 1992, 2005a). By operating in this way, the compensation mechanism can capitalize on the fact that in real-world situations, reverberation is unlikely to change substantially during the course of a few words in a phrase or sentence. In this respect, reverberation differs from factors such as rate, which regularly changes from one syllable to the next (Summerfield, 1981). The decaying "tails" that reverberation adds at sharp offsets in sounds appear to be particularly salient information for compensation (Watkins, 2005b, 2005c). Such tails are a consequence of the smoothly decreasing energy decay that is typical of reverberation patterns in box-shaped rooms (Schroeder, 1965; Allen and Berkely, 1979).

While the presence or absence of tails over the course of a few words seems to be informative about the presence of reverberation, the tails can also obscure certain phonetic distinctions. This is because tails give rise to distortion of sounds' temporal envelopes as they reduce the steepness of offsets and fill gaps. This can affect phonetic distinctions that are cued by features of the temporal envelope, such as the [s] vs [st] distinction between "sir" and "stir" test words (Watkins, 2005a, 2005b, 2005c). These distorting effects of tails seem to arise through masking (Bolt and Macdonald, 1949; Nábělek *et al.*, 1989) and fusion (Haas, 1951; Watkins and Holt, 2000).

Tails from reverberation will be added at the ends of sounds when they have sharp offsets, but tails will also be introduced during the sound in some of its frequency bands. This is because the sound's short-term spectrum can change, forming a "spectral transition" (Furui, 1986), with a reduction of the power in some of the sound's frequency bands. These "narrow-band offsets" occur in continuous speech, as the inherently dynamic origins of this sound give rise to numerous spectral transitions. Tails at narrow-band offsets might therefore provide listeners with information about a room's acoustic properties, because the "auditory filters" that characterize peripheral auditory processing perform a narrow-band analysis of incoming sounds.

Experiments on compensation have used test words embedded in different types of "context." These contexts are typically a few spoken words in a short phrase, or they are a version of this speech that has been transformed by signal-processing manipulations. Such experiments can demonstrate compensation because they alter the reverberation between the context and the test word, which gives systematic changes in listeners' identifications when compensation occurs. These experiments have so far only shown a compensation effect when the context is a sound that contains speech-like spectral transitions, while steady-spectrum "noise contexts" give little or no compensation (Watkins, 2005a, 2005b). These noise contexts were generated from the speech contexts by turning them into signal-correlated noise (Schroeder, 1968) with an "SCN operation," which gives sounds that have the same temporal envelope as the broad-

[a]Electronic mail: syswatkn@reading.ac.uk

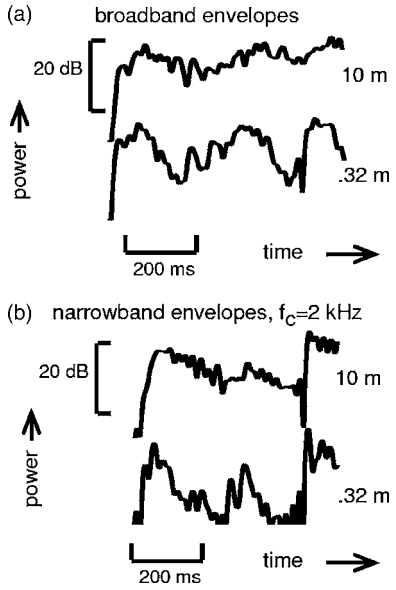


FIG. 1. (a) Temporal envelopes obtained from broadband speech contexts by full-wave rectification followed by a low-pass filter that has a corner frequency of 50 Hz. Here, the context is "Next you'll get…" that was used with the slowly spoken test words. The envelope's power (in dB) is plotted on the ordinate, and only the upper 30 dB is shown. (b) Temporal envelopes obtained as in (a), but after filtering the speech with an auditory filter centered at 2 kHz. The traces were obtained after the speech had been convolved with the left channel of a binaural room impulse response (BRIR). In both plots the source to receiver distance is 10 m in the upper trace and 0.32 m in the lower trace.

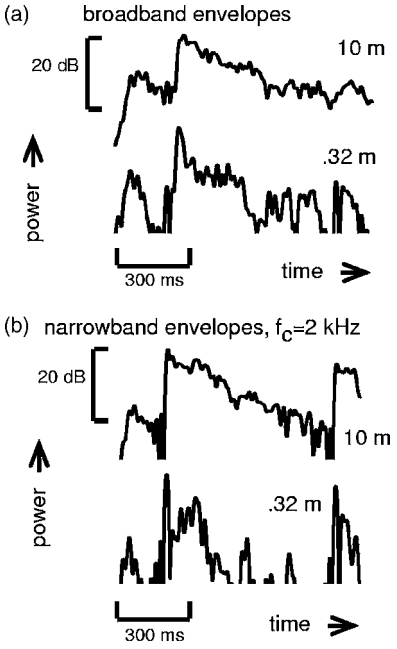


FIG. 2. Temporal envelopes from the context "OK, next you'll get…" used with the rapidly spoken test words. Other details are the same as Fig. 1.

band speech signal. However, these noise sounds have a steady spectrum in that the shape of the short-term spectral envelope is made to be the same throughout, and so precludes any spectral transitions.

Reasons for the ineffectiveness of noise contexts are investigated in experiment 1, which tests the following possibilities:

(1) It may be that speech-like spectral transitions in a context sound are necessary for there to be a compensation effect on a neighboring speech sound.
(2) Compensation might only affect test words that are sequentially grouped (Bregman, 1990) with the context, and test words might not group with a steady spectrum context because it has a "noise-like" character that is unlike speech. In this case, the context might be heard as a separate source, so that the compensation mechanism assesses the signal's reverberation from only the test word and does not include the context.
(3) The SCN operation might reduce information about reverberation from the sound's narrow-band offsets. This could happen because an effect of the SCN operation is to make the sound's narrow-band temporal envelopes to be all the same as the original broadband temporal envelope. Since the broadband temporal envelope of speech is generally smoother than its narrow-band temporal envelopes, the SCN operation on this signal will have a tendency to reduce the steepness of narrow-band offsets. When this happens, the presence (or absence) of tails from reverberation might be obscured. This effect is illustrated in Figs. 1 and 2, which plot the temporal envelopes of contexts used in the present experiments. These plots show that the envelope's downward-going slopes, where offsets occur, are less steep when there is more reverberation. However, these slopes are also less steep in the broadband envelopes than they are in the corresponding narrow-band envelopes.

To test these three possibilities, experiment 1 here uses both rapid and more slowly spoken test words together with different types of steady-spectrum noise contexts. These contexts have either the temporal envelope of the broadband speech signal, or else they have the temporal envelope of a narrow band of the speech that was obtained by playing it through an auditory-filter simulation.

Experiments 2 and 3 ask whether compensation for reverberation in speech is similar to effects of tails on loudness. Stecker and Hafter (2000) found that damped sounds with slowly decaying offsets (i.e., tails) were judged less loud than time-reversed "ramped" versions with rapid offsets (and slowly building "heads"). This effect can be explained by assuming that tails do not contribute as much to sounds' loudness because they resemble an effect of room reverberation. This loudness effect was just as prominent with narrow-band sounds as with broadband sounds. Experiments 2 and 3 therefore measure compensation effects with narrow-band contexts, which are generated with auditory-filter simulations to give bands centered at the auditory-filter's center frequency.

Experiment 3 also asks whether restrictions on temporal coding in the auditory nerve might influence compensation in speech. Such restrictions seem to affect the perceptual contrast between certain types of "ramped" (no tail) and "damped" (with tail) sounds when they are presented in a repeating sequence (Patterson, 1994a, 1994b). This contrast arises as the damped sounds appear less continuous than the

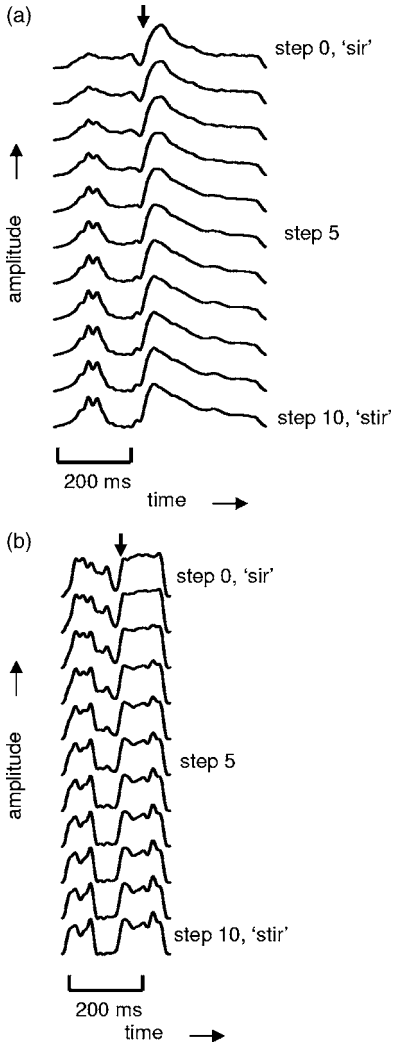


FIG. 3. Temporal envelopes of the sir to stir continua that were used as test words in the present experiments. These envelopes were obtained as described in Fig. 1, and are plotted here with amplitude on the ordinate. The traces run sequentially from step 0 at the top down to step 10 at the bottom. Arrows indicate voicing onset. (a) Slowly spoken test words. (b) Test words spoken more rapidly.

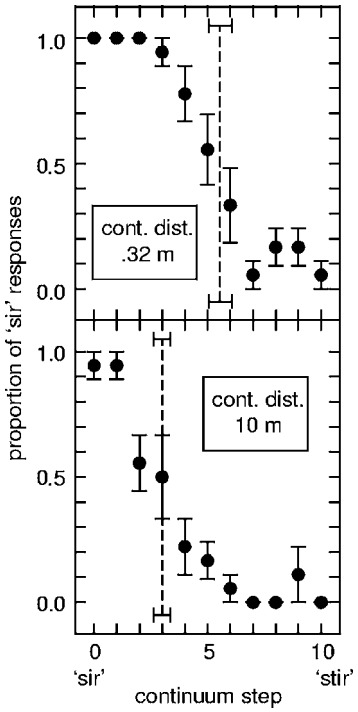


FIG. 4. Identification functions that show compensation for the distorting effects of adding reverberation to test words. Data points are means across six listeners and vertical bars are standard errors. Means of category boundaries are the vertical dashed lines, with their standard errors shown as horizontal bars. These examples are from conditions in experiment 2 where the context was wideband noise and the test words were spoken slowly. In both cases the test word's reverberation was from the 10-m distance. In the upper panel, the context's distance (cont. dist.) is 0.32 m, so it contains less reverberation than the test word. In the lower panel the context's distance is increased, to be the same as the test word's distance. This increase in the context's reverberation gives a compensation effect as more of the test words sound like stir, and so the category boundary moves leftwards, towards the sir end of the continuum.

ramped sounds, with a reduced prominence of damped sounds' tails. Lowering the center frequency of these sounds enhances the contrast, and this effect has been attributed to a more robust coding of time intervals in the auditory nerve at lower frequencies (Patterson, 1994a, 1994b). Experiment 3 therefore uses a lower center frequency for the auditory filter used to generate the narrow-band context, to determine whether compensation is similarly more prominent under these conditions.

II. EXPERIMENT 1

Test words were drawn from a continuum of steps between "sir" and "stir," which was formed using an amplitude modulation technique described by Watkins (2005b). In some conditions, listeners heard the rapidly spoken test words and contexts used in the earlier investigation of noise contexts (Watkins, 2005b), while in other conditions the more slowly spoken test words were heard. The temporal envelopes of these sounds are shown in Fig. 3.

The test words were used to measure distorting effects of adding reflected sound, as well as to measure any perceptual compensation for this distortion. Accordingly, the category boundary was measured, which is the continuum step where listeners switch from sir-to stir identifications. If distortion from reflections is sufficient to oppose the effects of the amplitude modulation that formed the continuum, by obscuring the gap that indicates the presence of the [t], then more steps will be identified as sir. When identification changes in this way, there is a "category boundary shift" as the boundary-step moves away from the sir end of the continuum to a higher step number.

Test words were embedded in a context phrase. When the amount of reflected sound in the context is kept at a minimal level, reflections added to the test word typically give a category boundary shift. If this happens, compensation may be seen when the amount of reflected sound in the context is also increased, to a level close to the amount in the test word. The compensation moves the category boundary back down towards sir, to a step close to that found when test words have a minimal amount of reflected sound. Figure 4 plots identification functions and category boundaries from conditions that show this compensation effect.

The context phrase was processed to form two types of steady-spectrum sound in which the shapes of the short-term spectral envelopes are all the same as the long-term averaged spectrum of the original speech. The temporal envelopes of these two sounds were different. One sound had the "broadband" temporal envelope of the original speech, while the other had a "narrow-band" temporal envelope, which arises at the output of a single auditory filter excited by the same speech. The auditory filter used had a center frequency at 2 kHz, which is in a range that is likely to be salient in hearing the [s] vs [st] distinction in test words. The experiment was designed to ask whether contexts with a narrow-band temporal envelope are more effective in generating compensation than the contexts with broadband temporal envelopes.

Listening conditions were chosen with the aim of providing substantial compensation effects. Accordingly, monaural presentation was used, as here effects of adding reverberation to test words are larger than in dichotic conditions (Watkins, 2005b). A further aim was to try and make category boundary shifts roughly comparable between conditions with test words spoken at the two rates, so that the amount of compensation could be compared. To this end, reverberation from a corridor was used in conditions with rapidly spoken test words, while reverberation from a larger, L-shaped room was used for the more slowly spoken test words (see Watkins, 2005b, experiment 1).

A. Method

1. Stimuli

The methods described by Watkins (2005b) were used to obtain recordings of context phrases containing sir and stir tokens. The recordings had a 16-bit resolution with a 48-kHz sampling rate, and were obtained in an IAC 1201 booth using a Sennheiser MKH 40 P48 cardioid microphone. The phrase "OK, next you'll get_to click on" was used for the more rapidly spoken test words that lasted 294 ms each. This duration was achieved by speaking so that only the "OK" was stressed, with no particular emphasis on the test word. The slower, 577-ms test words were spoken as the stressed word in the context "Next you'll get_to click on." The overall speaking rates of these phrases were a moderate 3 to 4 syllables per second, so that the large changes in test-word rates were mostly effected by the differences in stress. Clear stir sounds were obtained by amplitude modulation of the recordings of sir, using a modulation function that interchanged the temporal envelopes of the two words' waveforms. Interpolations between these end-point envelopes gave intervening steps to form the 11-step continua of test sounds. Dry test words from these continua were heard as sir at the lower step numbers, with a switch to stir at around step 3 or 4.

Test words were re-embedded in the context parts of the original utterance in a way that allowed different amounts of reverberation to be separately introduced into the two sounds. This was done by adding the context's waveform to the test word's waveform after silent sections had been added to each waveform to preserve the original temporal alignment.

TABLE I. The amount of reflected sound at the 10-m distances, as indicated by the time taken for the room's impulse-response energy to decay from its initial 0 to −10 dB. Values of this decay time at 0.32 m were all less than 0.01 s. These values are similar to early decay time, but see the text for caveats.

	Octave band		
Room	f_C=0.25 kHz	f_C=2 kHz	A-weighted
Corridor	0.15 s	0.12 s	0.12 s
L-shape	0.16 s	0.14 s	0.14 s

The methods described by Watkins (2005b) were also used to introduce reverberation into the dry contexts and test words by convolution with the left channel of a binaural room impulse response (BRIR). This gives the effect of monaural real-room listening over headphones. The impulse responses were obtained in rooms with dummy-head transducers (a speaker in a Bruel & Kjaer 4128 head and torso simulator, and Bruel & Kjaer 4134 microphones in the ears of a KEMAR mannequin), so that they incorporate the directional characteristics of a human talker and a human listener. To obtain signals at the listener's eardrum that match the signal at KEMAR's ear, the frequency-response characteristics of the dummy head talker and of the listener's headphone were removed using appropriate inverse filters.

BRIRs were obtained in two rooms of a disused office building. One room was a corridor that was 2 m wide, and about 35 m long, with a ceiling height of approximately 3.4 m. The other, larger room was L-shaped, comprising a 6×7 m rectangle and a 4×3 m rectangle with a ceiling height of about 3.4 m.

To obtain different amounts of reflected sound in a "natural" way, different distances between the dummy-head transducers of the talker and listener were used. The transducers faced each other, while the talker's position was varied to give distances from the listener of 0.32 or 10 m. The amounts of reflected sound at these distances, as indicated by measures of the time taken for the room's impulse-response energy to decay from its initial 0 to −10 dB, are shown in Table I. The slopes of the later parts of the energy decay are shown in Table II. These two sorts of measurement are similar to Early Decay Time (EDT) and Reverberation Time (T),

TABLE II. The slopes of the room impulse responses' energy decay. The table shows the time taken for a fall of 60 dB, extrapolated from measurements between −20 and −30 dB. These values are similar to reverberation time, but see the text for caveats.

		Octave band		
Room	Distance	f_C=0.25 kHz	f_C=2 kHz	A-weighted
Corridor	0.32 m	0.39 s	0.48 s	0.55 s
Corridor	10 m	1.14 s	1.08 s	0.96 s
L-shape	0.32 m	0.32 s	0.43 s	0.46 s
L-shape	10 m	0.99 s	0.99 s	0.90 s

respectively, in ISO 3382 (1997). However, these measurements do not comply with the standard's recommendations for omnidirectional transducers and spatial averaging, because the present purpose is to capture features that are present for listeners at different source-to-receiver distances.

To obtain steady-spectrum noise contexts whose broadband temporal envelopes were the same as the original speech, the speech contexts were processed with an "SCN operation." This involves reversing the polarity of a randomly selected half of the signal's samples to give signal-correlated noise (Schroeder, 1968). A filter was then applied to give these sounds the same long-term average spectrum as the original speech (Watkins and Makin, 1996).

To obtain steady-spectrum contexts that have the narrow-band temporal envelope of the original speech, the signals were first played through an auditory-filter simulation [gammatone, $n=4$, $f_C=2$ kHz, with a bandwidth equal to the "Cambridge ERB" (Hartmann, 1998)], followed by the SCN operation. This gives a sound with the same temporal envelope as the signal emerging from a single auditory filter when the context's speech is processed by the auditory periphery. A second stage of processing was then performed on this signal so that its temporal envelope would occur in the auditory filter of a listener. This processing compensates for the fact that the ear of the listener filters the sound again. Accordingly, the signal was played through a version of the gammatone filter whose impulse response was reversed in time. Finally, a second SCN operation gave a wideband signal, which was filtered to give the sound the same long-term average spectrum as the original speech signal. Temporal envelopes of these contexts are shown in Figs. 1 and 2.

2. Procedure

Category boundaries for the sir to stir continua were measured by asking listeners to identify test words from each continuum step. On each trial, listeners heard a context with an embedded test word. Listeners then identified the test word with a click of the computer's mouse, which they positioned while looking through the booth's window at the sir and stir alternatives displayed on the computer's screen. The computer waited for the listener to respond before presenting the following trial. Each of the 11 steps of a continuum was presented 3 times in a randomized sequence. The step corresponding to the category boundary was then found from the total number of sir responses to all of a continuum's steps, by dividing this total by 3 before subtracting 0.5, giving a boundary step number between −0.5 and 10.5.

Sounds were delivered to listeners at a peak level of 48 dB SPL through the left earpiece of Sennheiser HD480 headphones in the otherwise quiet conditions of an IAC 1201 booth. The experiment was administered to listeners in individual sessions by an Athlon 3500 PC computer with MATLAB 7.1 software and with an M-Audio Firewire 4100 sound card. Before the experimental trials, listeners were informally administered with a few randomly selected practice trials to familiarize them with the setup and attributes of the sounds such as their rate and their general variation across the different types of context.

3. Design

One group of six listeners identified rapidly spoken test words, while another group of six listeners identified slowly spoken test words. The test words were presented for identification in the speech context that was appropriate for their rate, as well as in the steady-spectrum versions of this context. BRIRs for the 0.32- and 10-m distances were applied to the context and to the test words. Over a sequence of trials, listeners heard all combinations of the context and test-word distances. This gives 2 test-word distances $\times$ 2 context distances $\times$ 3 types of context $\times$ 11 continuum steps $\times$ 3 repeats = 396 trials for a listener. Each listener received their trials in a different randomized order.

Listeners were Reading University Psychology undergraduates, who were native or fluent non-native English speakers and who did not report any hearing problems. They had had previous experience participating in category boundary experiments in their practical classes, but had not heard the stimuli used in this study before.

B. Results

Figure 5 shows means of category boundaries across the two groups of six listeners.

A four-way analysis of variance was performed on the category boundaries. There was a two-level between-subject factor, test-word rate, to indicate whether listeners heard rapid or more slowly spoken test words. Two within-subject factors were used to indicate the amount of reverberation in BRIRs by the distance at which they were recorded, giving two-level factors for context distance and test-word distance. The other within-subject factor, with three levels, was used to indicate the different types of context.

There were substantial compensation effects, as shown by the reductions in category boundaries when the context's distance was increased. These compensation effects were mainly confined to test words at the 10-m distance, and were more prominent with speech contexts than with the steady-spectrum noise contexts. This pattern of results shows in the analysis as a three-way interaction among the factors: context's distance, test word's distance, and context type, where $F(2,20)=9.30$, Huynh-Feldt epsilon $=0.8994$, and $p=0.0021$. Nevertheless, compensation effects were present to an extent with the steady-spectrum contexts, and were more pronounced when test words were slowly spoken, especially with the broadband contexts. This pattern shows in the analysis as a three-way interaction among: context's distance, context type, and test word's rate, where $F(2,20)=9.25$, Huynh-Feldt epsilon >1, and $p=0.0014$.

Effects of increasing the distance of the test word's BRIR were generally more substantial with speech contexts than they were with the steady-spectrum contexts, as indicated by the two-way interaction between context type and test word's distance with $F(2,20)=22.53$, Huynh-Feldt epsilon >1, and $p<0.0001$. There was also a two-way interaction between test-word's distance and test word's rate with $F(1,10)=5.16$, and $p=0.0464$, indicating that effects of increasing the distance of the test word were more prominent overall with the fast test words. However, compensation was

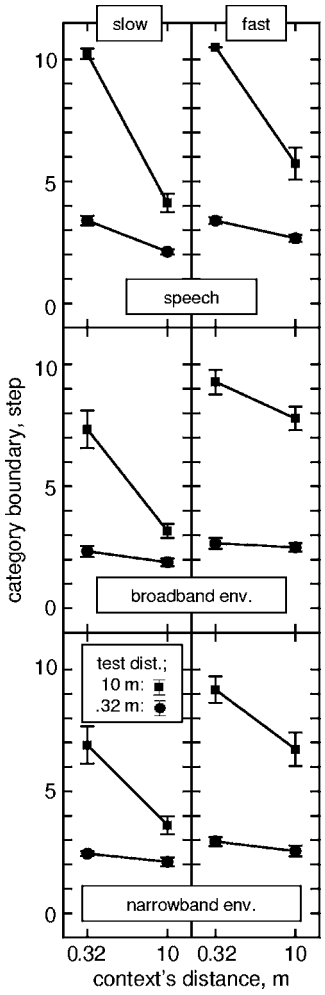


FIG. 5. Results of experiment 1, showing means and standard errors of category boundaries from the two groups of six listeners who identified test words with contexts that were speech (top panels) or steady-spectrum noise. The temporal envelopes of the noise contexts were from either the broadband speech signal (broadband env., middle panels) or from the speech played through an auditory filter simulation centered at 2 kHz (narrow-band env., bottom panels). The different groups heard test words that were spoken slowly (slow, left panels) or more rapidly (fast, right panels). Slow test words and their contexts were convolved with BRIRs from the L-shaped room while the fast test words were convolved with BRIRs from the corridor. The parameter in all the panels is the test word's distance (test dist.) shown in the bottom left panel.

generally more prominent with the more slowly spoken test words, as indicated by the two-way interaction between the context's distance and the test word's rate, where $F(1,10)=17.87$, and $p=0.0018$.

On average, compensation effects were more substantial with the speech contexts, as indicated by the two-way interaction between context type and context's distance where $F(2,20)=3.52$, Huynh-Feldt epsilon >1, and $p=0.049$. In conditions with steady-spectrum contexts, the more rapidly spoken test words were more influenced by an increase in the test word's distance, but at the same time there was less compensation than there was with the slow test words, especially for the broadband contexts. This combination of effects led to generally higher category boundaries for the fast test words in steady-spectrum contexts, which shows as an interaction between context type and test-word's rate with $F(2,20)=8.42$, Huynh-Feldt epsilon >1, and $p=0.0022$. There was an interaction between the context's distance and the test word's distance, with $F(1,10)=68.98$, and $p<0.0001$, which came about as effects of compensation were mostly confined to the test words with more reverberation, at the 10-m distance.

Main effects in the analysis arose largely through the nature of the interactions described above. Increasing the test word's distance raised category boundaries on average, which shows as a main effect of test word's distance with $F(1,10)=107.54$, and $p<0.0001$. The compensation tended generally to reduce category boundaries, as indicated by the main effect of context's distance with $F(1,10)=394.25$, and $p<0.0001$. Category boundaries were on average higher for fast test words, where the effects of increasing the test word distance were generally larger, while compensation effects were smaller than in conditions with the slow test words. This general pattern contributes to a main effect of test-word rate where $F(1,10)=26.83$, and $p=0.0004$. Overall, category boundaries were higher with speech contexts, giving a main effect of context type, with $F(2,20)=11.83$, Huynh-Feldt epsilon >1, and $p=0.0004$. This last effect seems mostly due to the relatively large effects of test-word distance with speech contexts. There were no other significant F-ratios.

These results replicate earlier findings with speech contexts (Watkins, 2005a, 2005b, 2005c) and show that steady-spectrum contexts can be sufficient for compensation, particularly with the more slowly spoken test words and their contexts. With these slowly spoken test words, steady-spectrum contexts reduced compensation a little, but the bulk of the effect remained with both types of temporal envelope. Compensation with the more rapidly spoken test words was found to be most apparent with speech contexts. However, with these fast test words, compensation was found with steady-spectrum contexts to some extent, and was more prominent with a narrow-band temporal envelope from an auditory filter centered at 2 kHz.

III. EXPERIMENT 2

Watkins (2005b) found that there were certain similarities between speech compensation effects and effects of sounds' tails in the loudness experiments of Stecker and Hafter (2000). These loudness experiments used noise bands or tones and found that the loudness of sounds with slowly decaying offsets, or "tails," was lower than the loudness of sounds with rapid offsets. This lowering of loudness occurred mainly when the preceding trial's sound had a tail, so this could be interpreted as an effect of "context" similar to that found in the speech compensation experiments of Watkins (2005b). The loudness effects were as prominent with the narrow-band tones as they were with broadband noise sounds. Here, experiment 2 asks whether narrow-band contexts are also effective in speech compensation, using narrow-band versions of the contexts used in experiment 1.

A. Method

Broadband contexts were obtained by the same processing used to obtain the steady-spectrum "narrow-band enve-

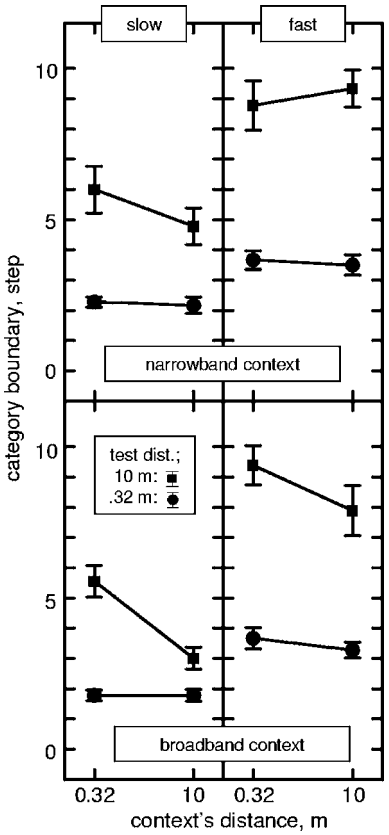


FIG. 6. Means and standard errors of category boundaries from experiment 2 where the two groups of six listeners identified test words with contexts that were steady-spectrum noise. The context had a broad bandwidth (lower panels) or a narrow bandwidth, centered at 2 kHz (upper panels). For both types of context, the temporal envelope was from the speech context played through an auditory filter centered at 2 kHz. Other details are the same as Fig. 5.

lope" contexts in experiment 1, using an auditory filter centered at 2 kHz as before. Narrow-band contexts centered at this frequency were obtained with the same processing, but with the final SCN operation omitted. As a result, the temporal envelopes of the broadband and narrow-band contexts were the same. Other aspects of the method were the same as for experiment 1.

There were two groups of six listeners who had not participated in experiment 1. These listeners identified either rapid or slowly spoken test words both in the narrow-band and broadband types of steady-spectrum contexts. BRIRs for the 0.32- and 10-m distances were applied to the context and to the test words. Over a sequence of trials, listeners heard all combinations of the context and test-word distances. This gives 2 test-word distances $\times$ 2 context distances $\times$ 2 types of context $\times$ 11 continuum steps $\times$ 3 repeats = 264 trials for a listener, and each listener received the trials in a different randomized order.

B. Results

Figure 6 shows means of category boundaries across the two groups of six listeners.

A four-way analysis of variance was performed on the category boundaries. The analysis used a two-level between-subject factor for test-word rate, along with two within-subject factors for context distance and test-word distance. These three factors were all the same as in experiment 1. One other within-subject factor, context bandwidth, had two levels for the narrow-band and broadband contexts.

Compensation effects were again apparent in this experiment, as there were reductions in category boundaries as the context's distance was increased. The compensation was largely confined to conditions with broadband contexts, and it occurred mostly with the test words at the 10-m distance. This pattern shows in the analysis as a three-way interaction among the factors: context's distance, context bandwidth, and test word's distance, with $F(1,10)=5.08$ and $p=0.0478$. The compensation effects were again more prominent with the slow test words, and at the 10-m test-word distance. This pattern is indicated by the three-way interaction among: context's distance, test-word rate, and test-word distance, where $F(1,10)=5.96$, and $p=0.0348$.

There was a two-way interaction between test word's distance and test word's rate with $F(1,10)=6.79$, and $p=0.0262$, indicating that effects of increasing the distance of the test word were more prominent overall with the fast test words. Conversely, compensation was generally more prominent with the more slowly spoken test words, as indicated by the two-way interaction between the context's distance and the test word's rate, where $F(1,10)=8.12$, and $p=0.0173$.

The overall effect of increasing the test word's distance was lower with broadband contexts, giving a two-way interaction between bandwidth and test-word distance, with $F(1,10)=7.25$, and $p=0.0226$. This effect seems partly due to there being little or no compensation in narrow-band conditions, along with the reductions in category boundaries brought about by the compensation effect in broadband conditions. There was also a two-way interaction between the context's distance and the test word's distance, with $F(1,10)=68.98$, and $p<0.0001$, indicating that when compensation occurred, it was generally confined to the 10-m test words.

Main effects in the analysis seemed largely to come about through the nature of the interactions described above. Increasing the test word's distance raised category boundaries on average, which shows as a main effect of test-word distance with $F(1,10)=34.57$, and $p=0.0002$. The compensation tended overall to reduce category boundaries, as indicated by the main effect of context's distance with $F(1,10)=87.31$, and $p<0.0001$. On average, category boundaries were higher with narrow-band contexts, giving an overall main effect of context type, with $F(1,10)=7.94$, and $p=0.0182$. This effect seems mainly due to there being little or no reduction in category boundaries from compensation in narrow-band conditions. Category boundaries were higher overall for fast test words, giving a main effect of test-word rate, where $F(1,10)=48.93$, and $p<0.0001$. This last effect seems mainly to arise as increasing the distance of fast test words led to larger increases in their category boundaries, while reductions in category boundaries through compensation were generally smaller than in conditions with the slow test words. In addition though, at 0.32 m, the category

boundaries tended to be somewhat higher for fast test words than for slow test words. There were no other significant F-ratios.

These results replicate the findings of experiment 1 with broadband steady-spectrum contexts, and they show that compensation is substantially reduced with narrow-band versions of these contexts.

IV. EXPERIMENT 3

There are some perceptual effects of sounds that have tails which seem to depend on the spectral location of the sound's frequency components. One is the perceptual difference between sounds with ramped and damped temporal envelopes presented in a repeating sequence (Patterson 1994a, 1994b). In Patterson's experiments, effects were confined to lower frequencies, and this was attributed to restrictions on the neural coding of temporal information in the "phase-locked" response of the auditory nerve's fibers. On the other hand, effects of sounds' tails that affect loudness judgments do not seem to depend on this sort of temporal coding, as they occur across a broad range of frequencies (Stecker and Hafter, 2000). Experiment 2 found that there was little compensation with narrow-band contexts centered at 2 kHz, and this finding might conceivably be due to some restriction on temporal coding. Therefore, experiment 3 investigates effects at a lower center frequency, as the phase-locked response of the auditory nerve's fibers becomes more pronounced towards lower frequencies.

A. Method

Broadband and narrow-band contexts were obtained with the same processing that was used to obtain the steady-spectrum contexts of experiment 2, but in experiment 3 the auditory filter involved was centered at 250 Hz. Temporal envelopes from the resulting contexts are shown in Fig. 7.

There were two groups of six listeners who had not participated in experiment 1 or 2. These listeners identified either rapid or slowly spoken test words both in the narrow-band and broadband types of steady-spectrum contexts. BRIRs for the 0.32- and 10-m distances were applied to the context and to the test words. Over a sequence of trials, listeners heard all combinations of the context and test-word distances. This gives 2 test-word distances $\times$ 2 context distances $\times$ 2 types of context $\times$ 11 continuum steps $\times$ 3 repeats = 264 trials for a listener, and each listener received the trials in a different randomized order.

B. Results

Figure 8 shows means of category boundaries for the two groups of six listeners.

A four-way analysis of variance was performed on the category boundaries, using the same factors as in experiment 2.

Compensation effects like those in experiments 1 and 2 were found, with reductions in category boundaries for the 10-m test words when the context's distance was increased. As in experiment 2, this pattern is largely confined to conditions with broadband contexts. These aspects of the data show in the analysis as a three-way interaction among the factors: context's distance, test word's distance, and context bandwidth, with $F(1,10)=25.81$, and $p=0.0005$. In one respect results differ from experiment 2, as in experiment 3 the compensation effect in broadband-context conditions with the fast test words seems to be substantially similar to the compensation in these conditions with slow test words.

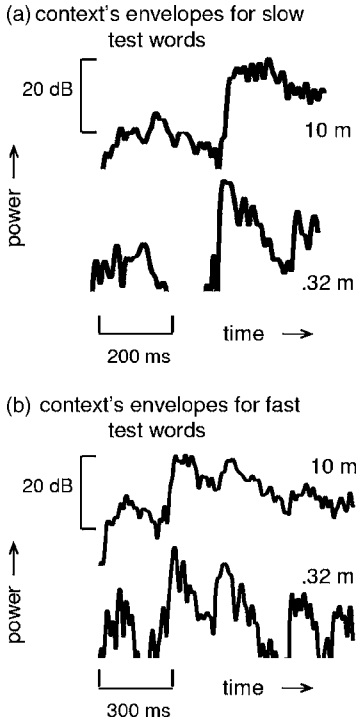


FIG. 7. Temporal envelopes obtained from contexts played through an auditory filter centered at 250 Hz. (a) "Next you'll get…" used with the slowly spoken test words. (b) "OK, next you'll get…" used with the more rapidly spoken test words. Other details are the same as Fig. 3.

Category boundaries were lower on average with broadband contexts, giving a two-way interaction between bandwidth and test-word distance, with $F(1,10)=33.84$, and $p=0.0002$. This effect seems partly due to there being little or no compensation in narrow-band conditions, coupled with the reductions in category boundaries through the compensation effect in broadband conditions. There was also a two-way interaction between the context's distance and the test word's distance, with $F(1,10)=17.27$, and $p=0.0020$, indicating that when compensation occurred, it was most prominent with the 10-m test words. On average, effects of the context's distance were greater with the more slowly spoken test words, as indicated by the two-way interaction between the context's distance and the test word's rate, where $F(1,10)=8.08$, and $p=0.0175$. This effect seems to arise here mainly through combining with narrow-band conditions, rather than from variations in compensation at the two rates.

Main effects in the analysis seemed largely to come about through the nature of the interactions described above. Increasing the test word's distance raised category boundaries on average, which shows as a main effect of test word's distance, with $F(1,10)=39.78$, and $p<0.0001$. The compensation tended overall to reduce category boundaries, as indi-

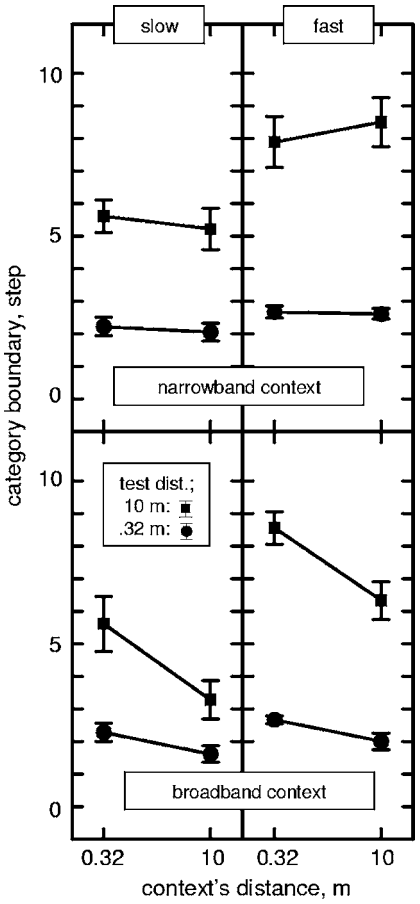


FIG. 8. Means and standard errors of category boundaries from experiment 3 where the two groups of six listeners identified test words in contexts with a broad bandwidth (lower panels) or a narrow bandwidth, centered at 250 Hz (upper panels). For both types of context the temporal envelope was from the speech context played through an auditory filter centered at 250 Hz. Other details are the same as in Fig. 6.

cated by the main effect of context's distance, with $F(1,10)=91.46$, and $p<0.0001$. On average, category boundaries were higher with narrow-band contexts, giving an overall main effect of context type, with $F(1,10)=7.94$, and $p=0.0182$, which seems mainly due to there being little or no reduction in category boundaries from compensation in narrow-band conditions. Category boundaries were generally higher with fast test words, giving a main effect of test-word rate, where $F(1,10)=15.18$, and $p=0.0030$. This last effect seems primarily due to the larger effects of test-word distance with fast test words. There were no other significant F-ratios.

These results again show that narrow-band versions of steady-spectrum contexts effect little or no compensation, even though there is substantial compensation from their broadband counterparts that have the same temporal envelope. This finding does not appear to be due to restrictions on temporal coding in the auditory nerve since the results found here, using narrow-band contexts at a low center frequency (250 Hz), are similar to results with narrow-band contexts in experiment 2 where a higher center frequency (2 kHz) was used.

V. DISCUSSION

The main finding of these experiments is that the perceptual compensation mechanism can pick up information about reverberation from different types of sounds that form a context for test words. Contexts that are effective in this respect include not only running speech, but also broadband noise-like sounds that have a steady spectrum with no changes in the spectral envelope's shape over time. In both cases, the effects of reverberation on a neighboring test word are reduced when the context's reverberation is increased to match the test word's reverberation. Therefore, offsets at the spectral transitions inherent in speech signals are not always necessary for compensation, as spectral transitions are absent from the noise contexts used in the present experiments.

These results with noise contexts indicate that a failure of perceptual grouping (Bregman, 1990) is unlikely to have been the cause of reductions in compensation in earlier experiments with similar noise contexts. Nevertheless, it may still be the case that compensation only affects test words that are sequentially grouped with the context, as it remains plausible that a noise-like context is grouped with the speech test word in these compensation experiments.

The present experiments suggest that it is the presence of sharp offsets in a context sound that allows it to inform compensation. Sharp offsets do occur in narrow frequency bands at spectral transitions, but they can also be present in a sound's broadband temporal envelope. These two factors are separated in the present experiments, where the noise contexts contained no spectral transitions while their temporal envelopes were varied. In this way, experiment 1 showed that the broadband temporal envelope of speech can contain offsets that are sufficient for compensation, and this result is clearest with the more slowly spoken test words.

In other conditions of experiment 1, with the rapidly spoken test words, compensation increased when the temporal envelope from a narrow frequency band of speech was used, compared with the smaller amount of compensation seen with the broadband temporal envelope. These results confirm that temporal envelopes of speech in the ear's numerous frequency channels can give more information about a room's acoustic properties than is apparent from the broadband temporal envelope.

These results with rapidly spoken test words in the broadband and narrow-band conditions of experiment 1 are also consistent with the ideas about effects of the SCN operation that were raised in the Introduction. When the SCN operation is used to generate a noise with the broadband temporal envelope of a speech signal, the effect appears in some ways to be similar to that of adding reverberation. Both of these operations will have a tendency to obscure offsets in narrow frequency bands, as they both reduce the steepness of downward-going slopes in the temporal envelope. The effect on temporal envelopes can be seen by comparing the corresponding traces in Figs. 1 and 2, and the effect on listeners is that there is less compensation when using the "smoother" broadband temporal envelope.

The speech contexts were more effective than steady-spectrum contexts in experiment 1, giving larger compensa-

tion effects with both the slowly- and rapidly spoken test words. It seems likely that this reflects the wide range of frequency channels available when listening to speech. The temporal envelopes in some of these channels might well be more informative about the presence of reverberation than the single channels that were selected for study in the present experiments.

Experiment 2 failed to demonstrate compensation with a narrow-band context. This result does not appear to be due to restrictions on temporal coding in the auditory nerve (Patterson, 1994a, 1994b), as the narrow-band contexts with a lower center frequency in experiment 3 were similarly ineffective.

The effects of tails on loudness reported by Stecker and Hafter (2000) occur with narrow-band sounds, unlike the speech compensation measured in experiments 2 and 3 where narrow-band contexts are ineffective. However, an important difference between these experiments is that the test sound in the loudness experiment is in the same narrow frequency-band as its "context" (i.e., the preceding trial), whereas a speech test word contains a wide range of frequencies. So, it is possible that a similar compensation is occurring in the two types of experiment if it is one that is confined to the peripheral frequency channels that the context occupies. This sort of mechanism would leave much of the test word's frequency content unaffected by the narrow-band contexts in experiments 2 and 3, and it seems likely that this is why there is little or no compensation in these conditions.

Although the present experiments have investigated the temporal-envelope distinction between only one pair of consonants, it seems likely that there are more diverse influences of the reverberation and compensation effects that are reported here. This is because information from the temporal envelope is pervasive in speech (Plomp, 1983; Rosen, 1992) to the extent that listeners can understand spoken messages on the basis of the temporal envelopes in only a few frequency bands (van Tasell *et al.*, 1992; Shannon *et al.*, 1995). Furthermore, the present experiments indicate that information about reverberation can be picked up from nonspeech sounds as well as from speech, and so suggest that compensation for reverberation may not be confined to speech.

ACKNOWLEDGMENT

This work was supported by a grant to the first author from EPSRC.

Allen, J. B., and Berkley, D. A. (**1979**). "Image method for efficiently simulating small-room acoustics," J. Acoust. Soc. Am. **62**, 943–950.

Bolt, R. H., and Macdonald, A. D. (**1949**). "Theory of speech masking by reverberation," J. Acoust. Soc. Am. **21**, 577–580.

Bregman, A. S. (**1990**). *Auditory Scene Analysis* (MIT, Cambridge).

Furui, S. (**1986**). "On the role of spectral transition for speech perception," J. Acoust. Soc. Am. **80**, 1016–1025.

Haas, H. (**1951**). "Über den einfluss des einfachechoes auf die hörsamkeit von sprache," (On the influence of a single echo on the intelligibility of speech) Acustica **1**, 49–58.

Hartmann, W. M. (**1998**). *Signals, Sound and Sensation* (Springer, New York).

ISO 3382 (**1997**). *Acoustics-Measurement of the reverberation time of rooms with reference to other acoustical parameters* (International Organization for Standardization, Geneva).

Nábělek, A. K., Letowski, T. R., and Tucker, F. M. (**1989**). "Reverberant overlap- and self-masking in consonant identification," J. Acoust. Soc. Am. **86**, 1259–1265.

Patterson, R. D. (**1994a**). "The sound of a sinusoid: Spectral models," J. Acoust. Soc. Am. **96**, 1409–1418.

Patterson, R. D. (**1994b**). "The sound of a sinusoid: Time-interval models," J. Acoust. Soc. Am. **96**, 1419–1428.

Plomp, R. (**1983**). "The role of modulation in hearing," in *Hearing-Physiological Bases and Psychophysics*, edited by R. Klinke and R. Hartmann (Springer, New York), pp. 270–276.

Rosen, S. (**1992**). "Temporal information in speech: Acoustic, auditory and linguistic aspects," Philos. Trans. R. Soc. London, Ser. B **336**, 367–373.

Schroeder, M. R. (**1965**). "New method of measuring reverberation time," J. Acoust. Soc. Am. **37**, 409–412.

Schroeder, M. R. (**1968**). "Reference signal for signal quality studies," J. Acoust. Soc. Am. **44**, 1735–1736.

Shannon, R. V., Zeng, F., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). "Speech recognition with primarily temporal cues," Science **270**, 303–304.

Stecker, G. C., and Hafter, E. R. (**2000**). "An effect of temporal asymmetry on loudness," J. Acoust. Soc. Am. **107**, 3358–3368.

Summerfield, Q. (**1981**). "Articulatory rate and perceptual constancy in phonetic perception," J. Exp. Psychol. Hum. Percept. Perform. **7**, 1074–1095.

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 evaluation of cochlear implants," J. Acoust. Soc. Am. **92**, 1247–1257.

Watkins, A. J. (**1992**). "Perceptual compensation for effects of reverberation on amplitude-envelope cues to the 'slay'-'splay' distinction," in Proceedings of the Institute of Acoustics **14**, 125–132.

Watkins, A. J. (**2005a**). "Listening in real-room reverberation: effects of extrinsic context," in *Auditory Signal Processing: Physiology, Psychoacoustics, and Models*, edited by D. Pressnitzer, A. de Cheveigné, S. McAdams, and L. Collet (Springer, New York, pp. 423–428).

Watkins, A. J. (**2005b**). "Perceptual compensation for effects of reverberation in speech identification," J. Acoust. Soc. Am. **118**, 249–262.

Watkins, A. J. (**2005c**). "Perceptual compensation for effects of echo and of reverberation in speech identification," Acta. Acust. Acust. **91**, 892–901.

Watkins, A. J., and Holt, N. J. (**2000**). "Effects of a complex reflection on vowel identification," Acust. Acta Acust. 86, 532–542.

Watkins, A. J., and Makin, S. J. (**1996**). "Some effects of filtered contexts on the perception of vowels and fricatives," J. Acoust. Soc. Am. **99**, 588–594.

Prediction of transmission loss of double panels with a patch-mobility method

Jean-Daniel Chazot[a] and Jean-Louis Guyader
Laboratoire Vibrations Acoustique de l'INSA de Lyon, 25 bis avenue Jean Capelle, 69621 Villeurbanne Cedex, France

(Received 21 April 2006; revised 20 October 2006; accepted 20 October 2006)

Sound transmission loss through double panels is studied with a patch-mobility approach. An overview of the method is given with details on acoustic and structural patch mobilities. Plate excitation is characterized by blocked patch pressures that take into account room geometry and source location. Hence, panel patch velocities before coupling can be determined and used as excitation in the mobility model. Then a convergence criterion of the model is given. Finally, transmission loss predicted with a patch-mobility method is compared with measurements.
 [DOI: 10.1121/1.2395920]

PACS number(s): 43.55.Rg [NX] Pages: 267–278

I. INTRODUCTION

The first studies on sound transmission were done using unbounded plates excited by plane waves. Cremer's[1] and London's works on single panels[2] and on double panels[3] enabled us to underline several important aspects: the critical frequency, the double panel resonance, and the existence of an incidence angle where sound was greatly transmitted. Modal behavior of plates was introduced by Vogel.[4] Later, coupling between a panel and a cavity has been studied[5–7] and sound transmission through nonresonant modes below critical frequency has been shown. Finally, several expressions of transmission loss through single and double panels were formulated by different authors like Beranek[8] or Fahy.[9] Their models, summed up and compared with experiment in Ref. 10, are based on the mass law with additional terms enabling to take into account some particular effects like critical frequency effect or reverberant field excitation. However, for more complicated systems, other tools are necessary like finite element models. A review of numerical solutions for low frequency structural acoustic problems and their limitations was done by Atalla and Bernhard.[11] Expansion of solution on a functional base to describe structural acoustic problems was presented in Refs. 12 and 13. At high frequency statistical energy analysis was used (Refs. 14 and 15). The different methods were then applied to design simple or double panels with optimized sound insulation properties on a given frequency band. Kropp and Rebillard[16] have for example studied the case of double panels with a transfer matrix method that is not very time consuming and give accurate results compared to the experiment. It is adapted to a description of diffuse field by uncorrelated plane waves summation. Unfortunately, it is not adapted to evaluate the influence of excitation conditions such as source room geometry, source location, and panel location on transmission loss of a panel.

Transmission loss of more sophisticated panels were also studied. Guyader[17–19] studied, for example, the case of orthotropic multilayered plates, and Nilsson[20] studied the case of sandwich structures with honeycomb and foam cores. Finally, developments of numerical poroelastic models (Refs. 21 and 22) enabled to study sound transmission through finite multilayered systems containing poroelastic materials.[23]

Practically, two main difficulties appear when dealing with calculation of transmission loss of double panels. The first is related to the excitation by diffuse field. The standard approach consist in decomposing the sound field in uncorrelated plane waves as in diffuse theory and cumulating the transmission loss calculated for each angle of incidence. This method has an extremely high computation cost that can be reduced drastically using the approach proposed by Spronck[24] and Guyader[18] using radiation impedance of plate modes in a reciprocal way. Another possibility of avoiding angle summation was proposed by Coyette,[25] based on a coherence length in diffuse field. A second difficulty with waves summation is the approximation of reality by diffuse field that can be bad especially at low frequency. A better comparison with the experiment was obtained by limiting the angle of incidence in order to suppress grazing waves excitation. In addition to the difficulty of fixing a limit angle, the numerical cost of cumulating results for different incident angles remains because the approach in Refs. 18 and 24 cannot be used in this case. The method presented in this paper avoids that difficulty, calculating the cavity boundary pressure field over the blocked panel, transmission loss is calculated without angle summation, and takes into account source room characteristics. A second difficulty associated with classical methods is the global modelization of the problem that necessitates a complete calculation when one element is modified. The use of a mobility technique permits us to characterize each component of the vibroacoustic problem separately, then to calculate the transmission loss solving the interaction equation. If one element is modified, only its own characterization has to be calculated before solving interaction equations. The mobility technique introduced in Refs. 26 and 27 is now widely used in structural coupling,[28–30] and an extension to energy mobility was presented by Orefice *et al.*[31] The mobility approach for vibroa-

[a]Electronic mail: jdchazot@lva.insa-lyon.fr

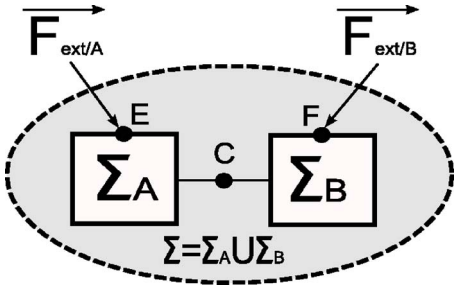


FIG. 1. Simple case of one coupling point.

coustic problems used here appears as an integral equation discretized by collocation leading to patch mobilities. This method was used previously by Ouisse *et al.*[32] for coupling acoustic subdomains.

II. PATCH-MOBILITY METHOD

A. Simple case of one coupling point

The mobility method enables us to calculate the response of vibrating coupled linear systems thanks to the knowledge of their isolated behaviors. The following example describes the principles of the mobility method. Let us consider a linear system Σ constituted of two subsystems Σ_A and Σ_B coupled at a point C (Fig. 1).

In all the following calculations, harmonic motions of angular frequency ω are assumed, and for sake of simplicity, time dependence $e^{j\omega t}$ is omitted. In a first step, transfer mobilities between two points M and Q are determined. This mobility being defined as the ratio of velocity at point M and force applied at point Q: $Y(M,Q)=V(M)/F(Q)$. From mobility $Y(M,Q)$, velocity at point M created by a force located at point Q can be calculated. However, excitation of a subsystem is produced by external forces, but also by coupling to other subsystems. This gives for subsystem A:

$$V_A(M) = \underbrace{Y_A(M,E)F_{\text{ext}\to A}(E)}_{\tilde{V}_A(M)} + \underbrace{Y_A(M,C)F_{B\to A}(C)}_{V_A^c(M)} \quad \forall M \in A. \tag{1}$$

Two velocities can be identified from Eq. (1): the velocity before coupling $\tilde{V}_A(M)$, and the velocity generated by the coupling $V_A^c(M)$. The second one is produced by the coupling force which is unknown, and appears as well in the subsystem B [notice that: $F_{B\to A}(C)=-F_{A\to B}(C)$]:

$$V_B(M) = Y_B(M,F)F_{\text{ext}\to B}(F) + Y_B(M,C)F_{A\to B}(C) \quad \forall M \in B. \tag{2}$$

To be able to find out the velocity at any point belonging to system A or B, the coupling force has to be determined. This is done with continuity condition at coupling point C: velocities of each subsystem have to be equal at the coupling point $[V_A(C)=V_B(C)]$. From these relations one can obtain the coupling force

$$F_{B\to A} = \frac{Y_B(C,F)F_{\text{ext}\to B}(F) - Y_A(C,E)F_{\text{ext}\to A}(E)}{[Y_A(C,C)+Y_B(C,C)]}. \tag{3}$$

Finally, the velocity of any point M can be calculated by using Eqs. (1) and (2).

B. Case of surface coupling with a patch-mobility method

When the coupling between subsystems is applied through a surface, Eqs. (1) and (2) must be adapted. External forces are considered surfacic as well. Tangential velocities and forces are not taken into account, that is realistic for acoustic subsystems and vibroacoustic coupling. Therefore the mobility $Y(M,Q)$ is only associated to transverse motion, and normal velocities and forces are defined as algebric values. Coupling velocities are then calculated by integrating the product of mobility and coupling (respectively, external) surfacic force over coupling (respectively, excited) surface

$$V_A(M) = \int_{S_A} [Y_A(M,E)F_{\text{ext}\to A}(E)]dS + \int_{S_c} [Y_A(M,C)F_{B\to A}(C)]dS \quad \forall M \in (A), \tag{4}$$

$$V_B(M) = \int_{S_B} [Y_B(M,F)F_{\text{ext}\to B}(F)]dS + \int_{S_c} [Y_B(M,C)F_{A\to B}(C)]dS \quad \forall M \in (B). \tag{5}$$

External forces of subsystem A (respectively, B) are applied on surface S_A (respectively, S_B), and coupling forces are applied on coupling surface S_c.

A collocation method can be used to solve the problem, separating surfaces such as $S_A=\cup_{i=1}^{N_A}S_{Ai}$ and $S_{Ai}\cap_{i\neq j}S_{Aj}=\emptyset$, $S_B=\cup_{i=1}^{N_B}S_{Bi}$ and $S_{Bi}\cap_{i\neq j}S_{Bj}=\emptyset$, $S_c=\cup_{i=1}^{N_c}S_{ci}$ and $S_{ci}\cap_{i\neq j}S_{cj}=\emptyset$. Let us first calculate averaged velocities on patch i:

$$\langle V_A\rangle_i = \int_{S_{Ai}} \sum_{j=1}^{N_A} \left\{ \int_{S_{Aj}} [Y_A(M,E)F_{\text{ext}\to A}(E)]dE \right\} dM + \int_{S_{Ai}} \sum_{j=1}^{N_c} \left\{ \int_{S_{cj}} [Y_A(M,C)F_{B\to A}(C)]dC \right\} dM \quad \forall M \in A, \tag{6}$$

$$\langle V_B\rangle_i = \int_{S_{Bi}} \sum_{j=1}^{N_B} \left\{ \int_{S_{Bj}} [Y_B(M,E)F_{\text{ext}\to B}(E)]dE \right\} dM + \int_{S_{Bi}} \sum_{j=1}^{N_c} \left\{ \int_{S_{cj}} [Y_B(M,C)F_{A\to B}(C)]dC \right\} dM \quad \forall M \in B. \tag{7}$$

Assuming in a second step that surfacic forces are constant on patches $[F(x,y)=\langle F\rangle_j$ if $(x,y)\in S_j]$, Eqs. (6) and (7) give

$$\langle V_A\rangle_i = \sum_{j=1}^{N_A} \langle\langle Y_A\rangle_i\rangle_j \langle F_{\text{ext}\to A}\rangle_j + \sum_{j=1}^{N_c} \langle\langle Y_A\rangle_i\rangle_j \langle F_{B\to A}\rangle_j, \tag{8}$$

$$\langle V_B\rangle_i = \sum_{j=1}^{N_B} \langle\langle Y_B\rangle_i\rangle_j \langle F_{\text{ext}\to B}\rangle_j + \sum_{j=1}^{N_c} \langle\langle Y_B\rangle_i\rangle_j \langle F_{A\to B}\rangle_j. \quad (9)$$

A new transfer mobility expression appears, it is the ratio of averaged patch i velocity to averaged patch j surfacic force

$$\langle\langle Y_A\rangle_i\rangle_j = \frac{\langle V\rangle_i}{\langle F\rangle_j}, \quad (10)$$

where

$$\langle V\rangle_i = \frac{\int_{S_i} V(x,y)dxdy}{S_i} \quad \text{and} \quad \langle F\rangle_j = \frac{\int_{S_j} F(x,y)dxdy}{S_j}. \quad (11)$$

Of course, assuming a constant surfacic force on each patch is an approximation, only valid if patch areas have small dimensions compared to wavelengths.

Then, velocity continuity over coupling patches gives N_c equations (one for each coupling patch)

$$\forall k \in [1,N_c] \quad \langle V_A\rangle_k = \langle V_B\rangle_k. \quad (12)$$

Hence, Eqs. (8) and (9) give $\forall k \in [1,N_c]$:

$$\sum_{j=1}^{N_c} \langle\langle Y_A\rangle_k\rangle_j \langle F_{B\to A}\rangle_j + \sum_{j=1}^{N_c} \langle\langle Y_B\rangle_k\rangle_j \langle F_{B\to A}\rangle_j = \sum_{j=1}^{N_B} \langle\langle Y_B\rangle_k\rangle_j \langle F_{\text{ext}\to B}\rangle_j - \sum_{j=1}^{N_A} \langle\langle Y_A\rangle_k\rangle_j \langle F_{\text{ext}\to A}\rangle_j. \quad (13)$$

The following matrices $[Y_A]$ and $[Y_B]$ are defined such as $Y_A(i,j)=\langle\langle Y_A\rangle_i\rangle_j$ and $Y_B(i,j)=\langle\langle Y_B\rangle_i\rangle_j$. Vectors $\{F_{B\to A}\}$, $\{F_{\text{ext}\to A}\}$, and $\{F_{\text{ext}\to B}\}$ are also defined such as $F_{B\to A}(j)=\langle F_{B\to A}\rangle_j$, $F_{\text{ext}\to A}(j)=\langle F_{\text{ext}\to A}\rangle_j$, and $F_{\text{ext}\to B}(j)=\langle F_{\text{ext}\to B}\rangle_j$.

Equation (13) can then be written in a matrix form, leading to the determination of coupling patch surfacic forces

$$\{F_{B\to A}\} = ([Y_A]+[Y_B])^{-1}([Y_B]\{F_{\text{ext}\to B}\} - [Y_A]\{F_{\text{ext}\to A}\}). \quad (14)$$

This extension of the mobility method, for surface coupling, is named the patch-mobility method and is applied in the following to sound transmission through double panels.

III. VIBROACOUSTIC RESPONSE OF A DOUBLE PANEL CALCULATED BY THE PATCH-MOBILITY METHOD

In this section, the patch-mobility method is applied to the vibroacoustic response of a double panel coupled to a cavity and radiating in the open external space.

In a first step, the system is divided into subsystems (cf. Fig. 2): A: emission chamber, B: first panel, C: cavity between the two panels, D: second panel, and E: semi-infinite medium.

Coupling surfaces are defined as: S_{AB} (respectively, S_{BC}, S_{CD}, S_{DE}) is the coupling surface between subsystems A and B (respectively B and C, C and D, D and E).

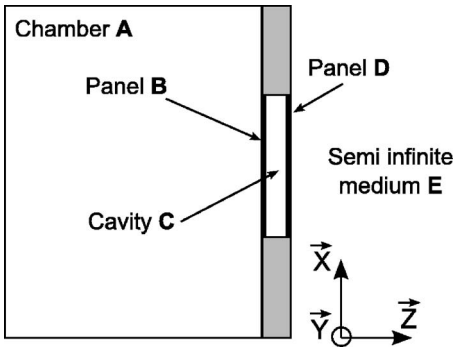


FIG. 2. Sketch of double panel coupled with a source room and a semi-infinite medium.

Equations governing patch velocities of plates, room, cavity, and infinite acoustic medium, write:

For the source room coupling surface

$$\langle V_A\rangle_i = \langle \tilde{V}_A\rangle_i + \sum_{j=1}^{N_{AB}} \langle\langle Y_A\rangle_i\rangle_j \langle F_{(B\to A)}\rangle_j. \quad (15)$$

For the first panel

$$\langle V_B\rangle_i = \langle \tilde{V}_B\rangle_i + \sum_{j=1}^{N_{AB}} \langle\langle Y_B\rangle_i\rangle_j \langle F_{(A\to B)}\rangle_j + \sum_{j=1}^{N_{BC}} \langle\langle Y_B\rangle_i\rangle_j \langle F_{(C\to B)}\rangle_j. \quad (16)$$

For the cavity coupling surfaces

$$\forall i \in S_{BC}: \langle V_{C1}\rangle_i = \langle \tilde{V}_{C1}\rangle_i + \sum_{j=1}^{N_{BC}} \langle\langle Y_{C1}\rangle_i\rangle_j \langle F_{(B\to C)}\rangle_j + \sum_{k=1}^{N_{CD}} \langle\langle Y_{C1}\rangle_i\rangle_k \langle F_{(D\to C)}\rangle_k, \quad (17)$$

$$\forall i \in S_{CD}: \langle V_{C2}\rangle_i = \langle \tilde{V}_{C2}\rangle_i + \sum_{j=1}^{N_{CD}} \langle\langle Y_{C2}\rangle_i\rangle_k \langle F_{(D\to C)}\rangle_k + \sum_{k=1}^{N_{BC}} \langle\langle Y_{C2}\rangle_i\rangle_j \langle F_{(B\to C)}\rangle_j. \quad (18)$$

For the second panel

$$\langle V_D\rangle_i = \langle \tilde{V}_D\rangle_i + \sum_{k=1}^{N_{CD}} \langle\langle Y_D\rangle_i\rangle_k \langle F_{(C\to D)}\rangle_k + \sum_{k=1}^{N_{DE}} \langle\langle Y_D\rangle_i\rangle_k \langle F_{(E\to D)}\rangle_k. \quad (19)$$

For the semi-infinite medium coupling surface

$$\langle V_E\rangle_i = \langle \tilde{V}_E\rangle_i + \sum_{k=1}^{N_{DE}} \langle\langle Y_E\rangle_i\rangle_k \langle F_{(D\to E)}\rangle_k. \quad (20)$$

The complete set of equations [Eqs. (15)–(20)] describe all the physical interactions taking part in the vibroacoustic response of the double panel. Nevertheless, some of these

phenomena can be neglected. For example, the first panel excitation can be separated in two acoustic fields: acoustic room response with blocked panel and pressure radiated by the panel. The radiated field, coming from the plate induced vibrations, is usually neglected as a second order phenomenon. On the contrary, the coupling between the cavity C and the two panels B and D is much more important. The small volume of the cavity C produces a high cavity stiffness that has a real effect upon the vibrations of the double panel.

After the discussed simplifications, Eqs. (16)–(19) write:

First panel

$$\langle V_B\rangle_i = \langle \tilde{V}_B\rangle_i + \sum_{j=1}^{N_{BC}} \langle\langle Y_B\rangle_i\rangle_j \langle F_{(C\to B)}\rangle_j. \tag{21}$$

Cavity

$$\forall i \in S_{BC}\colon \langle V_{C1}\rangle_i = \sum_{j=1}^{N_{BC}} \langle\langle Y_{C1}\rangle_i\rangle_j \langle F_{(B\to C)}\rangle_j + \sum_{k=1}^{N_{CD}} \langle\langle Y_{C1}\rangle_i\rangle_k \langle F_{(D\to C)}\rangle_k, \tag{22}$$

$$\forall i \in S_{CD}\colon \langle V_{C2}\rangle_i = \sum_{k=1}^{N_{CD}} \langle\langle Y_{C2}\rangle_i\rangle_k \langle F_{(D\to C)}\rangle_k + \sum_{j=1}^{N_{BC}} \langle\langle Y_{C2}\rangle_i\rangle_j \langle F_{(B\to C)}\rangle_j. \tag{23}$$

Second panel

$$\langle V_D\rangle_i = \sum_{k=1}^{N_{CD}} \langle\langle Y_D\rangle_i\rangle_k \langle F_{(C\to D)}\rangle_k. \tag{24}$$

Excitation appears in Eq. (21) as a patch velocity vector before coupling $\langle \tilde{V}_B\rangle_i$. It is calculated using blocked patch pressures (P_{Blocked}) applied to the first panel by the acoustic field in the source room. The first panel velocity before coupling is then calculated as follows:

$$\langle \tilde{V}_B\rangle_i = \sum_j \langle\langle Y_B\rangle_i\rangle_j \langle P\rangle_{j(\text{Blocked})} S_j. \tag{25}$$

Blocked patch pressures are calculated by integrating the pressure generated by a source in the emission chamber over the panel assumed to be rigid (blocked). Therefore room dimensions, excitation location, and position of the aperture in the room wall are, hence, taken into account. For rectangular cavity, modes are well known and calculation of room response is straightforward. Of course for large room the computation can be time consuming. However, for a given room configuration, blocked patch pressures are only calculated once whatever the plates under study.

Finally, continuity conditions over coupling surfaces S_{BC} and S_{CD} are written: $\forall i \in S_{BC}$: $\langle V_B\rangle_i = \langle V_C\rangle_i$, and $\forall i \in S_{CD}$: $\langle V_C\rangle_i = \langle V_D\rangle_i$. It yields to

$$\begin{Bmatrix} \langle \tilde{V}_1\rangle_i \\ 0 \end{Bmatrix} = \begin{bmatrix} \langle\langle Y_{C1}\rangle_i\rangle_j + \langle\langle Y_B\rangle_i\rangle_j & \langle\langle Y_{C1}\rangle_i\rangle_k \\ \langle\langle Y_{C2}\rangle_i\rangle_j & \langle\langle Y_{C2}\rangle_i\rangle_k + \langle\langle Y_D\rangle_i\rangle_k \end{bmatrix} \times \begin{Bmatrix} \langle F_{(B\to C)}\rangle_j \\ \langle F_{(D\to C)}\rangle_k \end{Bmatrix}. \tag{26}$$

Solving the linear system (26) allows us to get coupling forces $\langle F_{(B\to C)}\rangle_j$ and $\langle F_{(D\to C)}\rangle_k$ and then to calculate patch velocities after coupling by using Eqs. (21)–(24).

IV. PANEL PATCH MOBILITIES

In this section, panel mobilities are derived thanks to the Love–Kirchhoff equation of motion (flexural vibration of thin plates)

$$-\omega^2 \rho h U(x,y) + D^* \nabla^4 [U(x,y)] = P(x,y), \tag{27}$$

where $D^* = E^* h^3 / 12(1-\nu^2)$, and E^* is the complex Young modulus which takes into account structural damping such as $E^* = E(1+j\eta_s)$, and η_s is the damping loss factor.

The excitation term is defined on patch j as follows:

$$P(x,y) = \begin{cases} \dfrac{\langle F\rangle_j}{S_j} & \text{if } (x,y) \in \text{patch } j \\ 0 & \text{else} \end{cases}.$$

The modal expansion of the plate transverse displacement with simply supported boundary conditions writes

$$U(x_M, y_M) = \sum_{pq} a_{pq} \Phi_{pq}(x_M, y_M), \tag{28}$$

where mode shapes are given by

$$\Phi_{pq}(x,y) = \sin\left(\frac{p\pi}{L_x}x\right)\sin\left(\frac{q\pi}{L_y}y\right). \tag{29}$$

Incorporating Eq. (28) into Eq. (27), and using modes orthogonality yield to the modal amplitude

$$a_{pq} = \frac{\langle F\rangle_j \left[\int_j \Phi_{pq}(x,y)dS\right]}{S_j (K^*_{pq} - \omega^2 M_{pq})} \tag{30}$$

where $K^*_{pq} = \int_S D^* \Phi_{pq} \nabla^4(\Phi_{pq}) dS$, $M_{pq} = \int_S \rho h \Phi^2_{pq} dS$, and $\omega^*_{pq} = \sqrt{K^*_{pq}/M_{pq}}$.

The patch mobility as defined by Eq. (10) is then obtained

$$\langle\langle YP\rangle_i\rangle_j = \frac{j\omega}{S_i S_j} \sum_{pq} \frac{\left[\int_i \Phi_{pq}(x,y)dS\right]\left[\int_j \Phi_{pq}(x,y)dS\right]}{M_{pq}[\omega^*_{pq} - \omega^2]}. \tag{31}$$

Figure 3 shows an example of input and transfer panel patch mobilities versus frequency calculated for an aluminum panel ($Lx = 1.5$ m, $Ly = 0.96$ m, $Lz = 0.002$ m). As expected, transfer patch-mobility magnitude (thin curve) is lower than input patch-mobility magnitude (thick curve). It can also be noticed that as frequency increases, mobility curve is smoothed, due to modal overlap. However, this is more marked on input patch mobility than on transfer patch mobility.

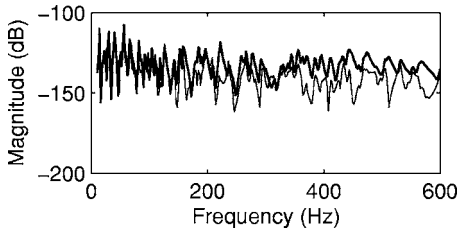


FIG. 3. Panel patch mobility—Thick line: Input patch mobility (patch 42), Thin line: Transfer patch mobility (patch 42/72)—Aluminium panel: $Lx=1.5$ m, $Ly=0.96$ m, $Lz=0.002$ m—Patch size: $\Delta x=7.9$ cm, $\Delta y=7.4$ cm—Patch 42: $X=[0.24\text{ m};\ 0.32\text{ m}]$, $Y=[0.15\text{ m};\ 0.22\text{ m}]$—Patch 72: $X=[1.11\text{ m};\ 1.18\text{ m}]$, $Y=[0.22\text{ m};\ 0.30\text{ m}]$.

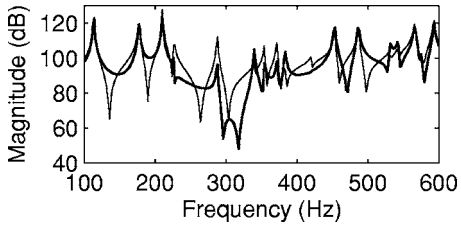


FIG. 4. Acoustic patch impedance—Thin line: Input patch impedance (patch 42), Thick line: Transfer patch impedance (patch 42/72)—Cavity: $Lx=1.5$ m, $Ly=0.96$ m, $Lz=0.01$ m—Patch size: $\Delta x=7.9$ cm, $\Delta y=7.4$ cm—Patch 42: $X=[0.24\text{ m};\ 0.32\text{ m}]$, $Y=[0.15\text{ m};\ 0.22\text{ m}]$—Patch 72: $X=[1.11\text{ m};\ 1.18\text{ m}]$, $Y=[0.22\text{ m};\ 0.30\text{ m}]$.

V. ACOUSTIC PATCH MOBILITIES OF A CAVITY

A. Problem description

Neumann inhomogeneous problem (32) is considered, that is to say an acoustic cavity with rigid wall boundary condition on $\partial\Omega_-$ and excited by a boundary velocity field on $\partial\Omega_+$. The aim is to find the pressure at any point M, radiated by a vibrating surface of normal velocity V_n:

$$\Delta P(M)+k^2P(M)=0 \quad \forall M\in\Omega,$$

$$\frac{\partial P(Q)}{\partial n}=-j\rho\omega V_n(Q) \quad \forall Q\in\partial\Omega_+,$$

$$\frac{\partial P(Q)}{\partial n}=0 \quad \forall Q\in\partial\Omega_-. \tag{32}$$

Acoustic damping is introduced with a complex wave number: $k^*=\omega/c^*=\omega/[c\sqrt{(1+j\eta_a)}]$, and to solve this problem, expansion on modes of undamped cavity is used.

B. Pressure modal expansion

Modal expansion is used to describe the cavity pressure field

$$P(M)=\sum_p a_p\psi_p(M). \tag{33}$$

After some calculations (cf. Appendix A), pressure radiated at point M is given by

$$P(M)=\sum_p\left(\frac{\int_{\partial\Omega_+}\{\psi_p(Q)[-j\rho\omega V_n(Q)]\}dQ}{(k^{*2}-k_p^2)N_p}\psi_p(M)\right). \tag{34}$$

To obtain the acoustic transfer impedance between two patches i and j, a uniform normal velocity V_n is given on patch j, and pressure (34) is integrated over the observation patch i:

$$\langle\langle Z\rangle_i\rangle_j=\frac{\int_{S_i}P(M)\bar{n}_i dS_i}{\frac{\int_{S_i}\bar{V}_n(M)dS_j}{S_j}}=\frac{\int_{S_i}P(M)\bar{n}_i dS_i}{\langle V_n\rangle_j}=\frac{\langle F\rangle_i}{\langle V_n\rangle_j}.$$

With pressure modal expansion, the following expression for patch acoustic impedance is obtained:

$$\langle\langle Z\rangle_i\rangle_j=-j\rho\omega\sum_p\frac{[\int_{Si}\psi_p(x,y)dS][\int_{Sj}\psi_p(x,y)dS]}{(k^{*2}-k_p^2)N_p}. \tag{35}$$

Equation (35) is the general cavity acoustic impedance expression. For a rectangular cavity with rigid walls, analytical expression of mode shapes can be derived

$$\psi_{pqr}(M)=\cos\left(\frac{p\pi x}{Lx}\right)\cos\left(\frac{q\pi y}{Ly}\right)\cos\left(\frac{r\pi z}{Lz}\right). \tag{36}$$

A mode being defined with three indices, expression (35) writes

$$\langle\langle Z\rangle_i\rangle_j=-j\rho\omega\times\sum_{pqr}\frac{[\int_{Si}\psi_{pqr}(x,y,Z_i)dxdy][\int_{Sj}\psi_{pqr}(x,y,Z_j)dxdy]}{(k^{*2}-k_{pqr}^2)N_{pqr}}. \tag{37}$$

The global impedance matrix describing the cavity behavior writes

$$\begin{pmatrix}\langle F\rangle_1\\ \vdots\\ \langle F\rangle_n\end{pmatrix}=\underbrace{\begin{bmatrix}\langle\langle Z\rangle_1\rangle_1 & \cdots & \langle\langle Z\rangle_1\rangle_n\\ \vdots & \ddots & \vdots\\ \langle\langle Z\rangle_n\rangle_1 & \cdots & \langle\langle Z\rangle_n\rangle_n\end{bmatrix}}_{[Z]}\cdot\begin{pmatrix}\langle V\rangle_1\\ \vdots\\ \langle V\rangle_n\end{pmatrix}$$

Then patch mobility matrix is obtained as the inverse of patch impedance matrix

$$\begin{pmatrix}\langle V\rangle_1\\ \vdots\\ \langle V\rangle_n\end{pmatrix}=\underbrace{\begin{bmatrix}\langle\langle Y\rangle_1\rangle_1 & \cdots & \langle\langle Y\rangle_1\rangle_n\\ \vdots & \ddots & \vdots\\ \langle\langle Y\rangle_n\rangle_1 & \cdots & \langle\langle Y\rangle_n\rangle_n\end{bmatrix}}_{[Y]=[Z]^{-1}}\cdot\begin{pmatrix}\langle F\rangle_1\\ \vdots\\ \langle F\rangle_n\end{pmatrix}.$$

Figure 4 presents two acoustic patch input impedances versus frequency of a cavity ($Lx=1.5$ m, $Ly=0.96$ m, $Lz=0.01$ m) that show its modal behavior. One can see also that modes are responding differently according to patch location. Acoustic patch impedances between patches located on opposite boundary surfaces such as $Z_i\neq Z_j$ are not presented since they are not different from impedances between patches located on the same boundary surface such as $Z_i=Z_j$. This is of course due to the small thickness of the cavity, and the studied frequency range.

Figure 5 presents two acoustic patch mobilities versus frequency for the same cavity ($Lx=1.5$ m, $Ly=0.96$ m, Lz

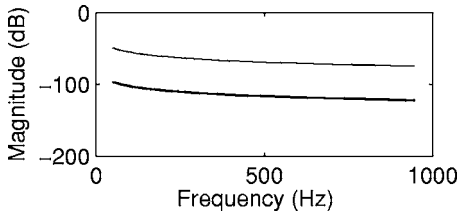


FIG. 5. Acoustic patch mobility—Thin line: Input patch mobility (patch 42), Thick line: Transfer patch mobility (patch 42/72)—Cavity: Lx=1.5 m, Ly=0.96 m, Lz=0.01 m—Patch size: Δx=7.9 cm, Δy=7.4 cm—Patch 42: X=[0.24 m; 0.32 m], Y=[0.15 m; 0.22 m]—Patch 72: X=[1.11 m; 1.18 m], Y=[0.22 m; 0.30 m].

=0.01 m). Results are surprising since patch mobilities do not exhibit the modal behavior that can be seen on patch impedances. To explain this strange result, one has to remember that direct patch mobility calculation will be obtained by solving the Dirichlet problem (null boundary pressure except on excited patch), and calculating the patch averaged velocity created at boundary. Cavity resonances with Dirichlet boundary conditions are obviously different from that of a cavity with Neumann boundary conditions. The first resonance of the rectangular cavity with Dirichlet boundary conditions is equal to $(c/2\pi)\sqrt{(\pi/Lx)^2+(\pi/Ly)^2+(\pi/Lz)^2}$, that is to say 17 000 Hz. That explains why no resonance appears in the studied frequency band.

VI. SEMI-INFINITE MEDIUM

This section describes the method used to calculate the power radiated into the semi-infinite receiving medium.

A. Radiated pressure

Radiation impedance is defined as the ratio of averaged patch i radiated pressure to averaged patch j velocity: $\langle\langle Z\rangle_i\rangle_j=\langle P_{\text{rad}}\rangle_i/\langle V\rangle_j$. The radiated pressure can thus be calculated from panel patch velocities (cf. Fig. 6).

Radiation impedance calculations are based on Rayleigh's integral and are detailed in Appendix B. A particular case is defined when emitting and receiving patch are the same (i.e., $i=j$). In this case radiation impedance expression is given by Eq. (38) where "a_i" is the radius of a circular patch of surface S_i. Otherwise, when emitting and receiving patch are different (i.e., $i\neq j$) Eq. (39) is used to calculate radiation impedance where d_{ij} is the distance between two patches central points

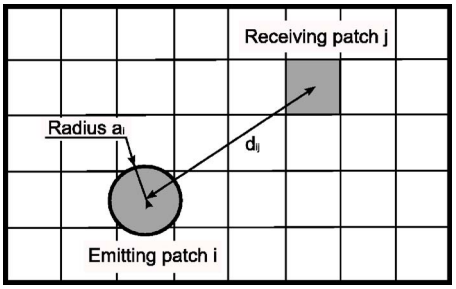


FIG. 6. Radiating panel description.

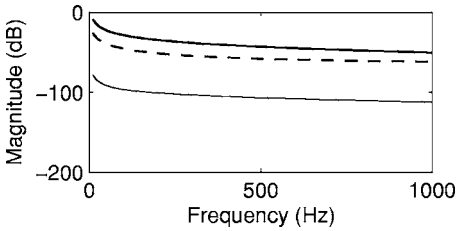


FIG. 7. Patch radiating mobility—Thick solid line: Input patch mobility, Thick dotted line: Transfer patch mobility with d_{ij}=0.08 m, Thin line: Transfer patch mobility with d_{ij}=0.84 m—Panel: Lx=0.96 m, Lz=1.5 m–Patch size: Δx=7.9 cm, Δy=7.4 cm.

$$\langle\langle Z\rangle_i\rangle_i=\frac{\langle P_{\text{rad}}\rangle_i}{\langle V\rangle_i}=\rho_0 c[1-\exp^{-jka}], \tag{38}$$

$$\langle\langle Z\rangle_i\rangle_j=\frac{\langle P_{\text{rad}}\rangle_i}{\langle V\rangle_j}=\frac{1}{2\pi}\rho_0 j\frac{\exp^{-jkd_{ij}}}{d_{ij}}S_j. \tag{39}$$

Radiation patch mobilities are obtained by inversion of impedance matrix calculated from previous Eqs. (38) and (39). Figure 7 illustrates radiation patch mobility behavior. As one can see, radiation patch mobility magnitude decreases with frequency and also when distance between radiating and receiving patch d_{ij} decreases.

B. Radiated power

Radiated power is calculated from patch velocities and radiated patch pressures, and can be written using radiation patch mobilities (40):

$$I_{\text{rad}}=\frac{1}{2}\sum_i \text{Re}\{\langle V\rangle_i^*\langle P_{\text{rad}}\rangle_i\}=\frac{1}{2}([Y_{\text{rad}}]^{-1}\{V\})'\{V\}^*. \tag{40}$$

VII. SOURCE ROOM MODELING

This section describes a simple way to model the source room as depicted in Fig. 8 using its modal behavior. Two quantities are presented: the mean quadratic room pressure that is used to calculate double panel transmission loss, and blocked patch pressures (BPP) that are used to calculate patch velocities before coupling in Eq. (25). To calculate these quantities one has to replace the flexible panel by a rigid wall.

Standard room response modal expansion writes

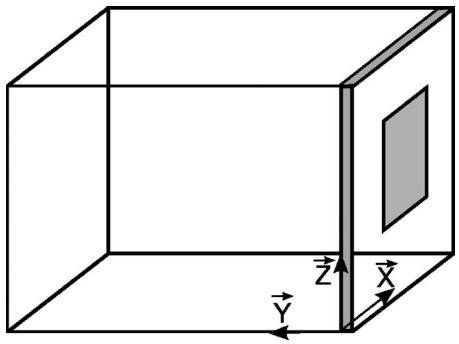


FIG. 8. Sketch of the excitation room.

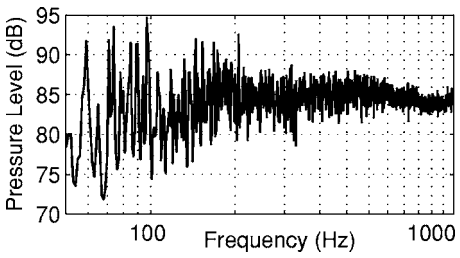


FIG. 9. Room pressure level—room dimensions: $lx=11.5$ m, $ly=8.69$ m, $lz=4.03$ m—Source at: $Xs=2$ m, $Ys=4$ m, $Zs=1$ m, amplitude: $S_0=2$—Cutoff frequency: 187 Hz.

$$P(x,y,z)=\sum_{p,q,r} A_{pqr}\psi_{pqr}(x,y,z), \tag{41}$$

where $A_{pqr}=\int_\Omega \psi_{pqr}(x,y,z)S(x,y,z)d\Omega/(k^{*2}-k_{pqr}^2)N_{pqr}$, and $N_{pqr}=\int_\Omega \psi_{pqr}^2(x,y,z)d\Omega$. A point source excitation is used such as: $S(x,y,z)=S_0\delta(x-x_0)\delta(y-y_0)\delta(z-z_0)$. The wave number k^* takes into account fluid damping with a complex sound speed: $k^*=\omega/c^*=\omega/c\sqrt{(1+j\eta_r)}$, where $\eta_r=2.2/fT_r$ and T_r is reverberation time.

The source room, being assumed to be a parallelepipedic-shaped room with rigid walls, has the following mode shapes:

$$\psi_{pqr}(x,y,z)=\cos\left(\frac{p\pi}{lx}x\right)\cos\left(\frac{q\pi}{ly}y\right)\cos\left(\frac{r\pi}{lz}z\right).$$

A. Quadratic room pressure

Mean quadratic room pressure calculation that is useful in the prediction of the transmission loss, is straightforward

$$P_r^2=\sum_{pqr}\frac{|A_{pqr}^2|\varepsilon_p\varepsilon_q\varepsilon_r}{8}$$

$$\text{and } \varepsilon_p=\begin{cases}1 & \text{if } p=0\\ 2 & \text{if } p\neq 0\end{cases}. \tag{42}$$

Figure 9 shows the mean quadratic pressure given by Eq. (42). At low frequency, despite the large room volume ($\approx$400 m^3), modal behavior is predominant. Modes (2,1,1) at 55 Hz, (1,2,1) at 59 Hz, (2,2,1) at 65 Hz are clearly identified. On the contrary, at higher frequency modal overlap occurs, reducing the mean quadratic room pressure variations with frequency. The cut off frequency given by $f_c=\sqrt{c^3T_r/8.8\pi V}$ corresponds roughly to the limit between modal and diffuse field behavior. Above that frequency, modal separation is lower than the −3 dB bandwidth. In the present case ($T_r\approx 10$ s and $V\approx 400$ m^3), the cutoff frequency is equal to 187 Hz.

B. Blocked patch pressure

Modal expansion (41) of the room response gives the pressure distribution in the room. Blocked patch pressures over a panel are then calculated by integrating the room pressure over a patch surface located on a rigid wall

$$\langle P\rangle_i=\int_{S_i}\left[\sum_{p,q,r} A_{pqr}\psi_{pqr}(x,y_0,z)\right]dxdz. \tag{43}$$

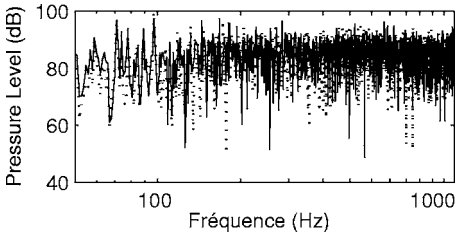


FIG. 10. Blocked patch pressure level on two patches—Dotted line: patch 124, Solid line: patch 1—Located at $X_{124}=6$ m, $Y_{124}=0$ m, $Z_{124}=1.75$ m, and $X_1=5.245$ m, $Y_1=0$ m, $Z_1=1.27$ m—Room dimensions: $lx=11.5$ m, $ly=8.69$ m, $lz=4.03$ m—Patch size: $\Delta X=0.08$ m, $\Delta Z=0.074$ m—Source position: $Xs=2$ m, $Ys=4$ m, $Zs=1$ m, amplitude: $S_0=2$—cutoff frequency: 187 Hz.

These blocked patch pressures are used in Eq. (25) to calculate panel patch velocities before coupling and enable to solve the set of Eqs. (21)–(24).

Figure 10 illustrates variations with frequency of two blocked patch pressures. Peaks and antipeaks can have different values, but both curves have the same averaged value over frequency (85 dB). These differences come obviously from mode shapes. Moreover, a large amplitude of variation with frequency is noticed even at high frequency, contrary to space averaged pressure in the room presented in Fig. 9. Blocked patch pressure behavior is a bit surprising compared to standard way of thinking on reverberant pressure in large room, however, it is a quasilocal quantity, contrary to reverberant pressure that is averaged over the room volume. Contrary to mean quadratic pressure, no obvious frequency limit between modal behavior and diffuse field, such as the cutoff frequency (187 Hz), can be observed on blocked patch pressure.

Figure 11 presents the blocked patch pressure distribution pattern at two frequencies for two window locations on source room wall. As a general rule, the blocked patch pressure patterns display clearly the acoustic wave length at each frequency. However, a change in pressure distribution is observed for the two window locations, that leads to transmission loss differences particularly at low frequency.

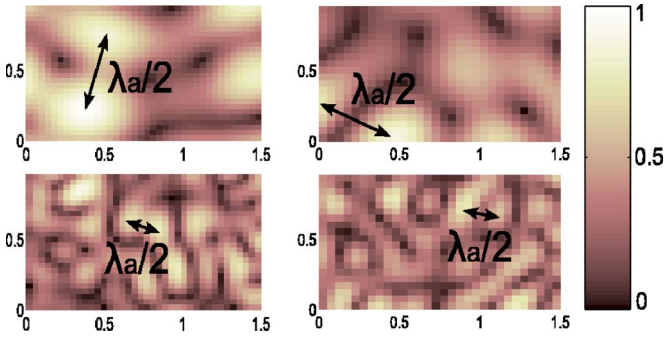


FIG. 11. (Color online) Blocked patch pressure magnitude for two window locations ($lx=0.96$ m, $lz=1.5$ m)—on the top: 400 Hz—on the bottom: 1000 Hz—on the left, the window is centered on point $X=6$ m, $Y=0$ m, $Z=1.75$ m—on the right the window is centered on point $X=6.96$ m, $Y=0$ m, $Z=2.78$ m—Room dimensions: $lx=11.5$ m, $ly=8.69$ m, $lz=4.03$ m—Patch size: $\Delta X=0.04$ m, $\Delta Z=0.037$ m—Source position: $Xs=2$ m, $Ys=4$ m, $Zs=1$ m, amplitude: $S_0=2$—cutoff frequency: 187 Hz.

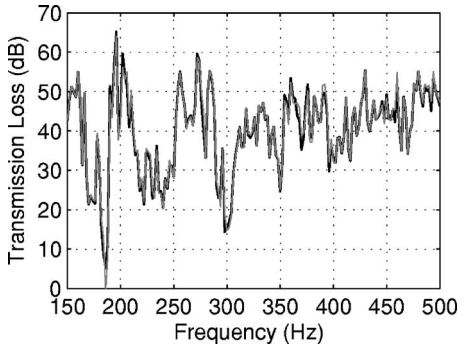


FIG. 12. Transmission loss comparison with two meshes—Thick line: mesh 30×20, Thin grey line: mesh 8×5—Double panels in steel ($Lx=0.6$ m, $Lz=0.4$ m), 2 and 1.5 mm thickness, separated by 1 cm of air.

VIII. ACOUSTIC TRANSMISSION THROUGH DOUBLE PANELS

This section presents the transmission loss calculation through a double panel.

A. Transmission loss calculation

Transmission loss τ is defined as the ratio of transmitted power Π_t to incident power Π_i. Transmitted power can be calculated thanks to the mobility method with Eq. (40). With the procedure of standard measurement of transmission loss, incident power is deduced from mean quadratic room pressure using a classical result of Sabine's theory: $\Pi_i = S(P_r^2/4\rho_0 c)$. Of course the assumption of Sabine's theory is not exactly verified for standard rectangular rooms specially at low frequency, but is nevertheless used for sake of simplicity.

This way of calculating transmission loss in a reverberant field (i.e., by using blocked patch pressures) avoids integration over incidence angles that is usually employed. When good accuracy is desired, several incidence angles must be considered with a plane wave summation method, and time needed to compute transmission loss becomes tremendous compared to time necessary with room blocked patch pressures excitation.

B. Convergence criterion

The usual mesh condition for finite element analysis is given by $\Delta L < \lambda_{\min}/6$, and sometimes $\lambda_{\min}/4$ is admitted.

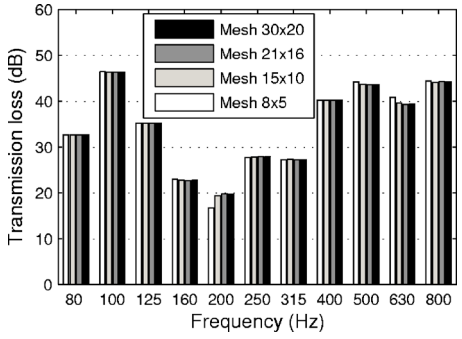


FIG. 13. Mesh comparison with third octave analysis—Double panels in steel ($Lx=0.6$ m, $Lz=0.4$ m), 2 and 1.5 mm thickness, separated by 1 cm of air.

TABLE I. Maximal frequency given by different criterions.

Criterion	$\lambda_a/6$	$\lambda_a/4$	$\lambda_a/2$	$\lambda_s/6$	$\lambda_s/4$	$\lambda_s/2$
Mesh 30×20	2833	4250	8500	1026	2309	9237
Mesh 21×16	1983	2975	5950	503	1132	4526
Mesh 15×10	1416	2125	4250	256	577	2309
Mesh 8×5	708	1062	2125	64	144	577

If we assume that the same criterion can be used for the patch mobility method, therefore the maximum patch length necessary to go up to 1000 Hz is 5.7 cm for acoustic media (8.5 cm with $\lambda_{\min}/4$). For the structure, one needs to look at the bending wave length $\lambda = c_{flx}/f = (1/f)\sqrt{\omega}\sqrt[4]{Eh^2/\rho\times12(1-\nu^2)}$.

Four meshes have been tested to find which criterion had to be taken with linear elements. Figure 12 presents the transmission loss calculated with two meshes, and Fig. 13 shows the transmission loss comparison with a third octave analysis for each mesh. Table I gives structural and acoustical criterions associated with each mesh.

Meshes 30×20, 21×16, and 15×10 give similar results: the three curves are superimposed and cannot be distinguished. The 8×5 mesh curve starts to be slightly different around 150–200 Hz. This frequency can be viewed as the limit frequency given by the $\lambda_s/4$ criterion. However, results obtained with mesh 8×5 can be accepted up to 577 Hz which is the limit given by criterion $\lambda_s/2$. This result agrees with criterions found in Ref. 33 where authors showed that a $\lambda/2$ criterion could be taken to couple linear acoustic problems with a similar method (patch transfer functions).

In the following, presented results are obtained with elementary plane rectangular patches verifying the $\lambda_s/2$ criterion.

C. Influence of room characteristics on acoustic transmission

Several studies have shown transmission loss measurements differences in interlaboratory comparisons at low frequencies not only for small reverberant rooms, but even for large ones. In this section, two large rooms of same volumes are studied (cf. Table II). Transmission loss results for both rooms are plotted in Fig. 14. Differences of transmission loss are up to 6 dB for the same double panel. At 78 Hz, it can be explained by looking at the BPP distribution (Fig. 15) that are quite different for the two source rooms. This result was expected since it is known that for frequencies below the cutoff frequency, the acoustic field cannot be considered as diffuse. At higher frequencies differences are still noticeable, that is more unexpected. As can be seen in Fig. 15, though

TABLE II. Test room dimensions, source location, and panel location.

	lx (m)	ly (m)	lz (m)	X_{source} (m)	Y_{source} (m)	Z_{source} (m)	$X_{\text{centralpatch}}$ (m)	$Z_{\text{centralpatch}}$ (m)
Room 1	4.03	5	10	0.8	2	2.5	1.75	5
Room 2	10	5	4.03	2	2	1	5	1.75

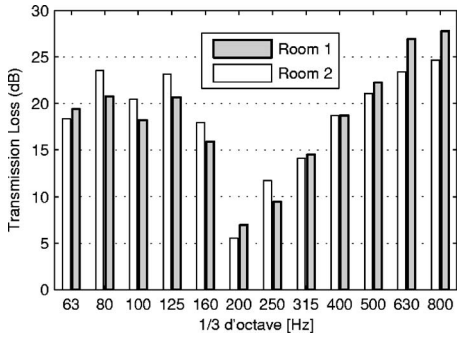


FIG. 14. Comparison between two rooms—Double panels in aluminum (Lx=0.96 m, Lz=1.5 m), 2 and 1.5 mm thickness, separated by 3 cm of air—Patch size: ΔX=0.08 m, ΔZ=0.074 m—Source amplitude: S_0=2—cutoff frequency: 187 Hz.

the field is more diffuse at 800 Hz, the two rooms do not provide the same pressure distribution over the panel. Moreover, as long as structural wavelength is not much higher than acoustical wavelength, differences in pressure distribution over the panel will have an effect on the panel response and on sound transmission. Therefore, even above the cutoff frequency, panel excitation depends on room characteristics, source location, and aperture location.

IX. COMPARISON WITH A FINITE ELEMENT MODEL

The patch-mobility method has been applied to the double panel studied by Panneton and Atalla with a finite element model in Ref. 23: simply supported aluminum plates (0.35 m by 0.22 m, 1 mm thickness) separated by 0.0762 m of air. Excitation is realized by a normal incident plane wave. Results obtained with patch mobility method are presented up to 4000 Hz in Fig. 16. A good agreement is observed with results obtained by Panneton and Atalla in Ref. 23, at least up to maximal finite element calculation frequency (500 Hz).

Two particular aspects of the patch-mobility method must be underlined in this comparison: the mesh criterion in $\lambda/2$ that is very coarse compared to the criterion in $\lambda/6$ used in finite element method (FEM) and the possibility to characterize each subsystem separately with the patch mobility method contrary to FEM (if one element is modified, only its own characterization has to be calculated before solving interaction equations).

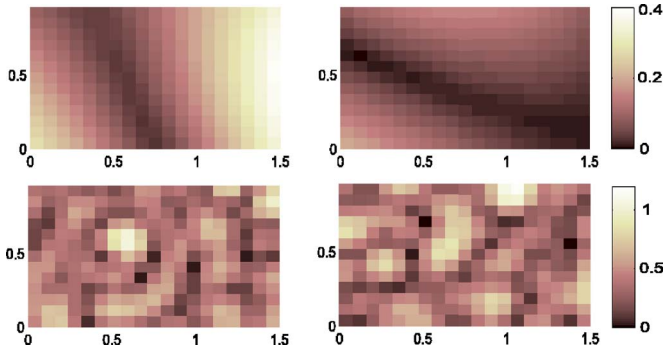


FIG. 15. (Color online) Blocked patch pressure distribution—Top: 78 Hz—Bottom: 800 Hz—Left: room 1, Right: room 2.

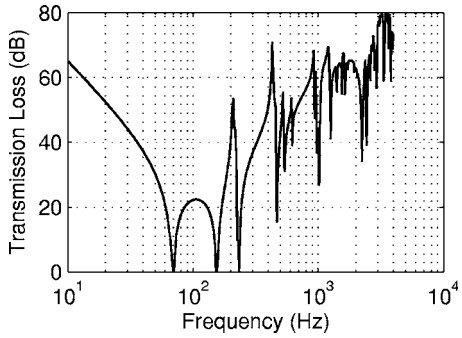


FIG. 16. Normal incidence sound transmission loss of a double aluminum panel (0.35 m by 0.22 m, 1 mm thickness) separated by 0.0762 m of air—Results obtained with the patch-mobility method.

X. EXPERIMENTAL RESULTS

Figure 17 presents the experimental setup. Measurements of transmitted power were done using intensity technique with a 50 mm intensity probe located at 20 cm from the radiating panel. 100 points have been used to measure radiated intensity. Source room is a reverberant room (lx =11.5 m, ly=8.69 m, lz=4.03 m, cutoff frequency=187 Hz) where walls are not parallels and with diffusers in the corners. Five microphones in the reverberant chamber enabled to measure reverberant pressure. Receiving room is not reverberant and is treated to be the most isolated from exterior sound and with a high sound absorption to limit reflections of radiated sound by tested panel. It can thus be considered as a semi anechoic enclosure.

Transmission loss of a double aluminum panel was then measured. The double panel dimensions were: width 0.96 m, length 1.5 m, thickness 2 mm and 1.5 mm, air gap thickness 1 cm. Panel critical frequencies are 7961 Hz for the aluminum panel of 1.5 mm thickness, and 5971 Hz for the aluminum panel of 2 mm thickness. The theoretical mass-spring-mass frequency for diffuse field is 328 Hz.

Figures 18 and 19 show comparisons of double panels transmission loss obtained by experiment and with the patch-mobility model. The model gives the same tendency as measurements and differences observed on some third octaves are quite reasonable. Indeed, compared to reproducibility and repeatability values mentioned in norm ISO 140-3, differences observed with third octave analysis between the model and the experiment are quiet good (example of repeatability values given by the norm 4.5 dB at 100 Hz and 1.5 dB at

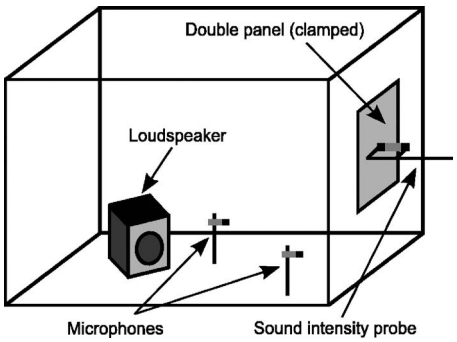


FIG. 17. Experimental setup.

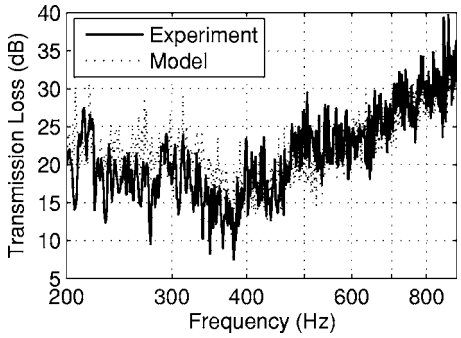


FIG. 18. Comparison with experiment—Solid line: experiment, dotted line: model—Double aluminum panel (Lx=0.96 m, Lz=1.5 m), 2 and 1.5 mm thickness, separated by 3 cm of air, centered on point X=6 m, Y=0 m, Z=1.75 m—Room dimensions: lx=11.5 m, ly=8.69 m, lz=4.03 m—Patch size: ΔX=0.08 m, ΔZ=0.074 m—Source at: Xs=2 m, Ys=4 m, Zs=1 m, amplitude: S_0=2—cutoff frequency: 187 Hz—structural damping: 0.06—fluid damping: 0.035.

1000 Hz). An intercomparison of laboratory measurements, realized in 23 European laboratories, shows also a high dispersion of airborne sound insulation measurements. This sudy is presented in Ref. 33. Dispersions are not negligible, and even at high frequencies differences up to 5 dB are observed. Therefore, the mobility model can be considered as reliable to predict transmission loss through double panels.

Concerning cavity damping, particular care is needed to determine its value. Indeed, cavity damping is known to be very influential upon sound transmission between the double panel resonance frequency and the critical frequencies of the panels. As mentioned in Ref. 34, cavity damping can be much higher than typical fluid damping due to neglected physical phenomenon like viscous and thermal dissipations near boundaries. In this reference, cavity damping is set to 10^{-3} while damping in surrounding fluid is set to 10^{-8}. In the following, cavity damping has been used to adjust the model with experiment, and a value of 0.035 has been found in the case of the double panel with a cavity of 1 cm thickness. It is also important to notice that cavity damping depends on the air layer thickness. Indeed, for thicker layers, effects of viscous and thermal dissipations will decrease leading to smaller cavity damping.

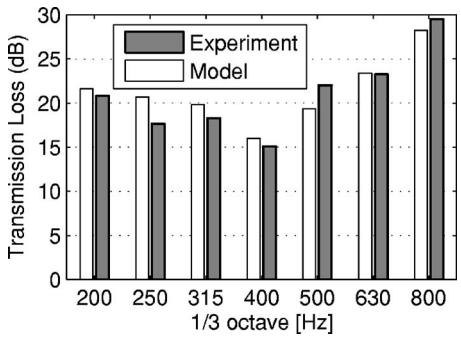


FIG. 19. Comparison with experiment—Double aluminum panel (Lx=0.96 m, Lz=1.5 m), 2 and 1.5 mm thickness, separated by 3 cm of air, centered on point X=6 m, Y=0 m, Z=1.75 m—Room dimensions: lx=11.5 m, ly=8.69 m, lz=4.03 m—Patch size: ΔX=0.08 m, ΔZ=0.074 m—Source at: Xs=2 m, Ys=4 m, Zs=1 m, amplitude: S_0=2—cutoff frequency: 187 Hz—structural damping: 0.06—fluid damping: 0.035.

XI. CONCLUSIONS

A patch-mobility method is presented in this paper and is used to predict transmission loss of finite double panels. Excitation is modeled with blocked patch pressures that are calculated from the source room response, taking thus into account source room dimensions, absorption, excitation position, and panel location. Calculation time is, hence, considerably reduced compared to classical plane waves summation technique to represent reverberant field. Moreover, the substructural approach used in the patch-mobility method enables to characterize each subsystem separately before coupling. Another powerful numerical aspect of this method is the mesh criterion in $\lambda_s/2$ that is very large compared to standard finite element analysis mesh criterion. The patch-mobility model and finite element model give the same results when comparing transmission loss of double panels. However, the patch-mobility method enables to study higher frequencies. Finally, a good agreement is found with experimental measurements realized on a double aluminium panel.

APPENDIX A: CAVITY MOBILITY

After having multiplied the Helmoltz Eq. (32) by the mode shape $\psi_p(M)$, the integration over the volume Ω yields to

$$\int_\Omega [\Delta P(M) + k^{*2} P(M)]\psi_p(M) d\Omega = 0. \tag{A1}$$

Using Green's formulas gives

$$\int_\Omega [\Delta \psi_p(M) + k^{*2}\psi_p(M)]P(M) dM - \int_{\partial\Omega} \left[\psi_p(Q) \frac{\partial P(Q)}{\partial n} - P(Q) \frac{\partial \psi_p(Q)}{\partial n} \right] dQ = 0. \tag{A2}$$

Conditions over boundary surface $\partial\Omega$ are defined over different surfaces: $\partial P(Q)/\partial n = -j\rho\omega V_n(Q)$ $\forall Q \in \partial\Omega_+ \partial P(Q)/\partial n = 0$ $\forall Q \in \partial\Omega_- \partial\psi_p(Q)/\partial n = 0 \forall Q \in \partial\Omega$.

Equation (A2) becomes

$$\int_\Omega [\Delta \psi_p(M) + k^{*2}\psi_p(M)]P(M) dM = \int_{\partial\Omega_+} \{\psi_p(Q)[-j\rho\omega V_n(Q)]\} dQ. \tag{A3}$$

Pressure is then replaced by its modal expansion (33):

$$\int_\Omega [\Delta \psi_p(M) + k^{*2}\psi_p(M)] \sum_m [a_m \psi_m(M)] dM = \int_{\partial\Omega_+} \{\psi_p(Q)[-j\rho\omega V_n(Q)]\} dQ. \tag{A4}$$

Thanks to modes orthogonality, this expression is simplified

$$\int_{\Omega}[-k_p^2\psi_p(M)+k^{*2}\psi_p(M)]a_p\psi_p(M)dM$$

$$=\int_{\partial\Omega_+}\{\psi_p(Q)[-j\rho\omega V_n(Q)]\}dQ. \quad \text{(A5)}$$

Mode amplitude a_p is then obtained

$$a_p=\frac{\int_{\partial\Omega_+}\{\psi_p(Q)[-j\rho\omega V_n(Q)]\}dQ}{(k^{*2}-k_p^2)N_p} \quad \text{(A6)}$$

with $N_p=\int_{\Omega}\psi_p^2(M)dM$.

Using this modal amplitude into Eq. (33) enables to find the pressure radiated at point M:

$$P(M)=\sum_p\left(\frac{\int_{\partial\Omega_+}\{\psi_p(Q)[-j\rho\omega V_n(Q)]\}dQ}{(k^{*2}-k_p^2)N_p}\psi_p(M)\right). \quad \text{(A7)}$$

APPENDIX B: RADIATION IMPEDANCE

Rayleigh's integral (B1) is used to calculate the radiated pressure of the panel. Two integrals can be written depending on the observation point location. Indeed, for a point located on the boundary surface, the Green's function singularity needs to be avoided, through the calculation of Cauchy's principal value integral (B2):

$$P(M_0)=\int_{\partial\Omega_+}\left[j\rho\omega V_n(Q)\frac{\exp(-jk\|QM_0\|)}{2\pi\|QM_0\|}\right]dQ \quad \forall M_0\in Q, \quad \text{(B1)}$$

$$P(Q_0)=\oint_{\partial\Omega_+}\left[j\rho\omega V_n(Q)\frac{\exp(-jk\|QQ_0\|)}{2\pi\|QQ_0\|}\right]dQ \quad \forall Q_0\in\partial Q. \quad \text{(B2)}$$

The vibrating surface $\partial\Omega_+$ is separated in patches S_i leading to equation

$$P(Q_0)=\sum_{j=1,j\neq i}^{N}\int_{S_j}\left[j\rho\omega V_n(Q_j)\frac{\exp(-jk\|Q_jM_0\|)}{2\pi\|Q_jM_0\|}\right]dQ_j+\oint_{S_i}\left[j\rho\omega V_n(Q_i)\frac{\exp(-jk\|Q_iQ_0\|)}{2\pi\|Q_iQ_0\|}\right]dQ_i. \quad \text{(B3)}$$

Let us introduce the distance d_{ij} between two patch centers. If the patch dimensions are small, the integral can be approximated by the value at the central point times the patch area. Moreover, the velocity $V_n(Q)$ being approximated by a uniform value over the patch, it can be put out of the integral. These assumptions lead to the classical approximation of the integral

$$\int_{S_j}\left[j\rho\omega V_n(Q_j)\frac{\exp(-jk\|Q_jM_0\|)}{2\pi\|Q_jM_0\|}\right]dQ_j=\frac{1}{2\pi}\rho_0j\omega\langle V\rangle_j\frac{\exp^{-jkd_{ij}}}{d_{ij}}S_j, \quad \text{(B4)}$$

where d_{ij} is the distance between two patches.

This approximation is valid for small enough patch surface, in general a discretization based on the standard $\lambda/6$ criterion is sufficient.

The use of cylindrical coordinates enables us to calculate Cauchy's principal value of the integral by considering a circular patch having the same surface as the original one. The result is immediate

$$\oint_{S_i}\left[j\rho\omega V_n(Q_i)\frac{\exp(-jk\|Q_iQ_0\|)}{2\pi\|Q_iQ_0\|}\right]dQ_i=\rho_0c\langle V\rangle_i[1-\exp^{-jka}]. \quad \text{(B5)}$$

[1] L. Cremer, "Theorie der schalldämmung dünner bei schragen einfall (theory of sound absorption of thin walls)," Akust. Z. **7**, 81–104 (1942).
[2] A. London, "Transmission of reverberant sound through single walls," J. Res. Natl. Bur. Stand. **42**, 605 (1949).
[3] A. London, "Transmission of reverberant sound through double walls," J. Acoust. Soc. Am. **22**, 270–279 (1950).
[4] T. Vogel, "Sur les vibrations de certains systèmes élastiques dans un champ sonore (vibrations of elastic systems in an acoustic field)," Ph.D. thesis, Marseille, 1947.
[5] M. Battacharya and M. Crocker, "Forced vibration of a panel and radiation of sound into room," Acustica **22**, 275–294 (1970).
[6] R. Guy and M. Battacharya, "The transmission of sound through a cavity-backed finite plate," J. Sound Vib. **27**, 207–223 (1973).
[7] K. Mulholland and R. Lyon, "Sound insulation at low frequencies," J. Acoust. Soc. Am. **54**, 867–878 (1973).
[8] L. Beranek, *Noise Reduction* (McGraw-Hill, NY, 1960).
[9] F. Fahy, *Sound and Structural Vibration: Radiation Transmission and Response* (Academic, New York, 1985).
[10] K. Ballagh, "Accuracy of prediction methods for sound transmission loss," Internoise Proceedings, 2004 (INCE, Ames, IA).
[11] N. Atalla and R. Bernhard, "Review of numerical solutions for low-frequency structural-acoustic problems," Appl. Acoust. **43**, 271–294 (1994).
[12] L. Gagliardini, J. Rolland, and J. Guyader, "The use of a functional basis to calculate acoustic transmission between rooms," J. Sound Vib. **145**, 457–478 (1991).
[13] J. Guyader and B. Laulagnet, "Structural acoustic radiation prediction: Expanding the vibratory response on a functional basis," Appl. Acoust. **43**, 247–269 (1994).
[14] A. Price and M. Crocker, "Sound transmission through double panels using statistical energy analysis," J. Acoust. Soc. Am. **47**, 683–693 (1970).
[15] R. Craik and R. Smith, "Sound transmission through leaf lightweight partitions—Part i: Airborne sound," Appl. Acoust. **61**, 223–245 (2000).
[16] W. Kropp and E. Rebillard, "On the air-borne sound insulation of double wall constructions," Acust. Acta Acust. 85, 707–720 (1999).
[17] J. Guyader, "Transparence acoustique de plaques multicouhes orthotropes, viscoélastiques, finies (acoustic transparency of multilayered orthotropic viscoplastic plates)," Ph.D. thesis, Université Claude Bernard, Lyon I, 1977.
[18] J. Guyader and C. Lesueur, "Acoustic transmission through orthotropic multilayered plates, part ii: Transmission loss," J. Sound Vib. **58**, 69–86 (1978).
[19] J. Guyader and C. Lesueur, "Transmission of reverberant sound through orthotropic viscoelastic multilayered plates," J. Sound Vib. **70**, 319–332 (1980).
[20] E. Nilsson and A. Nilsson, "Prediction and measurement of some dynamic properties of sandwich structures with honeycomb and foam cores," J. Sound Vib. **251**, 409–430 (2002).
[21] R. Panneton and N. Atalla, "An efficient finite element scheme for solving the three dimensional poroelasticity problem in acoustics," J. Acoust. Soc. Am. **101**, 3287–3298 (1997).
[22] N. Atalla, M. A. Hamdi, and R. Panneton, "Enhanced weak integral formulation for the mixed (u, p) poroelastic equations," J. Acoust. Soc. Am. **109**, 3065–3068 (2001).
[23] R. Panneton and N. Atalla, "Numerical prediction of sound transmission through finite multilayer systems with poroelastic materials," J. Acoust. Soc. Am. **100**, 346–354 (1996).

[24]F. Spronck, "Transparence au son des parois minces viscoélastiques finies et infinies, simples ou multiples (sound transparency of thin infinite viscoelastic plates)," Ph.D. thesis, Marseille, 1971.
[25]J. Coyette, G. Lielens, M. Robbé, and P. Neple, "An efficient method for evaluating diffuse field joint acceptance functions for cylindrical and truncated conical geometries," J. Acoust. Soc. Am. **117**, 1009–1019 (2005).
[26]J. O'Hara, "Mechanical impedance and mobility concepts," J. Acoust. Soc. Am. **41**, 1180–1184 (1967).
[27]B. Petersson, "A thin-plate model for the moment mobility at the intersection of two perpendicular plates," J. Sound Vib. **108**, 417–485 (1986).
[28]A. T. Moorhouse, "A dimensionless mobility formulation for evaluation of force and moment excitation of structures," J. Acoust. Soc. Am. **112**, 972–980 (2002).
[29]A. T. Moorhouse and B. Gibbs, "Calculation of the mean and maximum mobility for concrete floors," Appl. Acoust. **45**, 227–245 (1995).
[30]S. Naji, "Etude des transmissions vibratoires par une méthode de mobilité mixte dans les assemblages par surface (study on vibration transmissions with a mobility method)," Ph.D. thesis, Université Claude Bernard Lyon I, 1993.
[31]G. Orefice, C. Cacciolati, and J. Guyader, "The energy mobility," J. Sound Vib. **254**, 269–295 (2002).
[32]M. Ouisse, L. Maxit, C. Cacciolati, and J. Guyader, "Patch transfer functions as a tool to couple linear acoustics problems," J. Vibr. Acoust. **127**, 458–466 (2005).
[33]P. Fausti, R. Pompoli, and R. S. Smith, "An intercomparison of laboratory measurements of airborne sound insulation of lightweight plasterboard walls," Build. Acoust. **6**, 127–140 (1999).
[34]J. Brunskog, "The influence of finite cavities on the sound insulation of double-plate structures," J. Acoust. Soc. Am. **117**, 3727–3739 (2005).

A model of the VU (volume-unit) meter, with speech applications

Bryce E. Lobdell and Jont B. Allen
Beckman Institute, University of Illinois, Urbana, Illinois 61801

(Received 26 September 2005; revised 11 October 2006; accepted 12 October 2006)

The *Volume-Unit* (VU) meter, used in speech research prior to the advent of computers and modern signal processing methods, is described in signal processing terms. There are no known software implementations of this meter, which meet the 1954 ASA standard and provide the instantaneous needle level. Important speech applications will be explored, such as making comparisons of speech levels to earlier classic works, and measuring speech levels using traditional methods on modern computers. It is our intention to make this venerable method of measuring speech levels available once again. The VU meter is simulated and its properties are studied. A 1950s vintage and a recent vintage VU meter are studied by comparing the transient responses to tones and measurement of speech levels. Based on these measurements, a software VU meter (henceforth referred to as *VUSOFT*) is simulated, and verified. The method for reading the meter is explained, and simulated in software. The VU level for speech is shown to depend on the reading duration. The relationship between the root-mean-squared (rms) level of a signal and the VU level of a signal is determined, as a function of the meter-reading time.
[DOI: 10.1121/1.2387130]

PACS number(s): 43.58.Fm, 43.71.Gv [DOS] Pages: 279–285

I. INTRODUCTION

It is important to know how to make speech level measurements. Traditionally this was the job of the *VU meter*, an instrument which was used by radio engineers, audiologists, and speech perception scientists, to measure the level of speech sounds. Not every "sound level meter" having a microphone and needle is a VU meter. The VU meter is an industry standard device. Knowledge of speech VU levels are required for the proper interpretation of many speech perception experiments, since most of the early experiments depended on the VU speech levels (Castner and Carter, 1933).

Following the work of Fletcher and Steinberg (1930), the classic speech loudness measurements of Fletcher and Munson (1933) helped establish the importance of speech level measurements. French and Steinberg (1947) relied extensively on data from papers by Dunn and White (1940) and Sivian (1929). In particular, they used the average spectrum of speech and the cumulative level distribution versus long average intensity, in 1/8 s intervals. We shall show that the effect of a 1/8 s root-mean-square (rms) average is similar to that performed by a VU meter. During World War II, Harvard university adopted the methods developed at Bell Labs. For example, Miller and Nicely (1955) used a VU meter to control the signal-to-noise ratio (SNR) and speech level. When repeating such experiments, it is helpful (and arguably necessary) to have the VU meter measurement method available. As a result, a VU meter was obtained and simulated, as reported here.

Some of the issues developed here were touched upon by previous studies, namely Ludvigsen (1992) and Sjogren (1973). Sjorgren compared the consistency of eight different speech level measurements, including the VU meter, by measuring the level of consonant-vowel-consonant (CVC) sounds and monosyllabic words relative to the level of a carrier phrase. Ludvigsen went on to conclude that measurement methods that integrated in time, such as the VU meter, were preferable to "impulse" measurements. Thus, the need remains for a software simulation of the VU meter.

The VU meter standard is described in detail. A 1950s vintage VU meter and a recent vintage VU meter are measured and compared to the VU meter standard. A simulation of the VU meter, denoted *VUSOFT* is described and verified. Finally, the effect of the VU meter-reading method on the VU level is described, and comparisons between the root-mean-squared (rms) level and the VU level are presented.

II. SUMMARY OF THE VU METER STANDARD

In response to the need for a standard and effective way of measuring program levels (i.e., music and speech) for transmission purposes, Columbia Broadcasting Systems, the National Broadcasting Company, and the Bell Telephone Laboratories devised and published materials (Chinn *et al.*, 1940) describing the device that would later be called the VU meter.

As described in Bohn (2000), in 1942 the American Standards Association (ASA) published a standard for VU meters (ASA, 1942). This standard was followed by the IRE standard in 1953 (IRE, 1953) also known as IEEE Standard #152-1953, and another ASA standard in 1954 (ASA, 1954), upon which our investigations are based. The most recent standard IEC 60268-17 (IEC, 1990) is not relevant to work published prior to 1990.

According to the ASA standard (ASA, 1954), the VU meter is the output of a full wave rectifier followed by volt meter, comprised of a mass, spring constant, and damping of the meter movement, whose response to a sudden and steady input should reach 99% of its final value within 0.3 ± 0.03 s, and shall overshoot its final value by at least 1%, but not more than 1.5%. The response of the VU meter to steady sine

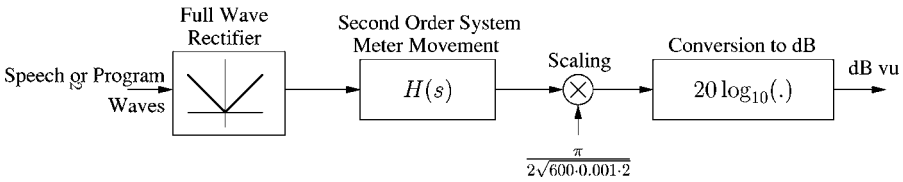


FIG. 1. Block diagram of the VU meter implied by the VU meter standard.

waves should not diminish more than 0.2 dB between 25 Hz and 10 kHz from the response to a 1 kHz sine wave. The VU meter output should be scaled to read in "dB vu" (while the standard says the unit is "VU," "dB vu" has become the accepted unit). Figure 1 shows the block diagram implied by the standard. The system described can be implemented on a computer by cascading the absolute value of the input voltage signal, with the proper second order system, and scaling and conversion to decibels. The second order system with the response described by the standard is a low-pass filter with a very low cutoff frequency (around 8 Hz). Conceptually, that means the VU level is a moving average of absolute value of the input signal. For periodic or steady signals such as a tone or noise, the VU level is the average absolute value of the signal. The parameters of the continuous and discrete time second order systems are derived in Appendix A. The MATLAB code that implements the VU meter standard is given in Appendix B, dubbed *VUSOFT*.

A VU meter reads in decibels, $20\log_{10}(V/V_{\text{ref}})$, where V is the meter voltage and V_{ref} is the level of a 1 kHz tone that will deliver 1 mW into a 600 Ω impedance. Thus $V_{\text{ref}} = (2/\pi)\sqrt{2\cdot600\cdot0.001}$ V, which is about −3 dBV.

A. Harmonic distortion

A full wave rectifier generates harmonics. In a discrete time simulation of a VU meter such harmonics alias, causing the simulated VU meter to breach the standard (i.e., no variations are allowed larger than 0.2 dB from the response to a steady tone at 1 kHz). This problem is solved by an upsample rate conversion of the discrete time input signal to at least eight times its original rate before the full wave rectifier (Oppenheim and Schafer, 1998).

B. Nonlinearity

The ASA standard refers to a nonlinearity in the rectifier used in VU meters "the exponent of whose characteristic is 1.2±0.2." A 1950s vintage VU meter was examined (further details in Sec. III and in Appendix C) to determine the effects of any such non-linearity on the ballistics of that VU meter. It was discovered that the VU meter faceplate is graduated in a way that removes the effect of the nonlinearity, and that has a negligible effect on the ballistics of the VU meter needle.

III. COMPARISON OF VU METERS TO THE STANDARD

VUSOFT was designed based on the specifications in Sec. II. The MATLAB code and derivation can be found Appendices B and A, respectively. To verify that *VUSOFT* implements the ASA VU meter standard correctly, it was compared with a 1950s vintage VU meter and a recent vintage VU meter. The 1950s vintage hardware VU meter was labeled "VOLUME INDICATOR, Type 911-B, Ser. No. D-8941, The Daven Co., Newark NJ." The recent vintage VU meter was manufactured by Simpson Electric Co. (520 Simpson Avenue, Lac du Flambeau, WI 54538). The transient responses of the three meters were compared, along with the peak VU level with short speech sounds.

A. Methods for transient response comparison

The response of a second order system can be described by any two of several parameters. The two easiest parameters to measure are the peak time t_p and the overshoot M_p. The peak time t_p is the amount of time it takes for the step response of a system to reach its highest level. The overshoot is the amount by which the step response of a system will exceed its final value. The overshoot is measured by applying a long-duration reference tone and then noting by how much the meter needle exceeds its final value. The peak time is measured by playing successively longer reference tones, until increasing the length of the reference tone no longer increases the maximum level the needle reaches. The length of the tone at which the maximum level reached no longer increases is taken as the peak time.

B. Results for transient response comparison

Figure 2 shows the step response of the three VU meters. Note that the 1950s vintage hardware VU meter used

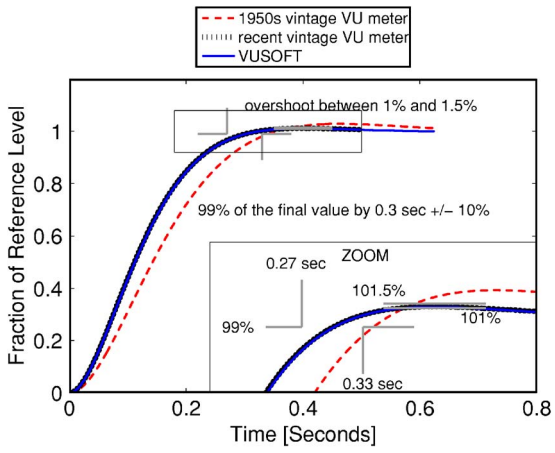


FIG. 2. (Color online) This figure shows the step response of *VUSOFT* and the two hardware VU meters. The stimulus is 1 kHz reference tone. This figure shows the instantaneous output of the second order system shown in Fig. 1 after scaling and before the conversion to decibels. The ordinate is scaled so that the reading is unity (0 dB vu) in response to the reference tone after the needle movement has had time to settle.

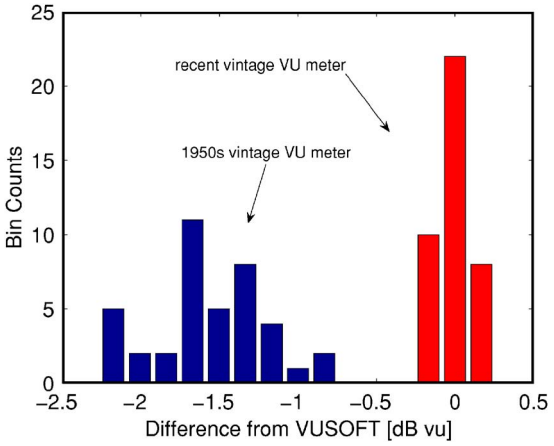


FIG. 3. (Color online) A histogram of the peak level recorded by the hardware VU meters in response to short speech recordings. The short speech recordings were scaled so that their peak level measured by *VUSOFT* was 0 dB vu.

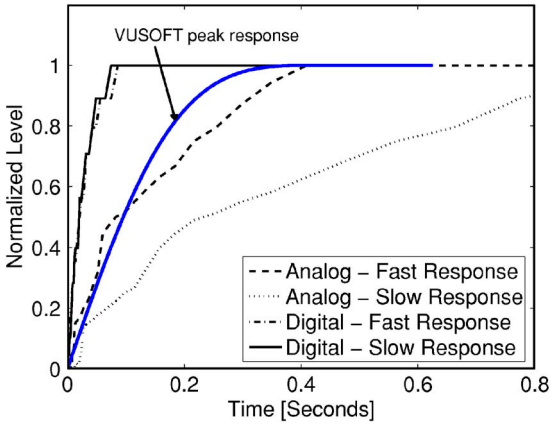


FIG. 4. (Color online) Peak response of Radio Shack sound level meters.

in this study does not meet the ASA specification, since the overshoot is too large, and it has a slightly longer rise time. This difference in the transient response causes an average difference in the VU reading for short syllables of −1.6 dB, as shown in Fig. 3. The recent vintage VU meter meets the specified transient response, as does *VUSOFT*.

To understand the sensitivity of these differences with speech as the input, we measured peak VU meter levels of 40 speech recordings. We tested speech material consisting of isolated consonant-vowel pairs. A computer was used to store and play back the sounds into the two hardware VU meters, and the largest displacement of the VU meter needle was recorded. All speech sounds were normalized to read 0 dB vu using *VUSOFT*. A calibration tone, specified by the ASA standard (ASA, 1954), was used to assure that all three VU meters were identically calibrated.

C. Speech level results

Figure 3 shows a histogram of the peak VU levels of the hardware VU meters and *VUSOFT*. The mean difference between 1950s vintage VU meter and *VUSOFT* is −1.6 dB vu, with a standard deviation of 0.37 dB vu. The mean difference between the recent vintage VU meter and *VUSOFT* is 0.009 dB vu with a standard deviation of 0.09 dB vu. The recent vintage VU meter provides readings that are more consistent with *VUSOFT*, because they have more similar transient responses.

D. Radio Shack "sound level meters"

Two Radio Shack digital and analog meters (catalog numbers 33-2055 and 33-4050) were purchased and tested to determine if they would be a suitable substitute for a VU meter. The peak responses of these instruments are shown in Fig. 4, compared to *VUSOFT*. The digital sound level meter has a peak response that rises much faster than *VUSOFT* and thus also faster than the ASA standard), while the analog meter response is slower. It was also determined that the response of both of the Radio Shack sound level meters depends on the SPL range setting (i.e., the transient response is different depending on whether it is set to read 60–70 dB or 70–80 dB, etc.). Thus, as noted in Radio Shack's manual, neither of these meters conforms to any VU meter standard specification.

IV. READING THE VU METER

Reading a VU meter is more of an art than a science. The duration of the recording turns out to be a critical variable, as we shall show next. With regard to the reading method, the ASA standard for VU meters reads as follows (ASA, 1954):

> The reading is determined by the greatest deflections occurring in a period of about a minute for program waves, or a shorter period (e.g., 5 to 10 s) for message telephone speech waves, excluding not more than one or two deflections of unusual amplitude.

The authors asked several "experts" how they read VU meters. We were told by to pick the three highest levels for a segment of speech material and average them together. This method is claimed to be less subjective, and purported by the experts to be the true "standard method" for reading the VU level of speech material.

Figure 5 shows the waveform of a speech signal along with the *VUSOFT* output. High speech levels occur less frequently than low speech levels. Due to the small probability of the tails of the probability distribution, the longer the recording, the higher the peak level. In other words, "The longer you measure, the larger the VU level you will record." The goal in the following study is to quantify the relationship between the rms level, the time duration of the speech sample, and the peak VU level.

Our results are derived from a histogram of the *VUSOFT* output for 26 hours of speech as well as a count of the *VUSOFT* output peaks and the amplitude of those peaks. VU levels reported on in this section were generated exclusively by *VUSOFT*. All the speech material was normalized to the same rms level (computed over the whole speech file, typically several minutes). The speech material was from a corpus titled "ICSI Meeting Speech" produced by the Linguistic Data Consortium (http://www.ldc.upenn.edu), catalog number 2004S02. The speech involved approximately equal

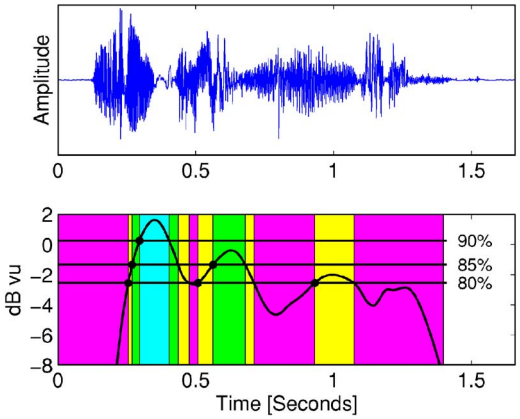


FIG. 5. (Color online) The top panel is the acoustic waveform for the phrase "No one pronounced zing seventh." The bottom panel is the *VUSOFT* output for that phrase time aligned with the acoustic waveform. The horizontal lines show the 80%, 85%, and 90% VU meter levels, denoted the *percentage of intervals*, as defined in the text.

numbers of male and female talkers conversing. This speech material was chosen because it was conversational in nature, involved a large number of speakers, and was never compressed or otherwise modified.

Figure 5 illustrates the peaks in the VU meter output for a particular speech phrase. The term *percentage of intervals* refers to the VU level compared to the distribution of VU levels (with VU levels sampled periodically). When we speak of a *percentage level of 90%*, the level is greater than 90% of other levels observed in speech for a fixed speech rms level. The horizontal lines in Fig. 5 show the 80%, 85%, and 90% levels for a particular speech recording.

A. VU meter and the level distribution of speech

The solid line in Fig. 6 shows the cumulative distribution of VU levels relative to the rms of speech. This figure was generated by computing levels for the speech material described above, and making a histogram of those levels. The histogram was converted to a cumulative level distribution where the levels are given relative to the rms level.

The dashed line in Fig. 6 is the result from Fig. 4 of French and Steinberg (1947), which was computed from the data of Dunn and White (1940) and Sivian (1929). It is not surprising that the relationship for the cumulative distribution of VU levels is similar to the result of Dunn and White (1940) because the meter has a similar frequency response to the 1/8 s window used by Dunn and White (1940), as illustrated in Fig. 7.

Figure 8 shows the relationship between the time duration that the VU meter level is monitored and the ratio of the VU peak level and the rms level, in dB. For each level the number of peaks of that level were counted. The average length of time between the peaks of each level was computed by dividing the length of the speech material by the number of peaks counted. This figure is particularly important because it allows one to compare the VU meter method described in the ASA standard to the rms level.

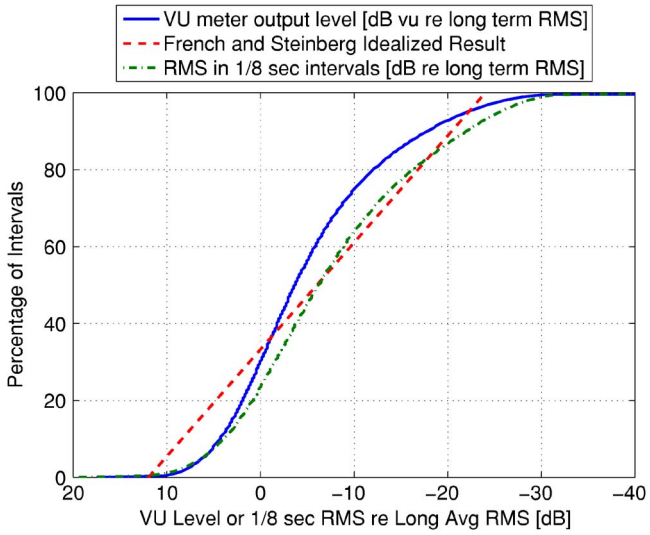


FIG. 6. (Color online) The solid line shows the cumulative distribution of VU levels (generated by *VUSOFT*) relative to the rms of speech, and compares it to the method of level measurement used by Dunn and White (1940). The dash-dotted line shows the cumulative distribution of rms levels in 1/8 s intervals, which is identical to the data shown in Fig. 4 of French and Steinberg (1947), taken from Dunn and White (1940). The idealized result of French and Steinberg is shown with the dashed line. For the solid line, the abscissa is the VU level (in dB vu) minus the long term rms level in decibels (computed over the whole speech recording, typically several minutes). For the dashed and dash-dotted lines, the abscissa is the ratio (in decibels) of the rms in 1/8 s intervals to the long term rms level. The ordinate is the percentage of 1/8 s intervals or VU levels (equally spaced in time) that are greater than the level shown on the abscissa.

V. RESULTS AND DISCUSSION

We are unaware of any ASA (1954) compliant software VU meter simulations that provide the instantaneous numerical needle position. Such a software simulation is necessary for comparison with other speech level measures (such as rms) and also automated level control using the VU meter in modern computer controlled speech experiments.

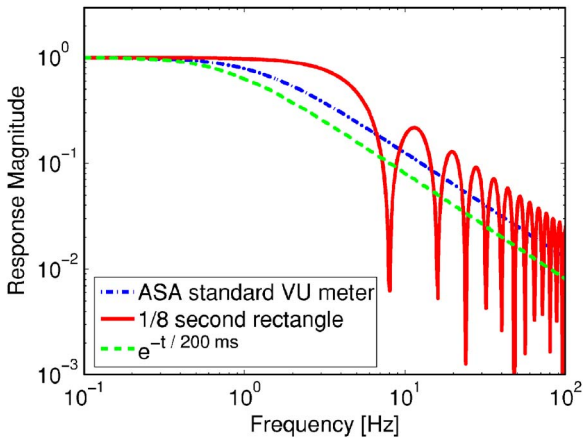


FIG. 7. (Color online) The dashed line shows the frequency response of the 200 ms integration related to the loudness of tones, described by Munson (1947). The solid line shows the frequency response of the 1/8 s window used by Dunn and White (1940) and Sivian (1929) to measure speech levels. The dash-dotted lines shows the frequency response of the second order system described by the ASA VU meter standard.

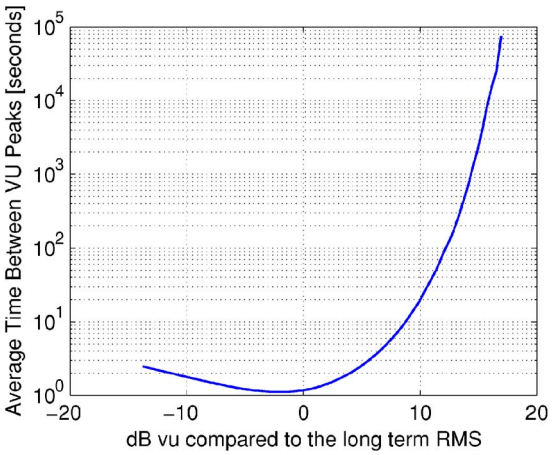


FIG. 8. (Color online) The duration between VU peaks (log s) as a function of the VU level peak level divided by the long terms rms (dB).

The ideal VU meter is a full wave rectifier followed by a second order low-pass system. The VU meter level is reported in dB vu referenced to a 1 kHz sin wave that will dissipate 1 mW into a 600 Ω resistor. A MATLAB© code (called *VUSOFT*) that implements the standard can be found in Appendix B.

Our VU meter reading method is to observe the highest peak. Figure 8 shows how the largest peak depends on observation duration. The ASA specified reading methods states that the VU level is the "greatest deflections occurring in a period of about a minute for program waves, or a shorter period (e.g., 5 to 10 s) for message telephone speech waves, excluding not more than one or two deflections of unusual amplitude." From Fig. 8 we conclude that the VU level observed over 5 to 10 s intervals will be 6–9 dB higher than the rms level, and that the VU level observed over a 1 min interval will be roughly 12 dB higher than the rms level.

The transient response of a 1950s vintage VU meter and a recent vintage VU meter were evaluated to confirm that we have accurately duplicated their behavior with *VUSOFT*. All three VU meters were very close to the standard specified response, leading us to conclude that we had properly interpreted the standard and duplicated it in *VUSOFT*. The 1950s vintage VU meter had an overshoot of which was 1.75% greater, and a peak time 0.06 s longer than that of the standard, while the recent vintage VU meter had a nearly identical transient response to the standard (Fig. 2). For short speech sounds, the peak level measured by the 1950s vintage VU meter was 1.6 dB vu lower on average than that measured by *VUSOFT*, while for the same set of speech sounds, the recent vintage VU meter differed from *VUSOFT* by 0.009 dB vu, on average.

The transient response of two Radio Shack "Sound Level Meters" were compared to the transient response of the VU meter to determine if they would make a suitable substitute for a VU meter. The sound level meters had a significantly different transient response and therefore would result in different observed levels.

The *sound level meter* standard published by the American National Standards Institute (ANSI) is different from the ASA VU meter standard, and will provide different level measurements for speech as a result of its different transient response. For example, the ANSI meter standard indicates that the needle level shall have an overshoot of 0 to 1.1 dB for the "fast response" setting and 0 to 1.6 dB for the "slow response" setting, which is significantly larger than the 0.09 to 0.13 dB overshoot specified for ASA standard VU meters. An ANSI sound level meter could potentially be used to measure speech levels, however, the specifications are less tight than the ASA VU meter standard and would therefore not be conducive to reproducibility between sound level meter instruments.

It is important when measuring speech levels to know that the transient response of the measurement device has a significant impact on the observed level, that the "VU meter" has tight specifications, and that not every level measurement device is a VU meter. Figure 2 and 3 illustrate how a small difference in transient response leads to an average difference of 1.6 dB vu for short speech sounds. The intensity just noticeable difference (JND) is less than this value.

The noise level and the signal-to-noise ratio (SNR) are critical components of many types of speech perception experiments; thus we would like to know how the rms measurement of noise compares to the VU-based measurement. For Gaussian noise the average absolute value is $\sigma\sqrt{2/\pi}$, where σ^2 is the variance of the noise (measured in volts squared). The VU level of the noise is then $20\log_{10}(\sigma/\sqrt{4\cdot 600\cdot 0.001/\pi})$ which is numerically equal to $20\log_{10}\sigma+1.17$ dB vu, where σ has the unit of volts (rms).

In summary, users of VU meters should be aware that the VU level still has important applications, that a VU meter is a standardized device with tight specifications, and that it is possible to relate the VU level for different methods of reading the VU meter with the rms level.

ACKNOWLEDGMENTS

The authors would like to thank Mead Killion for his expert advice and assistance in obtaining VU meters.

APPENDIX A: DERIVATION OF THE VUSOFT DIFFERENCE EQUATION

The ASA standard says that the VU meter needle should have a response that overshoots by 1% but not more than 1.5% and that reaches 99% of its final value in 0.3 s. Results from linear systems analysis can be used to derive the parameters for a continuous-time system which has the required response. Details can be found in Ogata (1997).

A second order mass-stiffness system has a frequency response defined by the Laplace transform,

$$H(s)=\frac{\omega_n^2}{s^2+2\zeta\omega_n s+\omega_n^2},\quad s=j2\pi f, \tag{A1}$$

where ω_n is the undamped natural frequency of the system, ζ is the damping ratio, s is the Laplace variable, and f is the frequency in Hz. The parameters ω_n and ζ conveniently specify the step response of a 2nd order system, which is

$$c(t)=1-e^{-\zeta\omega_n t}\left(\cos\omega_d t+\frac{\zeta}{\sqrt{1-\zeta^2}}\sin\omega_d t\right), \tag{A2}$$

$$\omega_d = \omega_n \sqrt{1-\zeta^2}.$$

The parameters for the second order system can be computed by combining Eq. (A2) with

$$M_p = e^{-(\zeta/\sqrt{1-\zeta^2})\pi} \quad \text{and} \quad c(t_r) = 0.99, \tag{A3}$$

where M_p is the overshoot (i.e., M_p=0.0125) and t_r is the time the system takes to reach 99% of its final value (i.e., t_r=0.3 s). The equation $c(t_r)$=0.99 is the constraint that the step response $c(t)$ reaches 99% of its final value in t_r seconds. Combining these equations, we find that the second order system that describes the VU meter needle ballistics has parameters ζ=0.812 72 and ω_n=13.512.

For our simulation we need a discrete time version of this system, which may be found using the bilinear transform having a z-transform (Oppenheim and Schafer, 1998), given by

$$H(z) = \frac{b_0 + 2b_1 z^{-1} + b_2 z^{-2}}{a_0 + a_1 z^{-1} + a_2 z^{-2}}, \tag{A4}$$

where $z=e^{j\omega}$. The corresponding difference equation is

$$a_0 y[n] = a_1 y[n-1] + a_2 y[n-2] + b_0 x[n] + 2b_1 x[n-1] + b_2 x[n-2]. \tag{A5}$$

The following parameters are computed using the bilinear transform: $b_0=2b_1=b_2=T_d^2\omega_n^2$, $a_0=4+4\zeta\omega_n T_d+\omega_n^2 T_d^2$, $a_1=-8+2\omega_n^2 T_d$, and $a_2=4-4\zeta\omega_n T_d+\omega_n^2 T_d^2$, and T_d is the sampling period for the discrete time system. For example, at a sampling rate of 44.1 kHz, $b_0=b_2=9.3876\times10^{-8}$, $b_1=4.6938\times10^{-8}$, $a_0=4.0010$, $a_1=-8.0000$, and $a_2=3.9990$.

APPENDIX B: MATLAB CODE FOR VUSOFT

The ASA VU meter specifications relevant to a software VU meter simulation are met by the following lines of the MATLAB code (©Bryce Lobdell 2006)

(http://www.auditorymodels.org/lobdell/vusoft/vusoft.m):

```
function y=vusoft (x, fs)
% Copyright 2006, Bryce Lobdell
% Parameters for the system:
% Td=1/fs/d⇒oversample by 8x to prevent aliasing.
wn=13.5119; eta=0.8127; D=8; Td=1/fs/D;
% Parameters for the filter:
B=Td^2*wn^2*[1 2 1];
A=[(4+4*eta*wn*Td+wn^2*Td^2)(-8
+2*wn^2*Td^2)...(4-4*eta*wn*Td+wn^2*Td^2)];
% Scale:
scaling=pi/2/sqrt(600*0.001*2);
% Upsample the input signal by 8x.
x_u=resample (x, D, 1, 50);
% Apply the absolute value, and the filter.
y1_u=scaling*filter(B,A,abs(x_u));
% Downsample back to the original rate.
y1=y1_u(1:D:end); y=20*log10(y1);
```

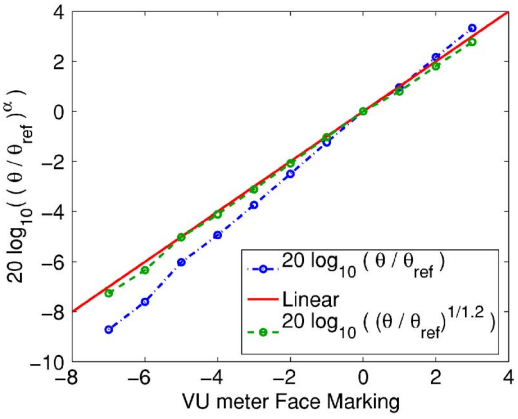


FIG. 9. (Color online) This figure shows the relationship between the needle angle and VU markings. If the dash-dotted line had a slope of one, the meter angle θ would be proportional to the voltage applied to the meter and rectifier package. The meter face is marked to compensate for the slope of 1.2, due to the meter's current I to voltage V relationship given by $I \propto V^{1.2}$.

APPENDIX C: NONLINEARITY OF THE RECTIFIER

The ASA standard ASA (1954) says that the VU meter shall be equivalent:

> "to the response with a direct current meter and a rectifier, the exponent of whose characteristic is 1.2±0.2"

We interpret this excerpt to mean that the current-voltage characteristic of the rectifier is θ (needle angle) $\propto I \propto V^{1.2}$. Several measurements were done to verify the exponent on the meter reading using a 1 kHz tone of varying level. Figure 9 shows the needle angle compared to the marked VU level. The dash-dotted line shows the needle angle θ measured with a protractor, compared to the marked VU level. The dashed line shows the relationship between $(\theta/\theta_{\text{ref}})^{1/1.2}$ and the VU level, which is linear. This implies that the meter current is proportional to $(V/V_{\text{ref}})^{1.2}$. It was verified that the VU meter face markings compensate for the non-linearity in the rectifier by comparing the VU level of tones of various levels.

The rectifiers used were most likely the copper-oxide type, as described in detail by Brattain (1951).

ASA (**1942**). *Volume Measurements of Electrical and Speech Program Waves*, American Standards Association.

ASA (**1954**). *Volume Measurements of Electrical and Speech Program Waves*, American Standards Association.

Bohn, D. A. (**2000**). "The Bewildering Wilderness: Navigating the complicated and frustrating world of audio standards." Sound and Video Contractor, pp. 56–64.

Brattain, W. (**1951**). "The copper oxide rectifier," Rev. Mod. Phys. **23**, 203–212.

Castner, T., and Carter, C. (**1933**). "Developments in the application of articulation testing," Bell Syst. Tech. J. **12**, 347–372.

Chinn, H. A., Gannett, D. K., and Morris, R. M. (**1940**). "A new standard volume indicator and reference level," Bell Syst. Tech. J. **19**, 94–137.

Dunn, H., and White, S. (**1940**). "Statistical measurements on conversational speech," J. Acoust. Soc. Am. **11**, 278–288.

Fletcher, H., and Munson, W. (**1933**). "Loudness, its definition, measurement, and calculation," J. Acoust. Soc. Am. **5**, 82–108.

Fletcher, H., and Steinberg, J. (**1930**). "Articulation testing methods," J. Acoust. Soc. Am. **1**, 17–64.

French, N., and Steinberg, J. (**1947**). "Factors governing the intelligibility of

speech sounds," J. Acoust. Soc. Am. **19**, 90–119.
IEC (**1990**). "Sound System equipment. Part 17: Standard volume indicators," International Electrotechnical Commission.
IRE (**1953**). *IEEE Standard for Audio Program Level Measurement*.
Ludvigsen, C. (**1992**). "Comparison of certain measures of speech and noise level," Scand. Audiol. **21**, 23–29.
Miller, G., and Nicely, P. (**1955**). "An analysis of perceptual confusions among Some English Consonants," J. Acoust. Soc. Am. **27**, 338–352.
Munson, W. (**1947**). "The growth of auditory sensation," J. Acoust. Soc. Am. **19**, 584–591.
Ogata, K. (**1997**). *Modern Control Engineering*, Prentice–Hall, Upper Saddle River, NJ, 3rd ed., pp. 234–259.
Oppenheim, A. V., and Schafer, R. W. (**1998**), *Discrete-Time Signal Processing* (Prentice–Hall, Upper Saddle River, NJ, 2nd ed., pp. 450–465.
Sivian, L. (**1929**). "Speech power and its measurement," J. Acoust. Soc. Am. **1**, 1–16.
Sjogren, H. (**1973**). "Objective measurements of speech level," Audiology **12**, 47–54.

Variance analysis of arrival time and amplitude estimates from random speckle signal

Svein-Erik Måsøy and Bjørn Angelsen
Department of Circulation and Imaging, Norwegian University of Science and Technology, Olav Kyrres Gate 9, 7489, Trondheim, Norway

Trond Varslot
Department of Mathematical Sciences, Norwegian University of Science and Technology, ALfred Getz Vei 1, 7034, Trondheim, Norway

(Received 16 March 2006; revised 26 September 2006; accepted 5 October 2006)

An aberration estimation algorithm, previously developed in the frequency domain, is implemented in the time domain. The algorithm is used to estimate arrival time and amplitude fluctuations with signals from random scatterers. Simulations have been performed to investigate the variance of the estimates. Stability differences between the two implementations are also explored. Eight body wall models, emulating the human abdominal wall, were used for this purpose. The variance was investigated as a function of the number of independent scatterer realizations, used to obtain an estimate. Such signals may be acquired by imaging moving scatterers, e.g., blood or contrast agent. Alternatively they can be obtained by using different nonoverlapping beams from a sector/linear scan. The results show only minor differences between the two implementations with respect to stability. This means that the algorithm can be readily implemented in the time domain for real-time applications. The standard deviation of arrival time and amplitude fluctuation estimates decrease, when the number of independent signals increase. Using only one signal for estimation produces a relatively high standard deviation, but an iterative transmit-beam aberration correction scheme still converges to a properly corrected focus using between one and four transmit-beam iterations for the investigated aberrators. [DOI: 10.1121/1.2382748]

PACS number(s): 43.58.Ry, 43.20.Fn, 43.60.Cg, 43.60.Tj, 43.80.Vj [TDM] Pages: 286–297

I. INTRODUCTION

Aberration estimation in medical ultrasound imaging is dependent on the backscatter. The scatterers' may be modeled as randomly distributed in space, and such a distribution changes for each direction of the ultrasound beam due to the changes in scatterer type and position.

In Ref. 1, a method for estimating arrival time and amplitude fluctuations using backscatter from random scatterers was developed. The method averaged cross spectrums from independent realizations of the backscatter. In the present article, a variance analysis of estimated arrival time and amplitude fluctuations is performed in order to quantify the number of independent receive signals needed to obtain an adequate and stable estimate. This information is important when using a method which is dependent upon several receive signals for aberration estimation, as this will degrade the frame rate of the imaging system and may make the method impractical for real-time applications.

In addition, a time-domain implementation of the aberration estimation algorithm presented in Ref. 1 is introduced. This algorithm was earlier developed in the frequency domain. Necessary assumptions made in the time-domain implementation make the algorithm different from the frequency-domain implementation, and a subgoal of the presented work is to study how these differences affect the variance in the obtained estimates. Having the possibility to choose between the time domain or frequency domain, may be of importance regarding the potential for creating fast real-time processing hardware or software solutions of the algorithm.

In a practical situation, statistically independent signals can be obtained by scanning a region containing moving scatterers, e.g., the heart or a large blood vessel. Given the assumption that all blood cells within the beam volume have moved significantly between each transmit signal, the different received signals can be assumed to be statistically independent. Ultrasound contrast agents can be used to enhance such a signal.

Alternatively, it is possible to use signals from different regions in a sector scan.[2–4] By choosing beams which do not overlap, each received signal can be assumed to be independent. If the beams are overlapping. e.g., if neighboring beams are used, the signals will not be totally independent, and some correlation will exist between them.

In Ref. 2, a sum of absolute differences (SAD) method was developed for estimating relative phase errors between adjacent array elements. The variance of the estimator was investigated by using signals from five different nonadjacent scan angles, and two different focusing depths, in a scattering phantom. The obtained estimates were used in an iterative adaptive aberration correction scheme, and the variance of the estimates decreased sharply between the first and second iterations. Only smaller changes were observed when more iterations were performed.

Walker and Trahey[5,6] derived the Cramér-Rao lower bound for correlation based time-delay estimation from speckle signals. They investigated the limit as a function of several parameters such as correlation coefficient, window length, signal-to-noise ratio (SNR), bandwidth, and frequency. Their results showed that the lower bound for the variance of a time-delay estimate increased rapidly with a reduced correlation coefficient. They also showed that a SNR greater than 15 dB had little effect on the error of correlation-based phase estimates.

Ng *et al.*[7] compared the neighbor correlation method introduced by Flax and O'Donnell[8] with a partial array reference correction method, and showed that the latter method was more stable in the presence of noise. Overall the partial array reference correction method performed better than neighbor correlation and had a lower variance in the estimates as a function of SNR.

Rigby *et al.*[3] performed *in vivo* adaptive aberration correction with a time-delay filter using a 1.75-dimensional (1.75D) array. In order to reduce image artifacts due to strong off-axis scatterers, they averaged time-delay estimates over five transmit beams in a linear array scan. The beam sum channel-correlation method used for estimating time delays proved to adaptively converge to a stable estimate, producing improved image quality in most of the investigated subjects.

Lacefield and Waag[9] investigated estimates of the coherence using a 2D array. They found that the coherence of data acquired on neighboring elements was not changed by time-delay correction of transmit and receive beams. On the other hand, time-delay correction improved the coherence in signals received over larger element separation, supporting the use of several array elements for correlation estimation. Coherence is an important measure of the accuracy of correlation-based time-delay estimation methods. It is directly linked to the lower bound of the variance of such an estimate,[10] and to the variance of the magnitude and phase of the cross spectrum.[11]

Varslot *et al.*[4] estimated cross spectra using independent signals obtained from different focal points inside the isoplanatic patch with a 2D array. The purpose was to use the estimated frequency-dependent phase and magnitude of the cross spectrum for aberration correction. Increasing the number of independent signals from 13 to 75 gave a significant improvement of the spectral estimates. It reduced the 95% confidence intervals, thus improving the stability of the estimate. Their results also showed that the major part of the frequency-dependent phase constituted a pure time delay, corroborating the results from simulations performed by Måsøy *et al.*[12]

Viola and Walker[13] investigated eight commonly used time-delay estimators employed in medical ultrasound, and compared their variance to the Cramér-Rao lower bound for the same parameters as in Refs. 5 and 6. Their simulations showed that correlation methods outperformed all other methods with respect to stability (variance).

Iteration of transmit-beam aberration correction (transmit-beam iteration) was studied in Ref. 14. Here, the same estimation method as in Ref. 1 was investigated. Results showed that in using a sufficient number (20) of independent receive signals in order to estimate the cross spectrum, the transmit-beam iteration process converged for eight different aberrators representing cases of weak and strong aberration. After convergence, the quality of corrected focus beam profiles was close to the correction with estimates obtained from a point source.

Employing a transmit-beam iteration process using 20 receive signals for correction of each transmit beam is not feasible for real-time imaging. It is therefore important to investigate the number of independent signals necessary to obtain a stable and adequate estimate. This will then enable a transmit-beam iteration process to converge. The manner in which the stability affects the quality of the converged aberration corrected transmit-beam profile, is also important.

In the present article a variance analysis of estimated arrival time and amplitude fluctuations is performed in order to quantify the number of independent receive signals needed to obtain an adequate and stable estimate. As stated earlier, this information is important when using a method which is dependent upon several transmit and receive signals. In addition, transmit-beam iteration is also necessary. This is time consuming and will lower the frame rate in a real-time imaging system. High frame rate is very important for imaging moving objects like the heart. When imaging stationary objects like the liver, a lower frame rate is acceptable.

A stable estimate can be defined as an estimate that provides some correction of the retransmitted aberration-corrected beam, compared to the aberrated beam. If iterative transmit-beam aberration correction is to be used, the first corrected transmit beam needs to provide some correction (compared to the aberrated beam) for this process to converge. If the variance is high, the initial correction could in some cases increase the aberration, making the correction scheme unstable.

The variance analysis is performed by simulating scattering of ultrasound waves from uniformly distributed point scatterers in space, and with a Gaussian intensity distribution. Absorption or electronic noise is not included in the simulations.

II. THEORY

In Ref. 1 the modified beam former output (MBFO) method was developed for estimating arrival time and amplitude fluctuations from random signals.

The MBFO estimator was developed in the frequency domain. It was based on calculating the cross spectrum between all elements on the receiving array and a modified beam-former output, in order to determine the time delay and amplitude of the received signal. The beam-former output b of the received signal is simply defined as

$$b = \sum_{n=1}^{N} y_n, \tag{1}$$

where y_n is the received signal on element n, and N represents the total number of elements on the array. Frequency dependence has been dropped for notational convenience.

In general, the MBFO estimation can be performed for each frequency component of the signal, thus determining the specific aberration of each frequency. In Ref. 12, it was shown that, if correctly estimated, a time delay and amplitude for each array element produced approximately ideal aberration correction of the retransmitted beam. The MBFO algorithm was thus only implemented at the center angular frequency of the received signal, in order to determine a time delay and amplitude in the received signal for each array element.

When calculating the MBFO, correlation information between all elements on the receiving array is used. The correlation information is weighted with the coherence of the cross spectrum squared, a measure of the variance in the estimate of the cross spectrum. An important issue is to obtain a reliable estimate of the cross spectrum. This was done, as explained in the Introduction, by averaging the estimated cross spectrum from several realizations of the scattering region; that is, from several statistically independent receive signals.

The current aim is to investigate how the stability of the estimator is affected by varying the number of independent realizations.

A. Time-domain implementation of the MBFO algorithm

The MBFO algorithm is based on cross-correlation estimation. In addition, a model for the received signal is used. In Ref. 1, the received signal y was written in the frequency domain as

$$y(\mathbf{r}_a;\omega) = s(\mathbf{r}_a;\omega) f(\mathbf{r}_a;\omega), \tag{2}$$

where $\mathbf{r}_a$ is the array coordinate, and $f(\mathbf{r}_a;\omega)$ represents the unaberrated receive signal. The function $f(\mathbf{r}_a;\omega)$ thus contains transmit aberration, but not receive aberration. The function $s(\mathbf{r}_a;\omega)$ is the frequency response of a filter producing distortion of the amplitude and phase of the signal as it propagates through the heterogeneous medium. This filter is denoted the *generalized frequency-dependent screen*. Note that due to the physics of aberration $s(\mathbf{r}_a;\omega)$ must be a causal filter.

In Eq. (2), aberration of the received signal on a specific array element is assumed to be independent of the spatial position of the scatterer. This is a general assumption for aberration correction as it is performed at the array surface as a filtering process, and is denoted in this paper as *scatterer independent aberration*. In this situation, the generalized frequency-dependent screen is denoted the *scatterer independent screen*.

In Ref. 1, a thorough discussion of the generalized frequency-dependent screen is given, as well as the basis and limitations of the receive scatterer independent aberration assumption.

Transforming Eq. (2) to the time domain yields

$$y(\mathbf{r}_a,t) = \int s(\mathbf{r}_a,t-\xi) f(\mathbf{r}_a,\xi)\mathrm{d}\xi, \tag{3}$$

where ξ is a convolution variable and the integration is taken over all time. This type of model can be solved with a blind system identification method,[15] but a different approach is chosen here.

Based on the investigations in Ref. 12 described above, the scatterer independent screen filter is modeled as a simple time delay and amplitude screen at the array. The scatterer independent screen is then approximated as

$$s(\mathbf{r}_a,t) \approx a(\mathbf{r}_a)\delta[t-\tau(\mathbf{r}_a)], \tag{4}$$

where $a(\mathbf{r}_a)$ is an amplitude and $\tau(\mathbf{r}_a)$ is a time delay, both a function of the array coordinate. This approximation introduces a loss of generality compared to the frequency implementation of the MBFO algorithm as described in Ref. 1.

Using Eq. (4) in (3) gives

$$y(\mathbf{r}_a,t) = a(\mathbf{r}_a) f[\mathbf{r}_a, t-\tau(\mathbf{r}_a)]. \tag{5}$$

It is convenient to introduce the analytic signal defined as

$$\hat{y}(\mathbf{r}_a,t) = y(\mathbf{r}_a,t) + i\mathcal{H}\{y(\mathbf{r}_a,t)\}, \tag{6}$$

where $\mathcal{H}\{\cdot\}$ denotes the Hilbert transform. For a band limited signal, the analytic signal can also be written as [Ref. 16 (p. AII.40)]

$$\hat{y}(\mathbf{r}_a,t) = y_e(\mathbf{r}_a,t)e^{i\omega_0 t}, \tag{7}$$

where $y_e(\mathbf{r}_a,t)$ represents the complex envelope, and ω_0 is the center angular frequency.

Introducing the narrow-band approximation,

$$\hat{y}(\mathbf{r}_a,t-\tau) \approx y_e(\mathbf{r}_a,t)e^{i\omega_0(t-\tau)} = \hat{y}(\mathbf{r}_a,t)e^{-i\omega_0\tau}, \tag{8}$$

a time delay τ of the envelope is assumed to be negligible compared to the same time delay of the carrier signal. Using this approximation in Eq. (5) gives

$$\hat{y}_p(t) = a_p\hat{f}_p(t-\tau_p) = a_p e^{-i\omega_0\tau_p}\hat{f}_p(t) \equiv s_p\hat{f}_p(t), \tag{9}$$

where the signals are written in spatial discrete form with p indicating the element number on the array. Here, s_p is the complex equivalence of the simple model of the scatterer independent screen in Eq. (4).

The signal model in the time domain as given by Eq. (9), is analogous to the frequency model in Eq. (2). The derivation of the estimation method in the time domain thus equals the derivation in the frequency domain as given in Ref. 1. It is briefly revised here for the purpose of continuity.

The definition of the spatial cross correlation between signals on elements p and n is

$$R_{pn} = \mathrm{E}[\hat{y}_p \quad \hat{y}_n^*], \tag{10}$$

where $\mathrm{E}[\cdot]$ denotes an expectation operator, while time dependence has been dropped for notational convenience. Inserting Eq. (9) gives

$$R_{pn} = s_p s_n^* F_{pn}, \tag{11}$$

where $F_{pn} = \mathrm{E}[\hat{f}_p \quad \hat{f}_n^*]$.

Solving for s_p gives

$$s_p = \frac{R_{pn}}{F_{pn}}\frac{1}{s_n^*} \equiv \frac{R_{pn}}{|F_{pn}|}\frac{1}{s_n^*}. \tag{12}$$

Now, F_{pn} can produce a phase due to refraction or aberration of the transmitted beam. Assuming that $F_{pn}=|F_{pn}|$ introduces an error in the estimates as discussed in Ref. 14.

In order to use all correlation information in the signal, a weighted mean estimate of s_p is introduced as[1]

$$\tilde{s}_p = \sum_{n=1}^{N} W_{pn}\frac{R_{pn}}{|F_{pn}|}\frac{1}{\tilde{s}_n^*}, \tag{13}$$

where W_{pn} is a weight function. The average is taken over all elements N on the array.

In Ref. 1, $|F_{pn}|$ was assumed to be a known variable and was calculated using the van Cittert-Zernike theorem for an incoherent homogeneous medium. By defining the weight function W_{pn} properly, this is not necessary and $|F_{pn}|$ is here assumed to be an unknown variable.

The correlation coefficient at zero lag is defined as

$$\rho_{pn} = R_{pn} \Big/ \sqrt{R_{pp}R_{nn}}. \tag{14}$$

By defining W_{pn} as

$$W_{pn} = |\rho_{pn}|^2 \Bigg/ \sum_{n=1}^{N} |\rho_{pn}|^2 \equiv |\rho_{pn}|^2/C_p, \tag{15}$$

the weight factor suppresses areas of low correlation, which contain little information of the phase and amplitude estimate. Krishnan *et al.*[17] showed that the variance of the phase estimate between two identical signals with a time shift, is inversely proportional to the magnitude of the correlation coefficient squared. A high correlation coefficient means a low variance in the phase estimate.

Taking the magnitude of the correlation coefficient and inserting the signal model from Eq. (9) gives

$$\begin{aligned}|\rho_{pn}| &= \frac{|R_{pn}|}{\sqrt{R_{pp}R_{nn}}} \\ &= \frac{|s_p s_n^* F_{pn}|}{\sqrt{|s_p|^2|s_n|^2 F_{pp}F_{nn}}} \\ &= \frac{|e^{i\omega_0(\tau_n-\tau_p)}F_{pn}|}{\sqrt{F_{pp}F_{nn}}} = \frac{|F_{pn}|}{\sqrt{F_{pp}F_{nn}}}.\end{aligned} \tag{16}$$

Assuming F_{pn} to be scaled such that $F_{pp}=1$, and inserting Eqs. (15) and (16) into Eq. (13) gives

$$\tilde{s}_p = \frac{1}{C_p}\sum_{n=1}^{N} |\rho_{pn}|R_{pn}\frac{1}{\tilde{s}_n^*}. \tag{17}$$

By choosing the weight factor W_{pn} as in Eq. (15), no assumptions need be made on $|F_{pn}|$.

Using the ergodic hypothesis,[11] an unbiased estimate of the cross correlation is defined as

$$\tilde{R}_{pn} = \frac{1}{K}\sum_k \frac{1}{T}\int_T \hat{y}_{kp}(t)\hat{y}_{kn}^*(t)\mathrm{d}t, \tag{18}$$

where k denotes different realizations of the received signal $\hat{y}$, and K the total number of realizations. Note that the integral part of Eq. (18) represents the general cross correlation calculated at zero lag. This is sufficient for determining a time delay in the signal due to the introduction of the complex envelope and the narrow band approximation.

Using the estimate of the cross correlation $\tilde{R}_{pn}$, the final expression for $\tilde{s}_p$ is

$$\tilde{s}_p = \frac{1}{C_p}\sum_{n=1}^{N} |\tilde{\rho}_{pn}|\tilde{R}_{pn}\frac{1}{\tilde{s}_n^*}, \tag{19}$$

where $\tilde{\rho}_{pn}$ is an estimate of ρ_{pn} based on $\tilde{R}_{pn}$.

Equation (19) is an implicit equation and is solved with an iteration scheme as outlined in Ref. 1. Zero time delay and unity amplitude across the array are used as initial values for the iteration scheme. Estimates of arrival time and amplitude fluctuations are then calculated from $\tilde{s}_p$ as the magnitude and time delay of the phase, as defined in Eq. (9).

B. Comparison of frequency-domain and time-domain implementations

Using Parseval's theorem, the integral in Eq. (18) gives

$$\int_{-\infty}^{\infty} \hat{y}_p(t)\hat{y}_n^*(t)\mathrm{d}t = \frac{1}{2\pi}\int_{-\infty}^{\infty} \hat{y}_p(\omega)\hat{y}_n^*(\omega)\mathrm{d}\omega. \tag{20}$$

The index k has now been dropped for notational convenience.

The receive signal can be assumed to be infinite in time (e.g., windowed part of an infinite signal). Dropping the integration period, and using the result from Parseval's theorem in Eq. (18) yields

$$\begin{aligned}\tilde{R}_{pn} &= \frac{1}{K}\sum_k \frac{1}{2\pi}\int_{-\infty}^{\infty} \hat{y}_{kp}(\omega)\hat{y}_{kn}^*(\omega)d\omega \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{1}{K}\sum_k \hat{y}_{kp}(\omega)\hat{y}_{kn}^*(\omega)d\omega = \frac{1}{2\pi}\int_{-\infty}^{\infty}\tilde{R}_{pn}(\omega)d\omega.\end{aligned} \tag{21}$$

The time-domain implementation uses the integral of the cross-spectrum estimate over all frequency components.

Equation (21) shows that the time-domain implementation of the MBFO estimation algorithm, produces a weighted estimate of both the phase and amplitude between the signals on elements p and n over all the frequency components. This is different from the frequency-domain implementation which only uses the phase and amplitude at the center frequency.[1]

As in Ref. 1, by inserting Eq. (18) into Eq. (19), $\tilde{s}_p$ can be shown to satisfy the relation

$$\tilde{s}_p = \frac{1}{C_p}\frac{1}{K}\sum_k \frac{1}{T}\int_T \hat{y}_{kp}(t)\hat{b}_{kp}^*(t)dt, \tag{22}$$

with

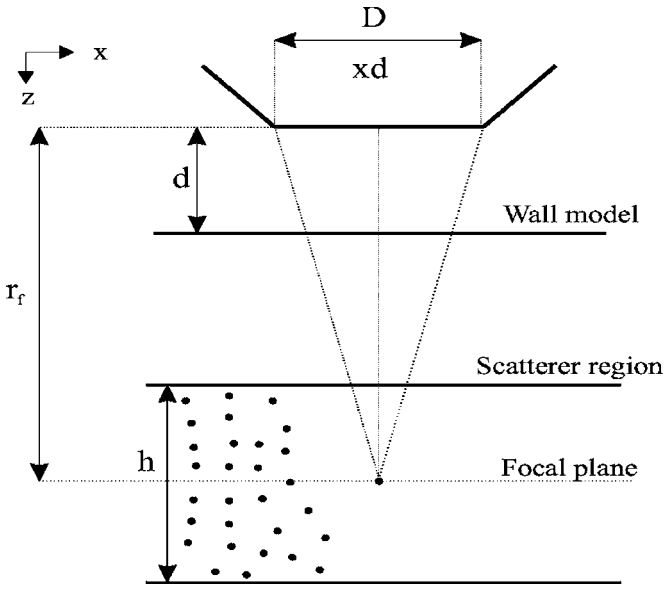


FIG. 1. Simulation setup.

$$\hat{b}_{kp}(t) = \sum_{n=1}^{N} \hat{y}_{kn}(t)|\tilde{\rho}_{pn}|\frac{1}{\tilde{s}_n}. \tag{23}$$

This relation is the time domain MBFO.

By using the magnitude of the correlation coefficient as a weight term in the beam-former output, the estimation process inherently performs an efficient subaperture correlation process. The magnitude of the correlation coefficient is unity for $n=p$, and falls quickly in value as the distance between n and p increases. The weight function $|\rho_{pn}|$ thus acts as a sliding window (following the element p where the estimation occurs) selecting a suitable subaperture for the beam-former output, from a correlation perspective.

III. SIMULATIONS

The same simulation setup as in Ref. 1 was used (see Fig. 1). An array of size D was placed onto a body wall model of thickness d. The signal was propagated with an angular spectrum operator[12] through the body wall model and to the scattering region of depth h, centered around the focal plane of the array in the range direction. Here the signal was scattered according to the Born approximation and propagated back to the array through the body wall model.

Eight body wall models (also denoted aberrators), emulating the human abdominal wall, were used. These were generated using a series of distributed time-delay screens. Two categories, representing weak and strong aberration, respectively, with four different body wall models in each category, were created. Details of the body wall modeling, properties of the wall models, and a discussion of their validity can be found in Ref. 12. The names of the body wall models are the same as in Ref. 12, and are given as $\{w2, w4, w6, w8\}$ and $\{s2, s4, s6, s8\}$. Here, the letter corresponds to the respective category, i.e., weak or strong. The number refers to the number of time-delay screens used in the aberrator.

The scatterers were modeled as uniformly distributed point scatterers in the x and z direction, and had a Gaussian distributed reflection strength. In order to determine the variance of the arrival time and amplitude fluctuations estimates, 400 realizations of the scattering region were generated for the $w6$ and $s6$ aberrator. A simulation, as described above, was then performed for all of the scatterer realizations generating 400 statistically independent receive signals. For the rest of the aberrators, this number was reduced to 100 realizations in order to save computation time. The reason for this will be duly described in Sec. V.

In order to obtain reference values for arrival time and amplitude fluctuations for the body wall models, one way point source simulations were performed: A point source was simulated in the focus $\mathbf{r}_f$ of the array. A pulse, equal to the transmit pulse from the array, was emitted from $\mathbf{r}_f$ and propagated through the aberrators to the array.

A. Simulation parameters and data processing

The simulation setup is the same as in Ref. 1 and is reiterated here for the convenience of the reader.

The simulations were implemented in MATLAB. The simulation area was 10.24 cm in the lateral direction (x direction in Fig. 1) with a resolution of 0.2 mm. To avoid reflections at the edges of the spatial fast Fourier transform (FFT) region (x direction), the signal was tapered to zero over 2.54 cm at each edge with a raised cosine window. The tapering was performed for a sufficient number of propagation steps in order to keep the noise level low.

The sampling frequency was 35.1 MHz providing a time window of 58.3 μs. The center frequency of the pulse was 2.5 MHz with a −6 dB bandwidth of 1.5 MHz. The transmitted pulse was filtered with a 100% fractional bandwidth bandpass filter at −6 dB. An array aperture size of 20 mm with pointlike elements was chosen. The focal depth $\mathbf{r}_f$ of the array was 60 mm. The thickness of the body walls were 20 mm. The medium through which the signals were propagated had a speed of sound equal to that of water (1523 m/s).

The scattering area had a range depth h of 30.5 mm (time window of 20 μs), 15.25 mm to each side of the focal plane (see Fig. 1). This range was chosen to be well inside the −3 dB focal depth of the array.[1] The scatterer density was approximately 1600 scatterers per square centimeter.

For the simulations using the scatterer region, dynamic focusing was applied on the receive signals prior to estimation of arrival time and amplitude fluctuations. Geometric focusing was removed from the point source simulations prior to processing of results.

For the time-domain implementation, the cross correlation was calculated according to Eq. (18). Trapezoid integration was used to evaluate the integral.

The cross spectrum for an element p was estimated by averaging the spectrum from overlapping segments of the signal (Welch's method). The element signal was separated into 64 segments (0.9 μs) with 50% overlap. Each segment was weighted with a Hanning window before calculation of the FFT.

A wave-front tracking method[12] was used to estimate arrival time fluctuations of the wave front from the point

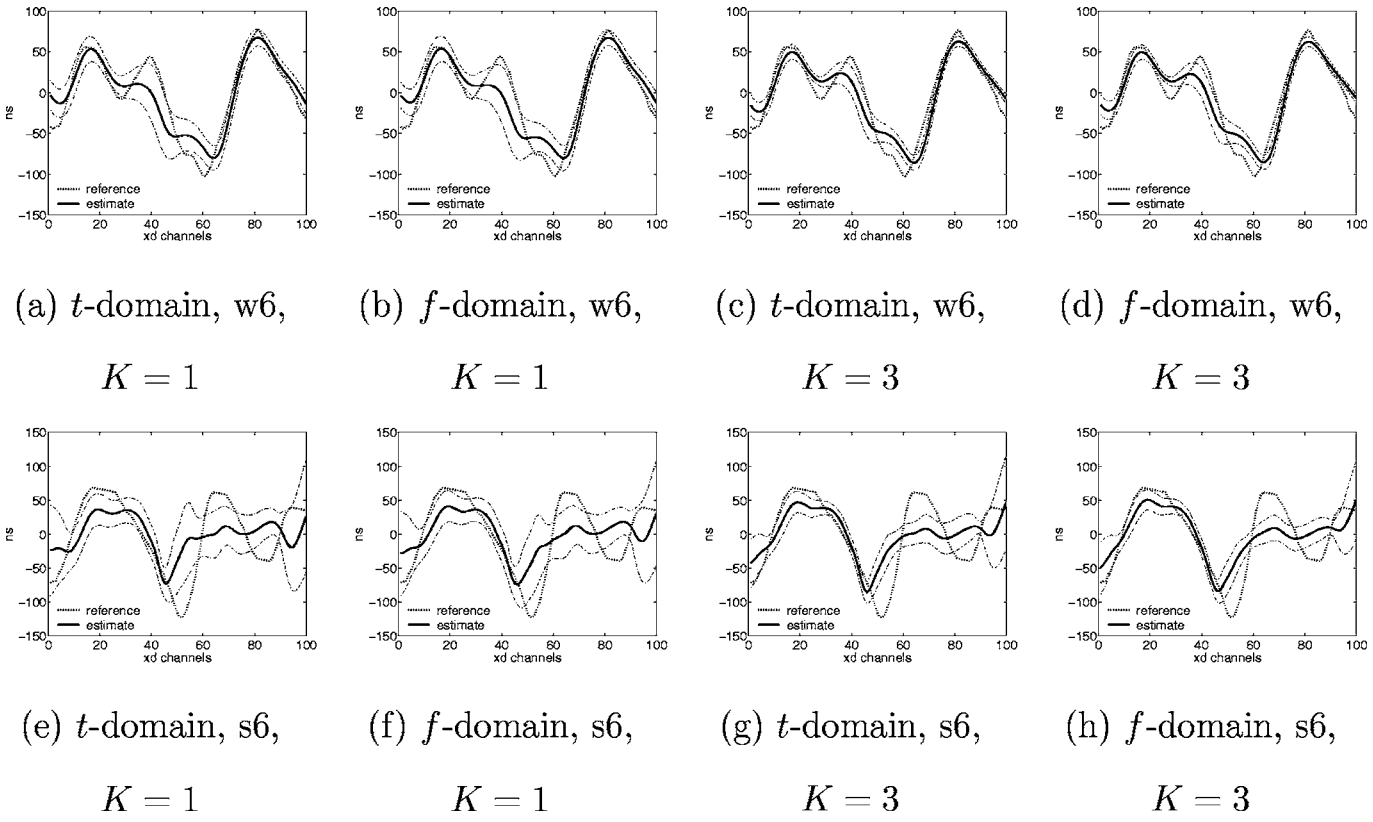


FIG. 2. Mean arrival time-delay fluctuation estimates with standard deviation for the *w*6 and *s*6 aberrators. Here, *K* signals per estimate are used. The reference curve denotes the arrival time fluctuations estimate from the point source. The solid line is the mean value of the estimate, and the dashed-dotted line shows the standard deviation of the estimates.

source simulations. This method provided an accurate description of the time-delay variation of the received wave front. Arrival amplitude fluctuations for each array element from the point source simulations were calculated as the arithmetic mean of the magnitude of the frequency spectrum over a band of frequencies ranging from 2–3 MHz. This band was chosen empirically.

Beam profiles in the focal plane of the array were calculated as the rms value in time of each spatial position. These profiles were used for the visual evaluation of the effect of the different aberration correction filters.

To analyze the variance of estimated arrival time and amplitude fluctuations, the standard deviation σ of the estimate for an array element p was calculated as

$$\sigma_p = \sqrt{\frac{1}{M-1}\sum_{m=1}^{M}(x_{mp}-\bar{x}_p)^2}, \tag{24}$$

where x denotes either an arrival time or amplitude fluctuation estimate, $\bar{x}$ is the mean value of the estimate, and M is the total number of estimates used for calculating the standard deviation. In all the simulations, 20 estimates ($M=20$) were used to calculate the mean and standard deviation. Thus when the number of independent receive signals K varied, M remained constant equal to twenty. The simulations were carried out for $K=\{1,3,5,10,15,20\}$ for the *w*6 and *s*6 aberrators, and for $K=\{1,3,5\}$ for the rest of the aberrators.

In addition, the average value of the standard deviation (average standard deviation) across the array was also calculated. This is defined as

$$\bar{\sigma} = \frac{1}{N}\sum_{p=1}^{N}\sigma_p, \tag{25}$$

where N is the total number of array elements.

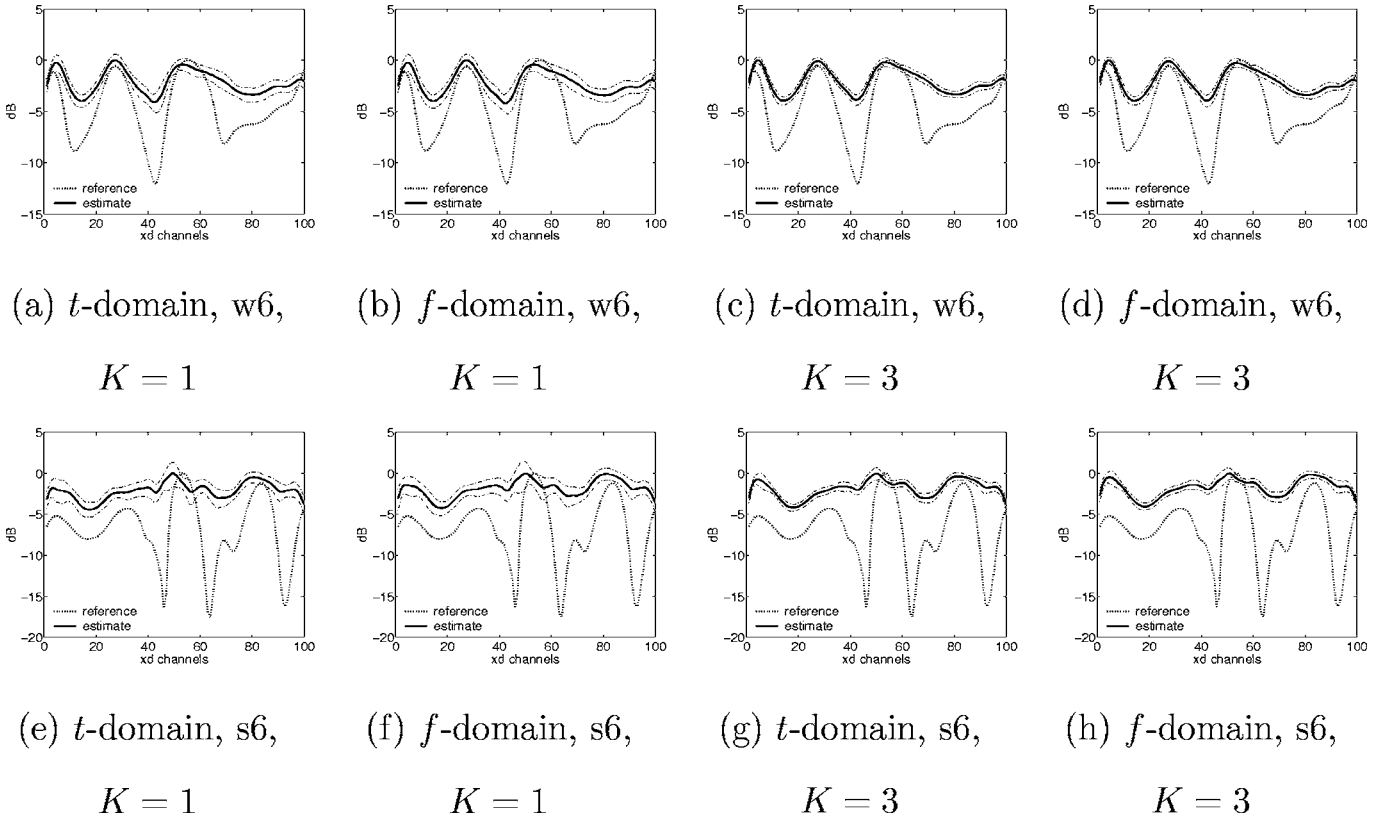
IV. RESULTS

For estimates of arrival time and amplitude fluctuations, and beam profiles, only selected data for the *w*6 and *s*6 aberrator are presented for $K=1$ and $K=3$, where K denotes the number of independent receive signals. These data are representative of all aberrators in the simulation study.

Average standard deviation curves, as a function K, are presented for all aberrators.

In all the figures presented here, *t*-domain denotes the time-domain implementation and *f*-domain represents the frequency-domain implementation.

Figure 2 shows mean arrival time fluctuation estimates with standard deviation for the *w*6 and *s*6 aberrators. The curves are plotted both for the time- and frequency-domain implementation of the MBFO algorithm. The mean values of estimates in the time and frequency domain are quite similar for both aberrators. There is clearly a higher standard deviation for the *s*6 aberrator compared to *w*6. The standard deviation is reduced when the number of independent signals is increased from $K=1$ to $K=3$. There is also a minor improve-



(a) t-domain, w6, $K = 1$ (b) f-domain, w6, $K = 1$ (c) t-domain, w6, $K = 3$ (d) f-domain, w6, $K = 3$

(e) t-domain, s6, $K = 1$ (f) f-domain, s6, $K = 1$ (g) t-domain, s6, $K = 3$ (h) f-domain, s6, $K = 3$

FIG. 3. Mean arrival amplitude fluctuation estimates with standard deviation for the $w6$ and $s6$ aberrators. Same notation as in Fig. 2.

ment in the mean estimate compared to the reference when K is increased. There are very small differences in the standard deviation between the time and frequency implementation.

Mean arrival amplitude fluctuation estimates with standard deviation are shown in Fig. 3. Here the standard deviation also decreases when K increases, but this does not seem to improve the mean value of the estimates. The mean estimate is poor compared to the reference. Here also, there are marginal differences between the time-domain and frequency-domain implementations. The above comments are valid for all aberrators investigated here.

Figure 4 shows the mean value of aberration corrected beam profiles, with standard deviation, in the focal plane of the array. As for the arrival time fluctuation estimates, there is some improvement in the corrected beam profiles for the $w6$ aberrator when K is increased. This is mainly due to the small improvement in the mean value of the arrival time fluctuation estimates. There are also minor improvements for the $s6$ aberrator. Note that for the $s6$ aberrator and $K=1$, the side-lobe level of the standard deviation curve is higher than the uncorrected beam in some parts of the profile.

Figure 5 presents the average standard deviation as a function of the number of independent signals for the $w6$ and $s6$ aberrators. The results for both arrival time and amplitude fluctuation estimates are given. The average standard deviation for the amplitude estimates is normalized to the mean value of the amplitude estimate for the point source reference. For the arrival time fluctuation estimates, the difference in average standard deviation between the $w6$ and $s6$ aberrator is approximately a factor of two. For both aberrators, there is little difference between the time and frequency implementation. In the case of arrival amplitude fluctuation estimates, there is also a substantial difference in average standard deviation between the aberrators. There is almost no difference between the implementations. In general (and as expected) for both arrival time and amplitude fluctuation estimates, the average standard deviation decreases with increasing K.

Figures 6 and 7 present the average standard deviation for the rest of the aberrators. These have only been simulated for $K=\{1,3,5\}$, since this is where the major changes occur. The general trend is that the average standard deviation decreases with increasing K, and that there exists only smaller differences between the two implementations. For the $s8$ aberrator, there are some differences visible in the average standard deviation for the arrival time fluctuation estimates.

V. DISCUSSION

As described in Sec. III, 400 independent receive signals were generated for the $w6$ and $s6$ aberrator, and only 100 for the rest. This was done because the results from the $w6$ and $s6$ aberrators showed that the major changes occurred when changing K from 1 to 5. In addition, performing a full simulation in order to generate one independent receive signal, needed approximately 19 min of computing on a Dell PC with two Intel Xeon 2.8 MHz EMS64 processors. Thus 400 realizations demanded more than five days of computing; more than 42 days for all eight aberrators. Therefore, it was deemed sufficient to only generate 100 realizations for the rest of the aberrators to capture the most important results.

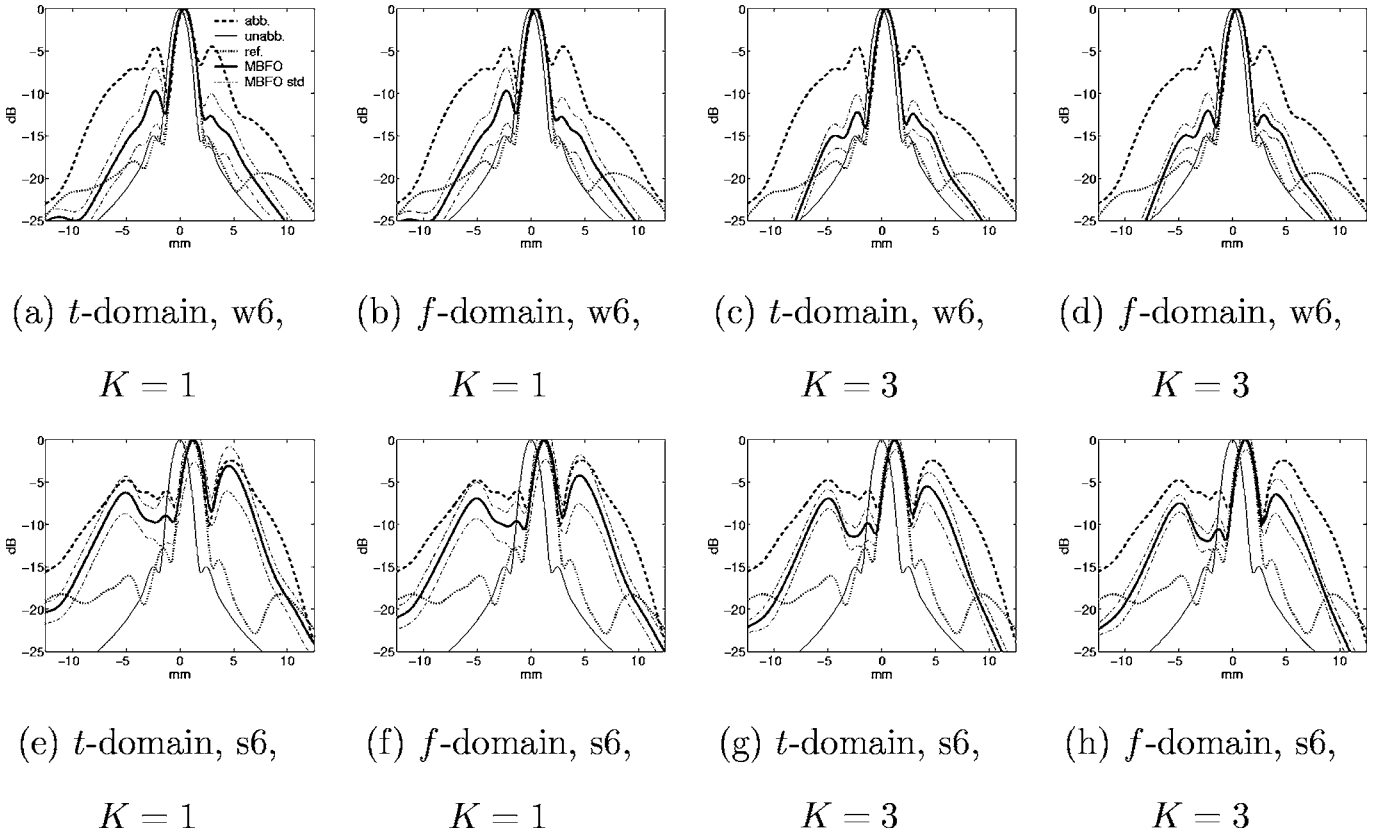


FIG. 4. Mean value of beam profiles in the focal plane of the array, with standard deviation, for the *w*6 and *s*6 aberrators. The vertical axis displays energy in decibels, and the horizontal axis displays the focal plane in millimeters. The curve *abb.* denotes the uncorrected transmit beam, *unabb.* denotes the unaberrated transmit beam, *ref.* is the transmit beam corrected with the point source reference, *MBFO* shows the mean value of the beam profiles corrected with the MBFO estimator, and the dashed-dotted line *MBFO std* represents the standard deviation. The legend is the same for all figures as for (a).

The development of the correlation coefficient in Eq. (15) is new, and can also be shown for the coherence in the frequency domain (confer Ref. 1). In the previous development of the algorithm,[1,14] the van Cittert-Zernike theorem for an incoherent homogeneous medium was used to calculate the value of $|F_{pn}|$, defined in Eq. (11). The development presented here shows that this is not necessary and it also makes the algorithm more intuitive. The correlation coefficient (or coherence in the frequency domain) acts as an efficient subaperture selector, choosing the optimum subaperture (from a variance perspective) to be used in the correlation estimation process, suppressing estimates from elements with high uncertainty.

As seen from the results in the previous section, there are only small differences between the time-domain and frequency-domain implementations. In general, the mean value and standard deviation of arrival times and amplitude fluctuations are approximately equal for all investigated aberrators. As shown in Eq. (21), the time-domain implementation uses the integral of the cross spectrum over all frequency components, and the only difference in the implementations lies in the estimation of the cross correlation and the correlation coefficient (cross spectrum and coherence in the frequency domain, confer Ref. 1). This provides freedom of choice for real-time implementations strategies of the algorithm either in hardware or software.

For proper aberration correction, a 1.5D or 2D array is necessary. Performing a full FFT for all channels may prove to be computationally expensive, and having the alternative of choosing a time-domain implementation is then of importance. It is also possible to count the number of real multiplications as an indicator of the computational complexity of the algorithm, but this may be overcome by parallelization. How to parallelize the algorithm may vary depending on the implementation. Choosing a hardware or software implementation is also an issue which will influence the degree of parallelization, where e.g. Field Programmable Gated Arrays (FPGA's) may be used. Strategies for optimization of the algorithm is a research topic of its own and is not further pursued here.

As expected, the average value of the standard deviation across the array is reduced when more independent signals K are averaged in the estimation process. Increasing this number from $K=1$ to $K=3$, reduced the average standard deviation of arrival time fluctuation estimates by $42.1 \pm 3.8\%$ for the time-domain implementation, and for all aberrators. For the amplitude estimates the reduction is $39.1 \pm 3.7\%$. The reduction from $K=3$ to $K=5$ is approximately 20% for both arrival time and amplitude fluctuation estimates. This very closely matches a $1/\sqrt{K}$ reduction, which is 42.3% for K increasing from 1 to 3, and 22.5% for 3 to 5.

An important point is that the standard deviation varies quite considerably between the aberrators. For $K=1$, using the time-domain implementation, the variation in average standard deviation for arrival time fluctuations between the

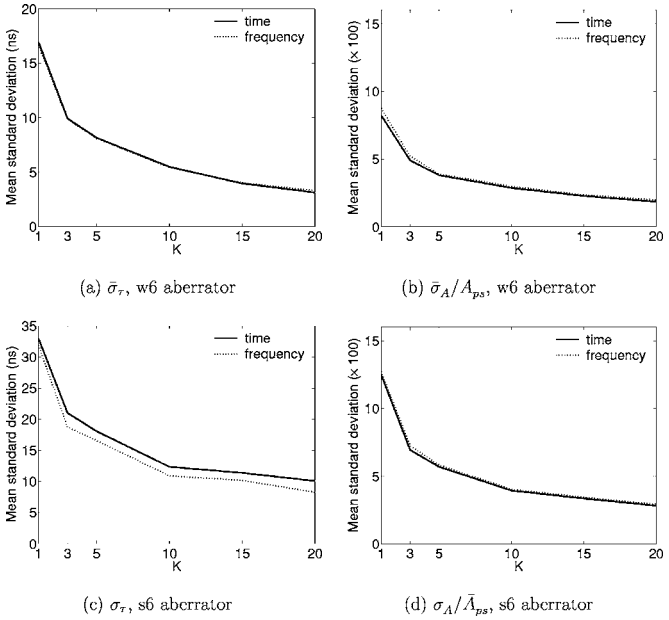


FIG. 5. Average standard deviation for arrival time ($\bar{\sigma}_\tau$) and amplitude ($\bar{\sigma}_A/\bar{A}_{\mathrm{ps}}$) fluctuations as a function of the number of independent signals (K) used in the estimation for the $w6$ and $s6$ aberrator. The average standard deviation for the amplitude is normalized to the mean value of the amplitude estimate ($\bar{A}_{\mathrm{ps}}$) from the point source simulation.

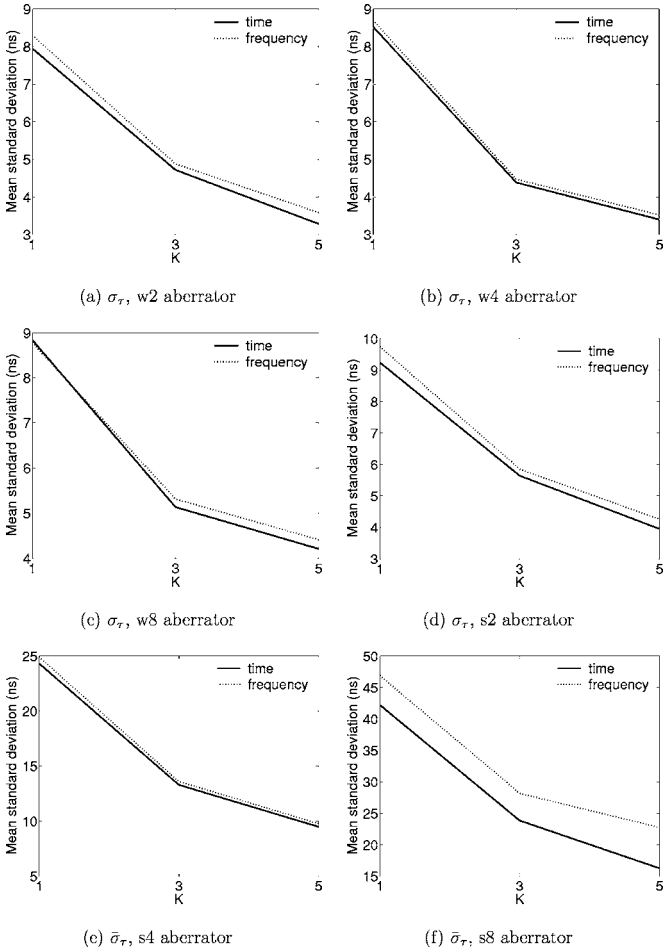


FIG. 6. Average standard deviation ($\bar{\sigma}_\tau$) for arrival time fluctuations as a function of the number of independent signals (K) used in the estimation for the $w2$, $w4$, $w8$, $s2$, $s4$, and $s8$ aberrators.

eight investigated aberrators ranged from approximately 8–43 ns. Also, there is a 42% difference in the average standard deviation for the amplitude fluctuations. These results show that the stability of the estimation is highly dependent on the aberrator. The aberration corrected beam profile for each estimate was therefore investigated for all aberrators. The general trend showed that all estimates produced an improvement of the corrected beam profile. This result was independent of K.

For the s6 aberrator, using $K=1$, many of the estimates yielded very little correction of the beam profile. This is clearly demonstrated by the standard deviation of the beam profiles in Fig. 4(e) and 4(f). If the transmit beam is not sufficiently corrected, a transmit-beam iteration process may not converge. Such a process was investigated for all the aberrators using $K=1$. Here, the transmit-beam iteration process was performed in the same manner as in Ref. 14, using five iterations. The simulations were only carried out with estimates of arrival time and amplitude fluctuations using the time-domain implementation, since there exists only minor differences between the two. The whole procedure was repeated 20 times for each aberrator in order to generate 20 different realizations of the transmit-beam iteration process. Two different scenarios were investigated

1. A new independent realization of the scattering region was generated for each iteration.
2. The same scattering region was used for each iteration.

Scenario 2 emulates a situation where the ultrasound beam is fired in the same direction and focused at the same spot for each transmit-beam iteration. Since the scattering region remained the same, the only change in scattering resulted from a change in the transmit beam due to the aberration correction. Note that for this scenario, a new scattering region was used for each of the 20 different realizations.

The quality, and convergence, of the transmit-beam iteration correction process was quantified by the focus quality parameter C introduced by Mallart and Fink.[18] This parameter has proven to be suitable for this purpose,[14] where $C=0$ describes total incoherence and $C=2/3$ maximum coherence for δ-correlated scatterers. The mean value (taken over the 20 different realizations) and standard deviation of C are plotted as a function of iteration number in Fig. 8, for the two scenarios and all aberrators. Iteration number 0 refers to the initial uncorrected transmit beam. There exist only minor differences between the two scenarios. Also, the figure shows that convergence is obtained after 1–3 iterations. This is the same result obtained in Ref. 14 using $K=20$. The quality of the convergence (the value of the C parameter) is very close to what is obtained using $K=20$. The results suggest that only minor improvements of the first corrected beam profile are necessary for a transmit-beam iteration process to converge rapidly with good quality, using the MBFO algorithm. Also, the results indicate that using one receive signal is sufficient. Finally, there seems to be no need for replacing the scatterers between each iteration. This can be explained by the fact that performing aberration correction changes the

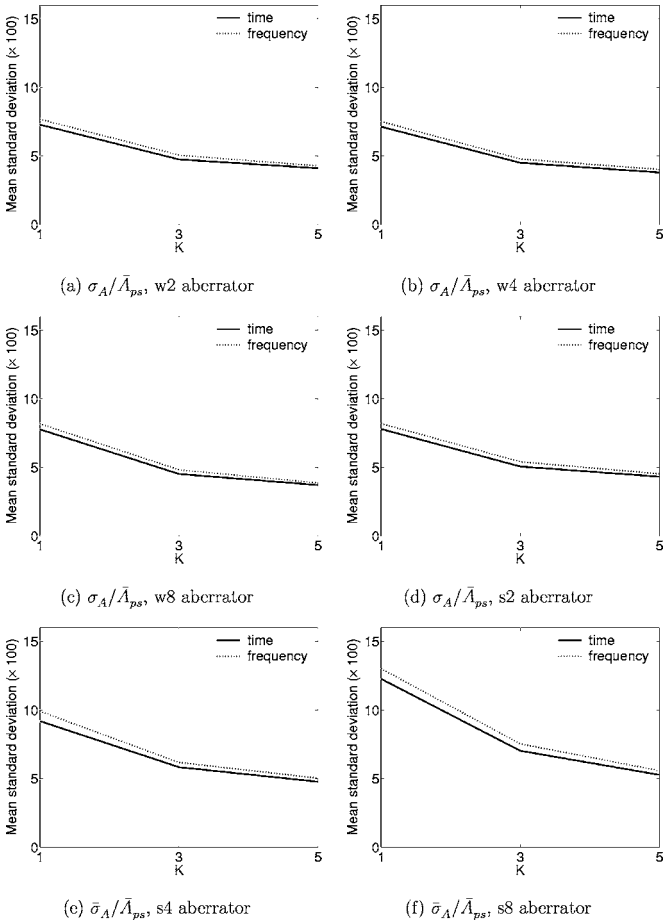


FIG. 7. Average standard deviation ($\bar{\sigma}_A/\bar{A}_{\mathrm{ps}}$) for amplitude estimates as a function of the number of independent signals (K) used in the estimation for the $w2$, $w4$, $w8$, $s2$, $s4$, and $s8$ aberrators. The average standard deviation is normalized to the mean value of the amplitude estimate ($\bar{A}_{\mathrm{ps}}$) from the point source simulation.

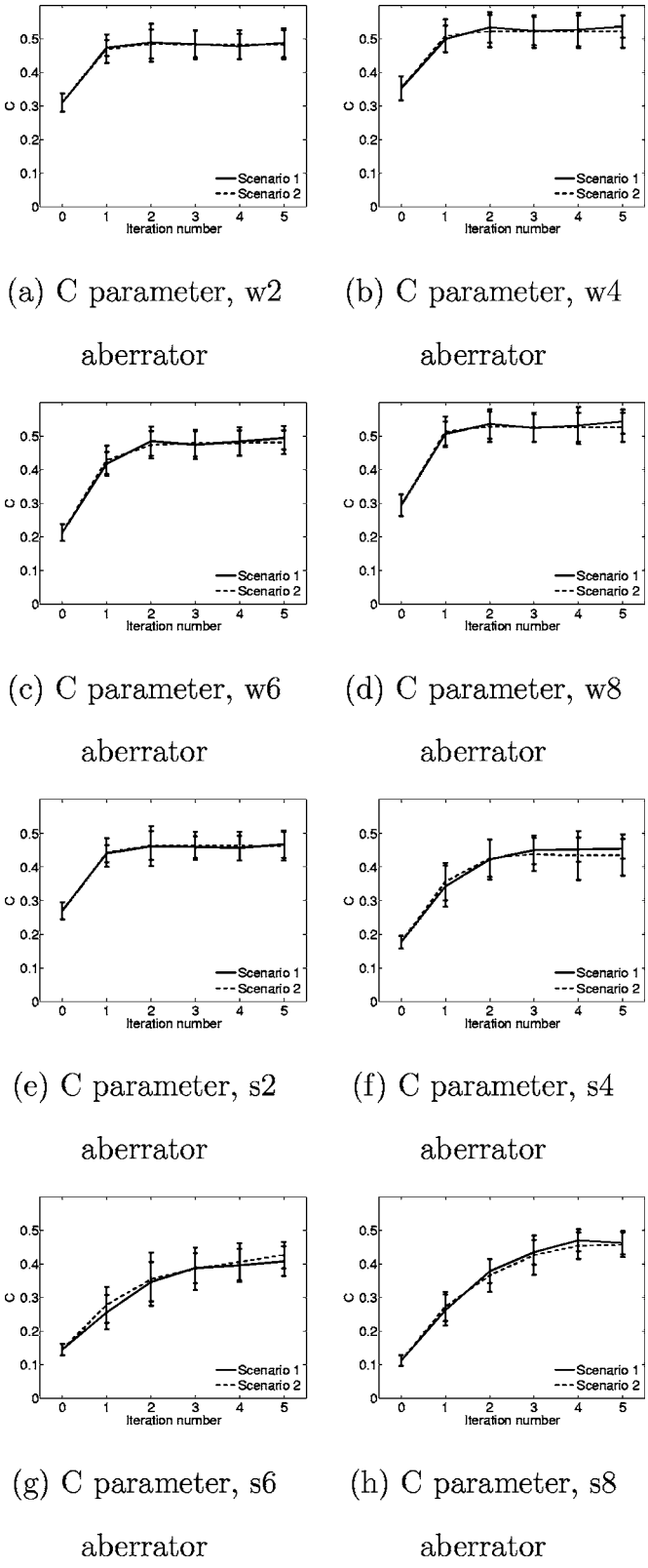


FIG. 8. Mean value of 20 estimates of the focus quality parameter C as a function of iteration number, using the time-domain implementation with $K=1$. The solid line represents scenario 1, and the dotted line scenario 2. The error bars show the standard deviation.

transmit beam. This again changes the region of the scatterers that are insonified, leading to a similar result as in scenario 1.

For the frequency-domain implementation, a Welch method with 50% overlap of segments was used to estimate the cross spectrum. Each segment was windowed with a Hanning window. The number of segments used was 64, which corresponds to a segment kernel length of 0.9 μs. The transmit signal was a 2.5 period pulse with 2.5 MHz center frequency, which gives a pulse length of 1 μs. The kernel length of each segment then corresponds to approximately 2.25λ, where λ represents the wavelength. Using a Welch method leads to a smoothing of the frequency spectrum which is determined by the kernel length. Even though the frequency-domain implementation of the MBFO algorithm was used at the center frequency of the signal, the magnitude and phase of this frequency has been mixed (averaged) with surrounding frequencies. Since a relatively small kernel length was used, the smoothing effect is limited to a narrow band around the center frequency used in the estimation.

The cross correlation in the time domain [Eq. (18)] yielded an average of the cross spectrum over all frequency components. Using the Welch method, as described above, leads to some smoothing for the frequency-domain implementation of the MBFO method. Still, with a narrow band signal, and a linearly varying phase with frequency, the obtained estimates with the time- and frequency-domain implementations should yield approximately the same result. In

Ref. 12, it was shown that using a time delay and amplitude aberration correction filter estimated from a point source, produced close to ideal aberration correction for the *w*6 and *s*6 aberrator. This is also true for the rest of the aberrators investigated here. Therefore, the phase of the aberration is adequately modeled with a pure time delay for the presented results, and both implementations produced the same results. If the phase of the aberration [as described by the generalized frequency-dependent screen in Eq. (2)] varies nonlinearly with frequency, the two implementations may produce different results due to differences in the averaging processes.

Absorption, electronic and thermal noise, and acoustic reverberations was not included in the simulations. A discussion about how these phenomenons affect aberration estimation can be found in Ref. 14.

To study the variance in estimates of arrival time and amplitude fluctuations, a spatially uniform distribution of scatterers was used. Averaging over a sufficient number of independent scatterer realizations (receive signals), produced a totally incoherent medium. This will not be the case in practical ultrasound imaging. Here, tissue structures of different sizes are present in the receive signal which introduce regions (of varying size) with high spatial correlation. This may render the correlation estimation less efficient than for uniform scatterers. To study this is beyond the scope of this article.

As shown by Rigby *et al.*,[3] image artifacts due to strong off-axis scatterers may be a problem in a transmit-beam iteration procedure. A strong off-axis scatterer can dominate the receive signal and affect the estimation of arrival time and amplitude fluctuations. The result is that the corrected retransmitted beam "locks" on to the strong scatterers, producing image artifacts. Such cases are not included in the analysis presented in this article.

VI. CONCLUSIONS

In this article the MBFO algorithm,[1] previously developed in the frequency domain, is implemented in the time domain. The MBFO method uses an average of cross-spectrum components (frequency domain), or cross correlations (time domain) between neighboring elements. The number of neighbor elements used in the averaging process is determined by the spatial coherence function in the frequency domain, and the cross-correlation coefficient in the time domain.

The time-domain implementation introduces a loss of generality. In the frequency domain, individual frequency dependent amplitude and phase correction could be performed on each frequency component in the signal. In the time-domain implementation presented here, only a simple time delay and amplitude correction (independent of frequency) is possible.

In order to improve the cross-spectrum or cross-correlation estimates, a method of averaging the correlation over independent receive signals was proposed in Ref. 1. The number K of independent signals used in the averaging process affects the stability of arrival time and amplitude fluctuation estimates. How this stability varied as a function of K was investigated for eight different aberrators, all emulating the human abdominal wall.

For practical purposes, the results show that the time- and frequency-domain implementations of the MBFO algorithm yields the same estimates of arrival time and amplitude fluctuations, independent of K. The standard deviation of these estimates is approximately the same for both implementations. Increasing K reduces the standard deviation by approximately $1/\sqrt{K}$.

There is a strong variation in standard deviation between the different investigated aberrators. Still, aberration corrected beam profiles are improved for all aberrators and all K.

Using just one receive signal for the estimation ($K=1$) provided adequate correction for transmit-beam iteration. The convergence rate, and quality, was approximately equal compared to using 20 independent signals for averaging the correlation process.[14] This also holds for the scenario where the scattering region remained the same throughout the transmit-beam iteration process. These results indicate that there may be no need for averaging signals to obtain a proper estimate of the aberration. Based on the results presented here, almost ideal aberration correction is obtained using from 2–5 transmit beams (counting iteration number zero) in each beam direction. In a practical situation, this resembles Doppler imaging, where, e.g., ten transmit beams for each scan direction is usual to obtain proper signal-to-noise ratio. This again indicates that the MBFO algorithm may be well suited for real-time imaging using aberration correction.

Also, the similar results obtained with the time- and frequency-domain implementation proves that different strategies can be chosen for real-time implementation of the algorithm.

ACKNOWLEDGMENTS

The work presented in this article was supported by the Medicine and Health program of the Research Council of Norway.

[1] S.-E. Måsøy, T. Varslot, and B. Angelsen, "Estimation of ultrasound wave aberration with signals from random scatterers," J. Acoust. Soc. Am. **115**, 2998–3009 (2004).
[2] M. Karaman, A. Atalar, H. Köymen, and M. O'Donnell, "A phase aberration correction method for ultrasound imaging," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **40**, 275–282 (1993).
[3] K. W. Rigby, C. L. Chalek, B. H. Haider, R. S. Lewandowski, M. O'Donnell, L. S. Smith, and D. G. Wildes, "Improved in vivo abdominal image quality using real-time estimation and correction of wavefront arrival time errors," Proc.-IEEE Ultrason. Symp. **2**, 1645–1653 (2000).
[4] T. Varslot, B. Angelsen, and R. C. Waag, "Spectral estimation for characterization of acoustic aberration," J. Acoust. Soc. Am. **116**, 97–108 (2004).
[5] W. F. Walker and G. E. Trahey, "A fundamental limit on the performance of correlation based phase correction and flow estimation techniques," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **41**, 644–654 (1994).
[6] W. F. Walker and G. E. Trahey, "A fundamental limit on delay estimation using partially correlated speckle signals," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **42**, 301–308 (1995).
[7] G. C. Ng, S. S. Worrel, P. D. Freiburger, and G. E. Trahey, "A comparative evaluation of several algorithms for phase aberration correction," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **41**, 631–643 (1994).
[8] S. W. Flax and M. O'Donnell, "Phase-aberration correction using signals

from point reflectors and diffuse scatterers: Basic principles," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **35**, 758–767 (1988).
[9]J. C. Lacefield and R. C. Waag, "Spatial coherence analysis applied to aberration correction using a two-dimensional array system," J. Acoust. Soc. Am. **112**, 2558–2566 (2002).
[10]G. C. Carter, "Coherence and time delay estimation," Proc. IEEE **75**, 236–255 (1987).
[11]M. B. Priestley, *Spectral Analysis and Time Series* (Academic Press, New York, 1988).
[12]S.-E. Måsøy, T. F. Johansen, and B. Angelsen, "Correction of ultrasonic wave aberration with a time delay and amplitude filter," J. Acoust. Soc. Am. **113**, 2009–2020 (2003).
[13]F. Viola and W. F. Walker, "A comparison of the performance of time-delay estimators in medical ultrasound," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **50**, 392–401 (2003).
[14]S.-E. Måsøy, T. Varslot, and B. Angelsen, "Iteration of transmit-beam aberration correction in medical ultrasound imaging," J. Acoust. Soc. Am. **117**, 450–461 (2005).
[15]F. Lin and R. C. Waag, "Estimation and compensation of ultrasonic wavefront distortion using a blind system identification method," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **49**, 739–755 (2002).
[16]B. Angelsen, *Ultrasound Imaging. Waves, Signals and Signal Processing*, Vol. **II** (Emantec, Trondheim, 2000), http://www.ultrasoundbook.com
[17]S. Krishnan, K. W. Rigby, and M. O'Donnell, "Improved estimation of phase aberration profiles," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **44**, 701–713 (1997).
[18]R. Mallart and M. Fink, "Adaptive focusing in scattering media through sound-speed inhomogeneities: The van Cittert Zernike approach and focusing criterion," J. Acoust. Soc. Am. **96**, 3721–3732 (1994).

Comparative study of audio spatializers for dual-loudspeaker mobile phones

Mingsian R. Bai,[a] Geng-Yu Shih, and Chih-Chung Lee
Department of Mechanical Engineering, National Chiao-Tung University, 1001 Ta-Hsueh Road, Hsin-Chu 300, Taiwan, Republic of China

(Received 16 May 2006; revised 11 October 2006; accepted 11 October 2006)

MPEG-1, layer 3 handsets equipped with dual loudspeakers and three-dimensional audio modules have received much attention in the market of consumer electronics. To create spatial impression during audio reproduction, the head-related transfer function (HRTF) and the crosstalk cancellation system (CCS) are key elements in many audio spatializers. However, there are many factors that one should take into account during the design and implementation stages of an audio spatializer in the handset application. In the paper, a comprehensive study was undertaken to compare various audio spatializers for use with dual-loudspeaker handsets, in the context of inverse filtering strategies. Two deconvolution approaches, the frequency-domain method and the time-domain method, are employed to design the required inverse filters. Different approaches to design audio spatializers with the HRTF, CCS, and their combination are compared. In particular, two modified CCS approaches are suggested. Issues in the implementation phase such as regularization, complex smoothing, and structures of inverse filters are also addressed in the paper. Comprehensive objective and subjective tests were conducted to investigate the aforementioned aspects of audio spatializers. The data obtained from the subjective tests are processed by using the multianalysis of variance to justify statistical significance of the results.
[DOI: 10.1121/1.2387121]

PACS number(s): 43.60.Dh, 43.60.Pt, 43.60.Qv, 43.60.Uv [EJS] Pages: 298–309

I. INTRODUCTION

Thanks to rapid advances of mobile communication technology, handsets have swiftly entered everyone's daily life. In addition to a simple phone, a nowadays' handset has to serve also as a camera, a personal digital assistant, MPEG-1, layer 3 (MP3) player, and even a video player in the third-generation application. In order to cater to the ever-increasing demands of high quality audio, three-dimensional (3D) audio reproduction for use with dual-loudspeaker handsets has emerged. In 3D audio reproduction, the head-related transfer function (HRTF) and the crosstalk cancellation system (CCS) are two core technologies. HRTF is a mathematical model representing the propagation process from a sound source to the human ears. HRTFs thus contain localization cues as a result of the propagation delay and the diffraction effects due to the head, ears, and even torso. This allows us to create a directional impression by properly synthesizing HRTFs at the prescribed direction.[1] Although this is effective in headphone reproduction, a crosstalk problem arises when loudspeakers are used as the rendering transducers.[2,3] To overcome this problem, the CCS based on inverse filtering are employed to minimize the effects due to crosstalk that can obscure sound image. In general, two types of deconvolution approaches, the frequency-domain method[4] and the time-domain method,[5,6] can be utilized to design the required inverse filters. Since the acoustic systems, or plants, are usually noninvertible, some regularization measures have to be taken in these methods to avoid excessive boosts for the inverse filters caused by overcompensating the acoustic system. As an effective alternative, excessive gain of the inverse filters can also be avoided by smoothing the frequency response functions of the acoustic system prior to the inversion process.[7]

In inverse filter design, Norcross *et al.* pointed out that the time-domain methods are subjectively more robust but computationally less efficient than the frequency-domain method.[8] The main difficulty in the inversion process lies in the fact that the acoustic plants are typically nonminimum phase, meaning that a causal inverse filter does not exist.[9] To cope with the problem, a modeling delay was first introduced by Clarkson *et al.*[10] Furthermore, Kirkeby *et al.*[11] used the least-squares method along with a modeling delay to find the causal inverse filters. Wang and Pai also applied the time-domain method to determine the optimal modeling delay for the inverse filters.[12]

Conventional inverse filtering leads to reduced crosstalk and equalized ipsilateral response. However, if the CCS is inadequately designed, the latter effect can result in audible high-frequency artifacts. To address the problem, two modified CCS are proposed in this paper. The idea underlying these modified methods is to eliminate the crosstalk of the contralateral paths from the loudspeakers to the listener's ears without equalizing the ipsilateral paths. The modified CCS methods also have a desirable property that the CCS is loudspeaker independent. Extensive tests were conducted in the work to compare different approaches of audio spatializers based on the HRTF, CCS, and their combination.

[a]Author to whom correspondence should be addressed; electronic mail: msbai@mail.nctu.edu.tw

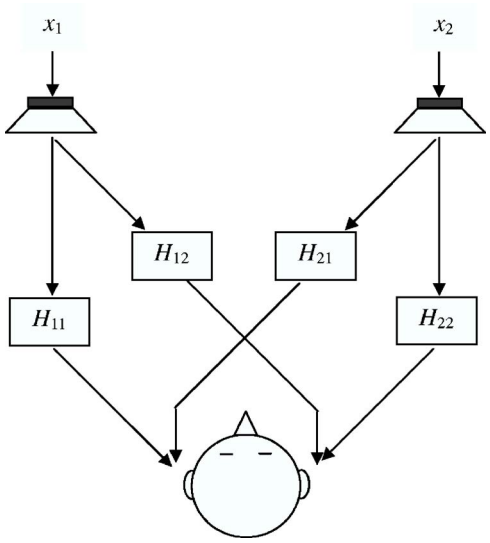


FIG. 1. Schematic diagram showing an audio reproduction system using two-channel stereo loudspeakers. Acoustic transfer functions between the loudspeakers and the listener's ears are indicated in the figure.

Another issue concerning the implementation phase is the structures of inverse filters. In considering psychoacoustic aspects and computational cost, the CCS can be implemented in a few different ways. Three structures of CCS are compared in this paper: the direct filtering method, the filter bank method,[13] and the simple lowpass mixing method.[14] The direct filtering method can be further divided into the full-band and the band-limited design.[3] The band-limited design limits the crosstalk cancellation to function only within the band 200–6 kHz.

In this work, comprehensive objective and subjective tests were conducted to investigate the aforementioned aspects of audio spatializers for mobile phones. The data of subjective tests are processed by using the multianalysis of variance (MANOVA) to justify the statistical significance of the results.[15]

II. CROSSTALK CANCELLATION SYSTEMS

A. Problem of crosstalk cancellation

Figure 1 shows a two-channel loudspeaker reproduction scenario, where H_{11} and H_{22} are ipsilateral transfer functions, and H_{12} and H_{21} are contralateral transfer functions from the loudspeakers to the listener's ears. The contralateral transfer functions, also known as the crosstalk, interfere with human's localization of sound sources when the binaural signals are reproduced by loudspeakers. In order to mitigate the effects of crosstalk, the crosstalk canceller is chosen to be the inverse of the acoustic plants such that the overall response becomes a diagonalized and distortionless response

$$\begin{bmatrix} \delta(n-m) & 0 \\ 0 & \delta(n-m) \end{bmatrix} = \begin{bmatrix} h_{11}(n) & h_{12}(n) \\ h_{21}(n) & h_{22}(n) \end{bmatrix} \otimes \begin{bmatrix} c_{11}(n) & c_{12}(n) \\ c_{21}(n) & c_{22}(n) \end{bmatrix}, \tag{1}$$

where $\otimes$ denotes convolution operation and $h_{ij}(n)$, $c_{ij}(n)$, and $\delta(n-m)$ represent the impulse responses of the respective acoustic paths, the inverse filters, and the discrete delta function delayed by m samples of delay to ensure a causal inverse filter. On the basis of inverse filtering, two deconvolution schemes along with regularization techniques are described in the following.

B. Multichannel inverse filtering with regularization

1. Frequency-domain deconvolution

The first method to be considered is the frequency-domain method[4] suggested by Kirkeby *et al.* In this method, a cost function J is defined as the sum of the "performance error" $\mathbf{e}^H\mathbf{e}$ and the "input power" $\mathbf{v}^H\mathbf{v}$,

$$J(e^{i\omega}) = \mathbf{e}^H(e^{j\omega})\mathbf{e}(e^{j\omega}) + \beta(\omega)\mathbf{v}^H(e^{j\omega})\mathbf{v}(e^{j\omega}) \tag{2}$$

with ω being the angular frequency. A regularization parameter $\beta(\omega)$ which varies from zero to infinite weighs the input power against the performance error. This is a well known Tikhonov regularization procedure. The optimal inverse filters obtained by minimizing J can be written in terms of discrete frequency index k as follows:

$$\mathbf{C}(k) = [\mathbf{H}^H(k)\mathbf{H}(k) + \beta(k)\mathbf{I}]^{-1}\mathbf{H}^H(k), \quad k = 1,2,\ldots,N_c, \tag{3}$$

where N_c-point fast Fourier transform (FFT) is assumed, and $\mathbf{H}(k)$ is the transfer matrix of acoustic plant. The coefficients of inverse filters can be obtained using the inverse FFT of the frequency response in Eq. (3), with the aid of appropriate windowing. In order to ensure the causality of the CCS filters, circular shift ($N_c/2$ maximum) of the resulting impulse response is needed to introduce a modeling delay.[16]

2. Time-domain deconvolution

The time-domain method is based on a matrix formalism of Eq. (1). In this method, a single-channel inverse filter can be obtained by solving the following matrix equation:[5,6]

$$\begin{bmatrix} d(0) \\ \vdots \\ d(N_h+N_c-2) \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} h(0) & & 0 \\ \vdots & \ddots & \vdots \\ h(N_h-1) & \ddots & h(0) \\ \vdots & \ddots & \vdots \\ 0 & & h(N_h-1) \\ \varepsilon & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \varepsilon \end{bmatrix} \times \begin{bmatrix} c(0) \\ \vdots \\ c(N_c-1) \end{bmatrix}, \tag{4}$$

or simply

$$\mathbf{d} = \mathbf{hc}. \tag{5}$$

In the preceding two equations, the vector $\mathbf{d}$ represents the desired response, the matrix $\mathbf{h}$ is composed of the impulse responses $h(n)$ of acoustical plants measured *a priori*, N_h is the length of the plant impulse response $h(n)$, the vector $\mathbf{c}$ represents the impulse response of the inverse filters, and N_c

is the length of the inverse filter. The parameter ε in the lower part of the matrix $\mathbf{h}$ is a small regularization constant. The forgoing single-channel deconvolution technique can be readily extended to the two-channel case described by the following matching matrix:

$$\begin{bmatrix} \mathbf{d} \\ 0 \\ 0 \\ \mathbf{d} \end{bmatrix} = \begin{bmatrix} \mathbf{h}_{11} & 0 & \mathbf{h}_{12} & 0 \\ 0 & \mathbf{h}_{11} & 0 & \mathbf{h}_{12} \\ \mathbf{h}_{21} & 0 & \mathbf{h}_{22} & 0 \\ 0 & \mathbf{h}_{21} & 0 & \mathbf{h}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{c}_{11} \\ \mathbf{c}_{12} \\ \mathbf{c}_{21} \\ \mathbf{c}_{22} \end{bmatrix}, \tag{6}$$

where $\mathbf{h}_{ij}$ and $\mathbf{c}_{ij}$ represent the matrices composed of the impulse responses $h_{ij}(n)$ and the coefficient vectors of the filters $c_{ij}(n)$.

The size of the matrix in Eq. (6) can be quite large. Instead of brute-force inversion, more efficient iteration techniques are employed in the work. By exploiting these properties one may use the iterative algorithms such as steepest descent and conjugate-gradient (CG) method to calculate the solution.[6] In both methods, a residual vector $\mathbf{R}$ is defined as

$$\mathbf{R} = \mathbf{D}_t - \mathbf{H}_t\mathbf{C}_t, \tag{7}$$

where $\mathbf{D}_t$, $\mathbf{H}_t$, and $\mathbf{C}_t$, represent the matrices in Eq. (6). In the steepest descent algorithm, the recursive relation for updating the coefficient of the inverse filters can be described as

$$\mathbf{C}_t(i+1) = \mathbf{C}_t(i) + \mu\mathbf{g}(i), \tag{8}$$

where i is the iterative index and $\mathbf{g}$ is the gradient vector of the cost function with a step size μ. Unlike the steepest descent algorithm, a plane search strategy based on the linear combination of gradient vectors consecutive iterations is used in the CG algorithm. Specifically, the coefficient update equation is given as

$$\mathbf{C}_t(i+1) = \mathbf{C}_t(i) + \mu\mathbf{g}(i) + \alpha\mathbf{s}(i), \tag{9}$$

where $\mathbf{s}$ is the gradient vector in last iteration and α is another step size parameter. In general, the convergence behavior of the CG method is superior to the steepest descent method due to the plane search nature of the former approach.

3. Generalized complex smoothing techniques

Due to the ill-conditioned nature of the acoustical system, how to properly limit the gain of the inverse filter is a critical issue in designing the CCS. One way to deal with this problem is the regularization method, as already mentioned in the previous section. Another simple but elegant way is to smooth the peaks and dips of the acoustic plant using the generalized complex smoothing technique suggested by Hatziantoniou and Mourjopoulos.[7] There are two alternative methods for implementing complex smoothing. The first method, uniform smoothing, is to calculate the impulse response using the inverse FFT of the frequency response. Then, apply a time-domain window to truncate and taper the impulse response, which in effect smoothes out the frequency response. Finally, recover the frequency response by FFT of the modified impulse response. Alternatively, a nonuniform smoothing method can also be used. This method performs smoothing directly in the frequency domain. The frequency response is circularly convolved with a window whose bandwidth increases with frequency. The choice of the window follows the psychoacoustics that the spectral resolution of human hearing increases with frequency. Therefore, the nonuniformly smoothed frequency response

$$H_{\mathrm{ns}}(m,k) = \sum_{i=0}^{N-1} H[(k-i) \bmod N] W_{\mathrm{sm}}(m,i), \tag{10}$$

where k, $0 \le k \le N-1$ is the frequency index and m is the smoothing index corresponding to the length of the smoothing window. The smoothing window $W_{\mathrm{sm}}(m,k)$ is given by

$$W_{\mathrm{sm}}(m,k) = \begin{cases} \dfrac{b-(b-1)\cos[(\pi/m)k]}{2b(m+1)-1}, & k = 0,1,\ldots,m \\ \dfrac{b-(b-1)\cos[(\pi/m)(k-N)]}{2b(m+1)-1}, & k = N-m, N-(m-1),\ldots,N-1 \\ 0, & k = m+1,\ldots,N-(m+1). \end{cases} \tag{11}$$

The integer, $m=m(k)$, can be considered as a bandwidth function by which a fractional octave or any other nonuniform frequency smoothing scheme can be implemented. The variable b determines the roll-off rate of the smoothing window. As a special case when $b=1$, the window reduces to a rectangular window.

C. Structures of inverse filters

There are a number of different ways to implement the inverse filters of CCS. The direct filtering method, the filter bank method, and the simple lowpass mixing method are three major filtering structures to discuss in this section.

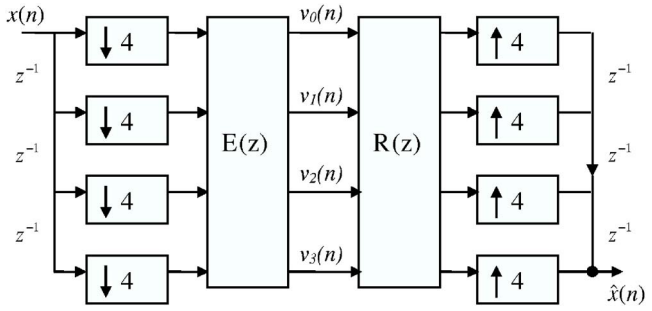


FIG. 2. The block diagram of a four-channel QMF bank using the polyphase representation.

1. Direct filtering method

In this structure, crosstalk cancellation is carried out by direct filtering using inverse filters. However, crosstalk cancellation can be demanded either for a full-band (200–24 kHz) performance or just a band-limited performance (200–6 kHz) in the design stage of inverse filters. The reason for the latter design is twofold. First, the sweet spot in which CCS is effective becomes impractically small at high frequencies. Second, a listener's head provides natural shadowing at high frequencies so that the need for cancellation becomes less important. The match equation appropriate for the band-limited design is written as[3]

$$\begin{bmatrix} \delta(n-m) & 0 \\ 0 & \delta(n-m) \end{bmatrix} = \begin{bmatrix} h_{11}(n) & h_{12}(n) \otimes f_{\mathrm{LP}}(n) \\ h_{21}(n) \otimes f_{\mathrm{LP}}(n) & h_{22}(n) \end{bmatrix} \otimes \begin{bmatrix} c_{11}(n) & c_{12}(n) \\ c_{21}(n) & c_{22}(n) \end{bmatrix}, \tag{12}$$

where $f_{\mathrm{LP}}(n)$ denotes the impulse response function of a lowpass filter. Thus, the inverse filters should in principle give rise to a flat response within the intended band after compensation.

2. Filter bank method

In the direct filtering approach, even if the inverse filters are deigned for band-limited performance, the filtering process is still carried out at a sampling rate of 48 kHz. To take advantage of the band-limited design, a subband filtering approach is exploited to simplify the computation. Specifically, a four-channel quadrature mirror filter (QMF) bank[13] is used to implement the CCS. For further enhancement of processing efficiency, the polyphase representation is employed to implement the QMF bank, as shown in Fig. 2. The block $\mathbf{E}(z)$ is the type 1 polyphase matrix for the analysis bank, and the block $\mathbf{R}(z)$ is the type 2 polyphase matrix for the synthesis bank. $v_i(n)$ represents the subband signal. The first subband signal is processed by the CCS and the other subband signals are simply delayed by the delay block $\mathbf{D}(z)$ and transmitted to the synthesis filter bank, as shown in Fig. 3.

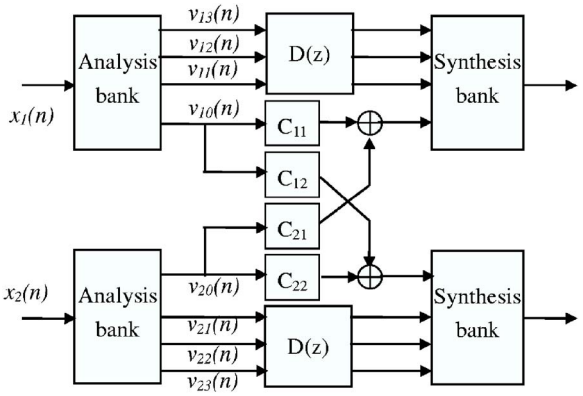


FIG. 3. Block diagram depicting the filter bank implementation of CCS.

3. Simple lowpass mixing method

For reference, a brief review of an alternative way of implementing the band-limited design originally proposed by Elliott *et al.* is also given (Fig. 4).[14] In this simple lowpass mixing approach, the input signal is lowpass filtered and down-sampled before sending to the CCS. Sufficient modeling delays must be inserted in the path. The CCS filters are adaptively updated by comparing the lowpass and delayed input and the lowpass plant out put at the control point (ears). Finally, the output of the CCS is up-sampled and remixed into the original full-band signal. The major difference between this method and the preceding filter bank method lies in the fact that the CCS-processed signal is mixed with the unprocessed full-band input in the simple mixing approach, while it is not the case in the filter bank method. This could have potential effect on the localization performance of spatializers.

D. Implemental issue

To facilitate the inverse filter design, the aforementioned smoothing techniques is employed to modify the impulse responses. On the other hand, the regularization parameters β and ε are selected to be 0.01 and 0.1 in the frequency-domain and time-domain deconvolutions, respectively, to limit the gain of the inverse filter to 10 dB maximum.

An objective index, channel separation, is employed to assess the cancellation performance

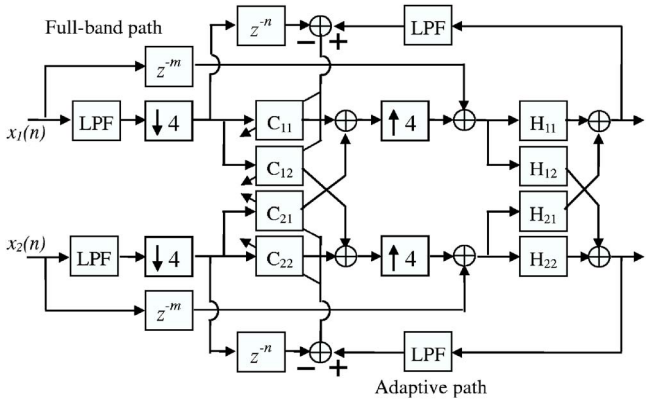


FIG. 4. Block diagram depicting the simple lowpass mixing implementation of CCS.

TABLE I. The test items used in the subjective evaluation.

Experiment 1	Test 1	Full-band frequency-domain CCS with uniform smoothing
	Test 2	Full-band time-domain CCS with uniform smoothing
Experiment 2	Test 1	Full-band conventional CCS with uniform smoothing
	Test 2	Full-band modified CCS-1 with uniform smoothing
	Test 3	Full-band modified CCS-2 with uniform smoothing
	Test 4	Commercial spatializer: DiMAGIC VX™ virtual sound imaging system
Experiment 3	Test 1	Full-band conventional CCS with uniform smoothing
	Test 2	Band-limited conventional CCS with uniform smoothing
	Test 3	Filter bank conventional CCS with uniform smoothing
	Test 4	Simple lowpass mixing conventional CCS with uniform smoothing
Experiment 4	Test 1	HRTF widening
	Test 2	Full-band conventional CCS with uniform smoothing
	Test 3	Full-band modified CCS-1 with uniform smoothing
	Test 4	HRTF+Full-band conventional CCS with uniform smoothing
	Test 5	HRTF+Full-band modified CCS-1 with uniform smoothing

$$S_{\mathrm{ep}}(j\Omega) = H_c(j\Omega)/H_i(j\Omega), \tag{13}$$

where $H_c(j\Omega)$ and $H_i(j\Omega)$ represent the contralateral (H_{12}, H_{21}) and the ipsilateral (H_{11}, H_{22}) frequency responses, respectively. According to the definition, a small (negative) value of channel separation indicates good cancellation performance.

III. DESIGN OF AUDIO SPATIALIZERS

A brief description of various approached based on HRTF and CCS will be given. For clarity, the experiments of audio spatializers were summarized in Table I.

A. HRTF

As mentioned previously, directional impression can be created by electronically synthesizing the HRTF in the desired angle. This is especially important in the case of mobile phones, where loudspeakers closely spaced. In this study, the HRTF database available in the website of the MIT media lab[1] was employed to "widen" the sound image. Each impulse response originally measured at a Knowles Electronic Mannequin for Acoustic Research (KEMAR) with a sampling frequency 44.1 kHz. HRTFs at the azimuth ±30° are implemented as 128-tapped finite impulse response (FIR) filters by which the audio input signals are filtered before sending to the loudspeakers. The processing can be written in matrix form as follows:

$$\begin{bmatrix} \hat{x}_1(n) \\ \hat{x}_2(n) \end{bmatrix} = \begin{bmatrix} h_{30\ \mathrm{ipsi}}(n) & h_{30\ \mathrm{contra}}(n) \\ h_{30\ \mathrm{contra}}(n) & h_{30\ \mathrm{ipsi}}(n) \end{bmatrix} \otimes \begin{bmatrix} x_1(n) \\ x_2(n) \end{bmatrix}, \tag{14}$$

where $h_{30\ \mathrm{ipsi}}(n)$ and $h_{30\ \mathrm{contra}}(n)$ denote the ipsilateral and contralateral HRTFs, respectively, at the azimuths ±30°.

B. CCS

The objective of CCS is to minimize the effect of crosstalk. A generic inverse filter of a two-channel CCS can be a factored into the following expression:

$$\mathbf{C} = \frac{1}{1 - \mathrm{ITF}_1\mathrm{ITF}_2} \begin{bmatrix} 1/H_{11} & 0 \\ 0 & 1/H_{22} \end{bmatrix} \begin{bmatrix} 1 & -\mathrm{ITF}_2 \\ -\mathrm{ITF}_1 & 1 \end{bmatrix}, \tag{15}$$

where $\mathrm{ITF}_1 = H_{12}/H_{11}$, $\mathrm{ITF}_2 = H_{21}/H_{22}$ are interaural transfer functions, and the ipsilateral transfer functions H_{11}, H_{22} and the contralateral transfer functions H_{12}, H_{21} are defined as in Fig. 1. The earlier expression reveals the fact that the inverse filters attempt not only cancel the crosstalk with delays (the third term on the right hand side) but also equalize the ipsilateral response (the second term on the right hand side). The poles of the comb filter of the first term on the right hand side give the *ringing frequency*.[17]

The ipsilateral equalization (the second term) in the inverse filters may not be always desirable in practical application. For example, coloration problem may arise at around 10 kHz when the inverse filters strive to compensate the concha dip in the ipsilateral responses, which is largely independent of loudspeaker span. In addition, the other dips and roll-offs particularly at the very at the low and high frequencies in the ipsilateral responses further aggravate this situation. Consequently, an unnatural change of sound quality is often audible during reproduction due to over-compensating the ipsilateral responses. To address the problem, two modified techniques of CCS are suggested in the following.

1. The modified CCS-1

In this method, the diagonal terms of the matching model in the left hand side of Eq. (1) are replaced with delayed ipsilateral impulse responses

$$\begin{bmatrix} h_{11}(n-m) & \gamma \\ \gamma & h_{22}(n-m) \end{bmatrix} = \begin{bmatrix} h_{11}(n) & h_{12}(n) \\ h_{21}(n) & h_{22}(n) \end{bmatrix} \otimes \begin{bmatrix} c_{11}(n) & c_{12}(n) \\ c_{21}(n) & c_{22}(n) \end{bmatrix}, \quad (16)$$

where γ is a small constant, e.g., 0.0001 and m is the modeling delay. This in effect modifies the transfer functions of inverse filters in Eq. (15) into

$$C \approx \frac{1}{1-\mathrm{ITF}_1\mathrm{ITF}_2}\begin{bmatrix} 1 & -\mathrm{ITF}_2 \\ -\mathrm{ITF}_1 & 1 \end{bmatrix}. \quad (17)$$

The modified CCS makes no attempt to compensate the ipsilateral responses when canceling the crosstalk. It follows that the sound quality can be better preserved by using this method.

There is another potential benefit in the use of this method. Assume that two speaker responses are displaced by a factor S. Neglecting the parameter γ, the z-domain version of Eq. (16) can be written as

$$\begin{bmatrix} z^{-m}\tilde{H}_{11}(z)S & 0 \\ 0 & z^{-m}\tilde{H}_{22}(z)S \end{bmatrix} \approx \begin{bmatrix} \tilde{H}_{11}(z)S & \tilde{H}_{12}(z)S \\ \tilde{H}_{21}(z)S & \tilde{H}_{22}(z)S \end{bmatrix}\begin{bmatrix} C_{11}(z) & C_{12}(z) \\ C_{21}(z) & C_{22}(z) \end{bmatrix}, \quad (18)$$

where $\tilde{H}_i$ represents the transfer function without loudspeaker responses. Thus, the factor S cancels out on both sides. The implication of this is that the CCS is loudspeaker independent as long as the characteristics of two loudspeakers are well matched. This could be a desirable property in practical applications in that a CCS designed off-line is applicable to all systems with different loudspeaker characteristics.

2. The modified CCS-2

Along the same line, another modified CCS is developed to underplay the equalization of ipsilateral response during cancellation of crosstalk. In this approach, the ipsilateral inverse filters are assigned to be a delayed discrete delta function, i.e., $c_{11}=c_{22}=\delta(n-m)$ such that the sound quality can be preserved because of the direct transmission of ipsilateral paths. In this setting, the match equation should be modified into

$$\begin{bmatrix} d_L(n) & 0 \\ 0 & d_R(n) \end{bmatrix} = \begin{bmatrix} h_{11}(n) & h_{12}(n) \\ h_{21}(n) & h_{22}(n) \end{bmatrix} \otimes \begin{bmatrix} \delta(n-m) & c_{12}(n) \\ c_{21}(n) & \delta(n-m) \end{bmatrix}, \quad (19)$$

where the diagonal terms d_L and d_R are the resulting ipsilateral responses. Expanding this equation only for the off-diagonal terms leads to two equations

$$-(h_{12}(n) \otimes \delta(n-m)) = -h_{12}(n-m) = h_{11}(n) \otimes c_{12}(n), \quad (20)$$

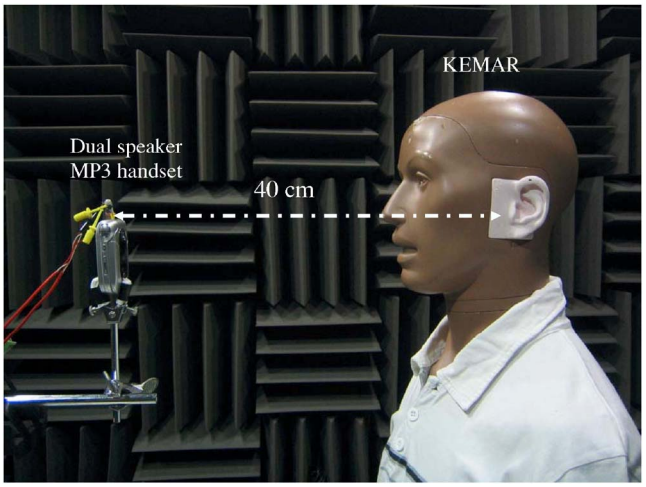


FIG. 5. Experimental arrangement for the dual speaker handset with a dummy head system inside an anechoic chamber.

$$-(h_{21}(n) \otimes \delta(n-m)) = -h_{21}(n-m) = h_{22}(n) \otimes c_{12}(n). \quad (21)$$

The contralateral inverse filters can be obtained by solving this inverse problem. By the same token, it can be shown that this modified CCS is also loudspeaker independent. However, this approach would possibly lead to poor bass response because the crosstalk canceller will no longer have the factor $1/(1-\mathrm{ITF}^2)$, which is essentially a bass boost.

IV. EXPERIMENTAL INVESTIGATIONS

A. Experimental arrangement

The experiments were conducted by using a dummy head system (KEMAR) inside a 4 m×4 m×3 m anechoic chamber, as shown in Fig. 5. An MP3 handset equipped with dual loudspeakers is mounted on a stand. The distance between the handset, and the dummy head is 40 cm. Binaural transfer functions from the loudspeakers to the microphone embedded in the dummy head's ears were measured by using a spectrum analyzer. The algorithms were implemented on the platform of a fixed-point DSP, ADI BF-533, operating at 48 kHz. The inverse filters were realized as 128-tapped FIR filters in the experiments.

B. Objective experiment

For simplicity, symmetrical acoustic plant is assumed. The head-related impulse responses measured by using the dummy head is shown in Fig. 6. The complex smoothing is applied prior to the design of CCS. In this regard, the CCS will prove more robust against misalignment of the listener's head than that designed for unsmoothed frequency responses.[10,18] Figure 7 shows frequency responses obtained using uniform smoothing and nonuniform smoothing. It can be seen that the frequency responses are effectively smoothed by both methods. However, an informal subjective test has indicated that the difference between the two smoothing techniques is hardly detectible. The uniform smoothing method, therefore, is used exclusively in the following experiments.

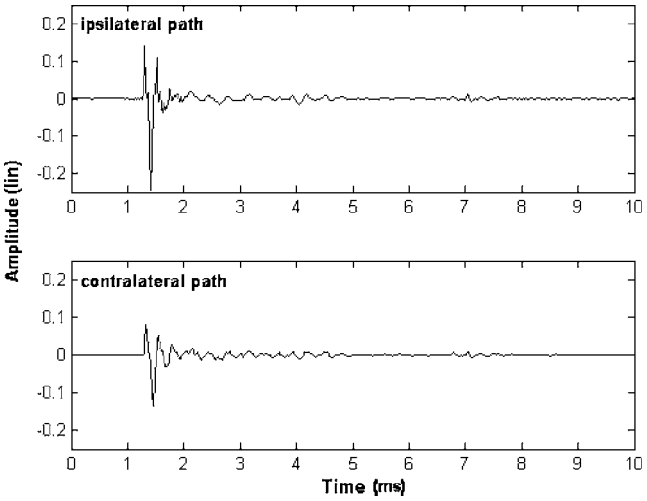


FIG. 6. Head-related impulse responses measured by using the dummy head system.

Another issue concerning the CCS design is a modeling delay that is necessary to ensure the causality of inverse-filters. This is of fundamental importance whether the frequency-domain method or the time-domain method is used. A simple experiment was conducted to examine the effect of different modeling delays on a 128-tapped filter and a 512-tapped taps filter obtained using the time-domain method. Average channel separation (Ave-Sep, dB) between 200 and 20 kHz is calculated to assess the cancellation performance. The result summarized in Table II reveals that the optimal modeling delay is approximately half of the length of the inverse filter.

The length of the inverse filter also affects the performance of CCS. The performance of inverse filters with different length of inverse filter is compared in Table III. As expected, the performance of CCS improves as the filter length is increased for both deconvolution methods. However, it is worth noting that the time-domain method outperforms the frequency-domain method for short filter length such as 128 taps. The frequency-domain method performs well only when a long filter is used. Another drawback of the frequency-domain method can be clearly seen by plotting the magnitudes of the equalized time responses on the dB scale, as suggested by Fielder.[19] In Fig. 8(a), preringing artifacts are visible (at 1–3 ms) in the equalized time responses when the frequency-domain method is used, while no such artifacts are found in the result of the time-domain method in Fig. 8(b).

Next, a useful variation of inverse filter design to enhance CCS performance is examined. Figure 9(a) shows the experimental results of the unprocessed and the processed frequency responses with the conventional CCS. While the flat spectrum is attained as expected in the compensated ipsilateral response, the contralateral response is not totally eliminated but amplified at the frequencies above 10 kHz. This incurs some audible coloration at high frequencies. To overcome the problem, the aforementioned modified approaches were employed to suppress the crosstalk while preserving the ipsilateral response. Figures 9(b) and 9(c) refer to the implementation of the modified CCS-1 and CCS-2, respectively. It is observed that not only the ipsilateral response remains largely unchanged but also the contralateral response is effectively attenuated without undesired amplification in high frequencies. To explore further the modified CCS, the time responses of the inverse filters of the modified methods are compared with those obtained using the conventional identity matching model. Figure 10(a) refers to the implementation of the conventional CCS. Figures 10(b) and

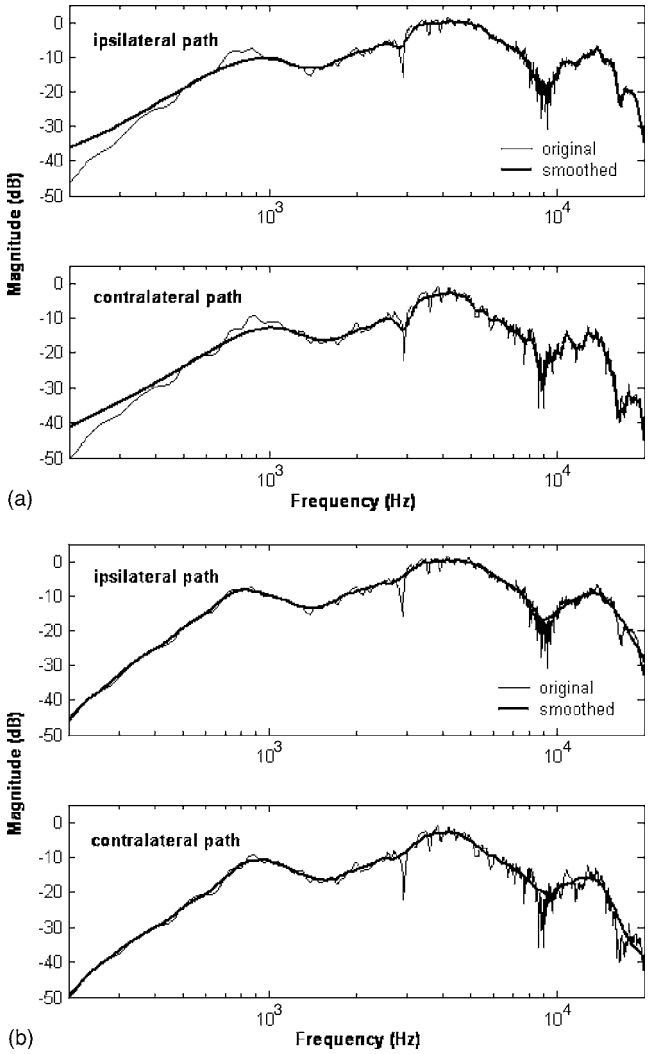


FIG. 7. Comparison between original and the complex smoothed magnitude spectrum. The thick line represents the complex smoothed magnitude response spectrum. (a) Result obtained using uniform smoothing. (b) Result obtained using nonuniform smoothing.

TABLE II. The average separation obtained using the time-domain method with different delays.

Filter length (N_c): 128 taps		Filter length (N_c): 512 taps	
Delay (m)	Average separation (dB)	Delay (m)	Average separation (dB)
16	−20.583	32	−20.799
32	−20.717	128	−21.592
48	−20.833	256	−21.701
64	−21.050	288	−21.706
80	−21.007	320	−21.692
96	−20.282	448	−20.771

TABLE III. The average separation obtained using inverse filtering with different filter length.

Filter length (N_c)	Average separation (dB)	
	Frequency domain	Time domain
128	−18.203	−21.050
256	−18.361	−21.608
512	−21.535	−21.705
1024	−22.329	−21.760
2048	−22.375	−21.870

10(c) refer to the implementation of the modified CCS-1 and CCS-2, respectively. The impulse responses of inverse filters designed using the modified methods are significantly shorter than those of the conventional method. This computational saving is a benefit for real-time implementation.

The inverse filters were implemented by using the band-limited design as detailed in the preceding section. It can be seen in the experimental result of Fig. 11 that the CCS maintains wideband equalization of the ipsilateral response to result in a flat spectrum, while the cancellation of crosstalk is only attained in low frequency range with some unwanted amplification in the high frequency range. Cancellation performance is confined in low frequency range as it should be for the filter bank method and the simple lowpass mixing method since they are essentially band-limit designs.

C. Subjective experiment

In order to assess the perceptual performance of the spatializers, subjective listening tests were conducted according to the double-blind triple stimulus with hidden reference method suggested in the standard ITU-R BS. 1116-1.[20] The listening tests were carried out inside the anechoic chamber. The program material consists of various instruments with significant dynamic variations between the two stereo channels. Both timbre-related and space-related qualities are considered. The loudness of each reproduced signal was adjusted with equal power. Nine subjective indices employed in the subjective tests are summarized as follows:

(1) Fullness: Dominance of low-frequency sound;
(2) Brightness: Dominance of high-frequency sound;
(3) Noise and distortion: Any extraneous disturbances to the signal are considered as noise. Effect on the signal that produces new sounds or timbre change is considered as distortion;
(4) Width of stage: Perceived angular width of extreme left to extreme right edges of the stage;
(5) Depth perception: Ability to hear that performers are appropriately localized from the front to the rear of the sound stage;
(6) Spaciousness: Perceived quality of listening within a reverberant environment. The sound is perceived as open, not constrained to the locations of the loudspeakers. The perception is an important part of the "you are there" sensation;
(7) Localization: Determination by a subject of the apparent direction or distance, or both, of a sound source;

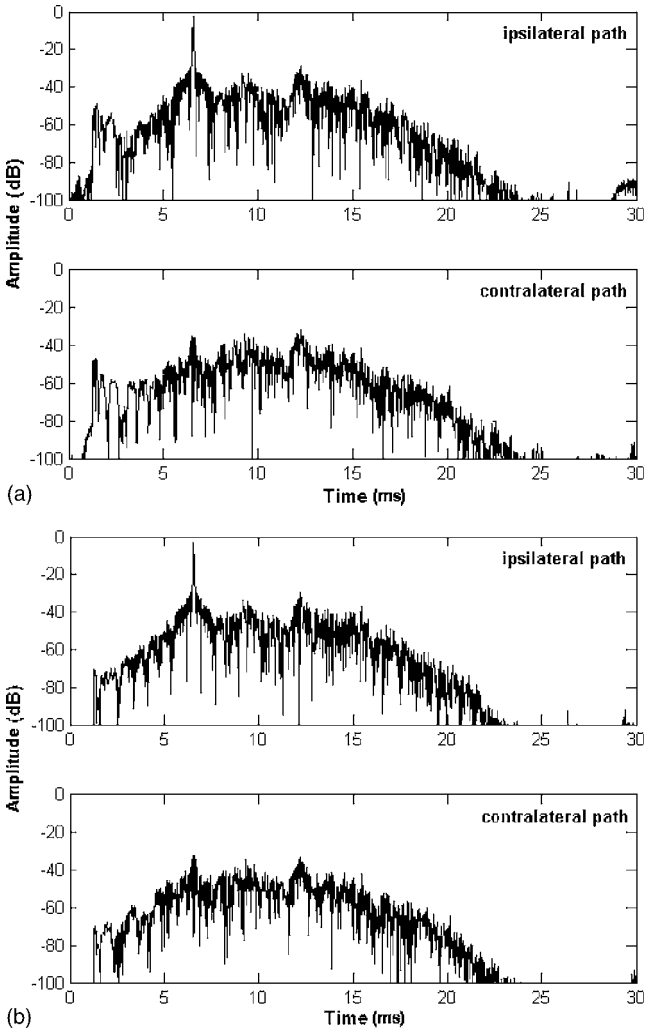


FIG. 8. Equalized time responses plotted on the dB scale. (a) Frequency-domain method. (b) Time-domain method.

(8) Robustness: Stability of performance with normal listener movements and listening locations. This index is assessed by 5 and 10 cm lateral movement of listener's head, and calculating the average grade; and
(9) Fidelity: The clarity of the reproduced signals.

Twenty experienced subjects participating in the tests were instructed with definition of the preceding subjective indices and the procedure before the listening tests. The subjects were asked to respond after listening in a questionnaire, with the aid of a set of subjective indices placed on a scale from −4 to 4. Positive, zero, and negative scores indicate perceptually improvement, no difference, and degradation, respectively, of the signals after processed by the spatializers. In order to justify the statistical significance, the scores were further processed by using the MANOVA.[15] Cases with significance levels below 0.05 indicate that statistically significant difference exists among methods. The experiments were summarized in Table I.

The first listening test was carried to compare the frequency-domain and the time-domain methods. The total grades are plotted in Fig. 12. The vertical bars denote 0.95 confidence intervals. The small significance level (s

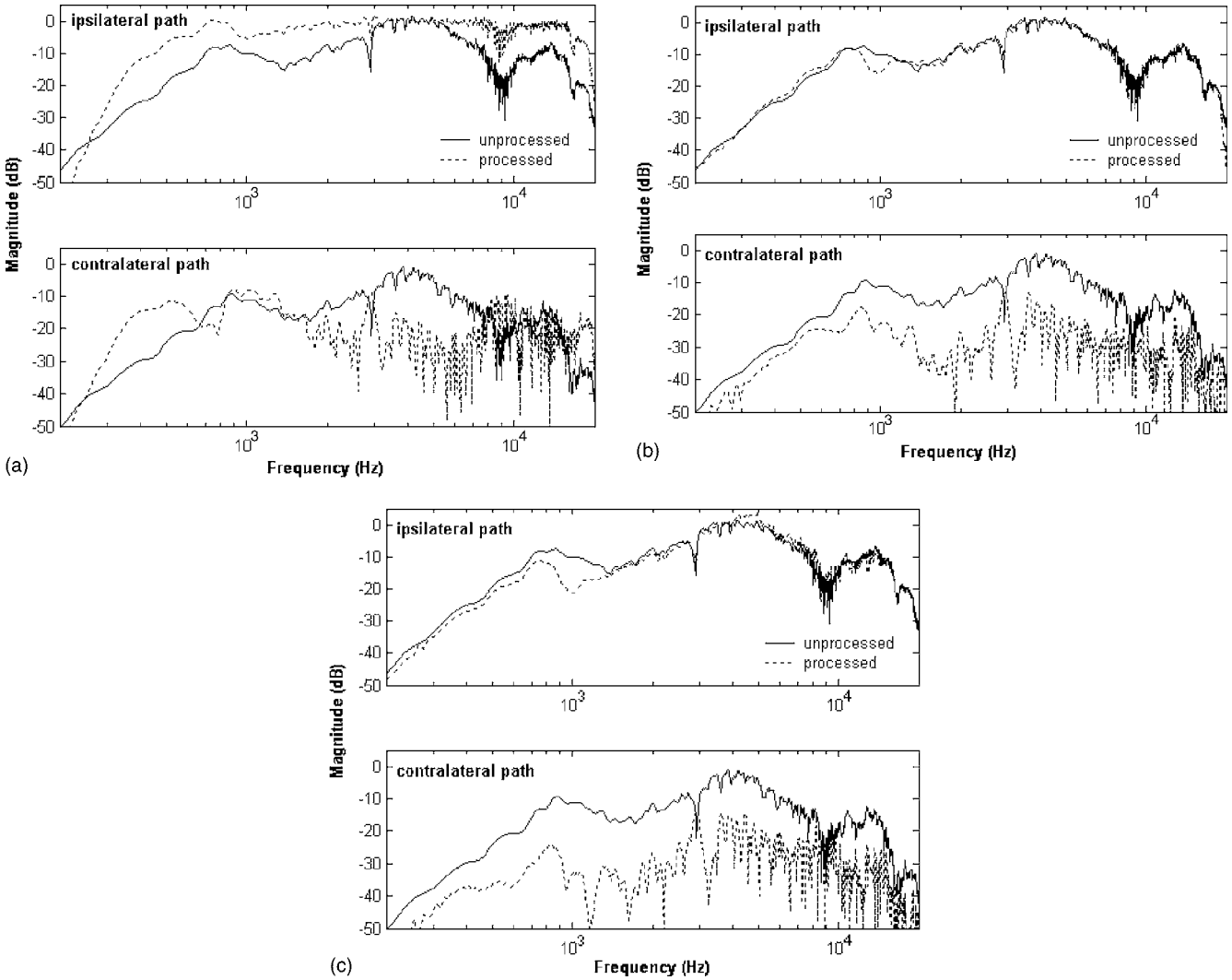


FIG. 9. Comparison between the unprocessed and the processed frequency responses. (a) The conventional CCS. (b) The modified CCS-1. (c) The modified CCS-2.

=0.041523) in the MANOVA output indicates that the difference among the methods is statistically significant. In particular, the time-domain method seemed to significantly outperform the frequency-domain method for inverse filters of this length (128 taps), which is in agreement with the observation in the preceding objective tests.

Next, the second experiment is performed to compare the modified CCS methods and a commercial spatializer[21] which is used in this experiment as the benchmark. The results shown in Fig. 13 revealed that the modified method-1 received the highest score among all approaches with strong statistical significance (s=0.000001). The modified method-1 is particularly advantageous when sound quality is used as the performance index in addition to the cancellation performance.

In the third listening test, different structures of CCS implementation are compared. The total grades are summarized in Fig. 14. The MANOVA output reveals that significant difference in performance (s=0.019207) does exist among the methods. The direct filtering method has attained the highest grade, while the simple lowpass mixing method received the lowest grade. In the direct filtering approach, there is no significant difference between the full-band and the band-limited designs. It is worth noting that the filter bank approach and the simple lowpass mixing approach did not attain the grades as high as two other direct filtering approaches. Possible explanations for this are that the crossovers in the filter bank are not adequately handled in the filter bank methods, and portion of the low-frequency signal is contaminated by crosstalk in the simple lowpass mixing method.

In the fourth listening test, various audio spatializers utilizing the HRTF, the conventional CCS method, the modified CCS method-1, and their combinations are compared. The total grades are summarized in Fig. 15. The MANOVA output reveals that significant difference in performance (s =0.000001) exists among the methods. It is observed from the result that the HRTF approach receives the lowest grade. The "widening" effect provided by the HRTF solely is obviously insufficient to spatialize the sound image due to the severe crosstalk between the closely spaced loudspeakers. In contrast to the HRTF approach, there is a leap in performance when the CCS comes into play. In particular, the spatializer combining the HRTF and the conventional CCS

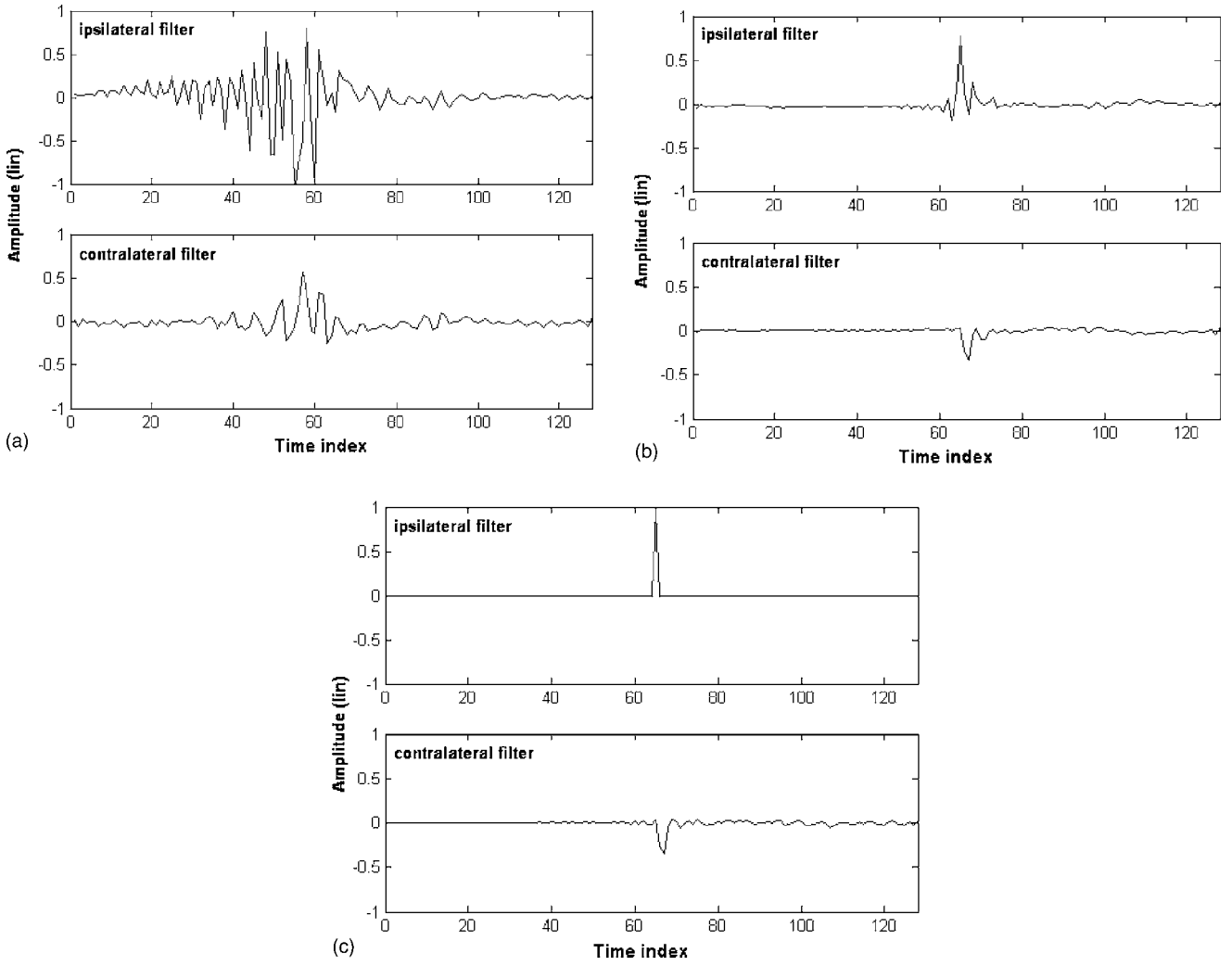


FIG. 10. Impulse responses of the inverse filters. (a) The conventional CCS method. (b) The modified CCS-1. (c) The modified CCS-2.

method has achieved the highest grade in both spatializing performance and sound quality. Surprisingly, when the modified CCS method is used in combination with the HRTF, there is a sudden drop in performance. It is suspected that double HRTF filtering effect may have contributed to this result. That is, while the sound quality has already been preserved by plugging the HRTF in the matching model for the modified CCS, the additional HRTF filtering becomes superfluous and may adversely affect the sound quality of the processed signal.

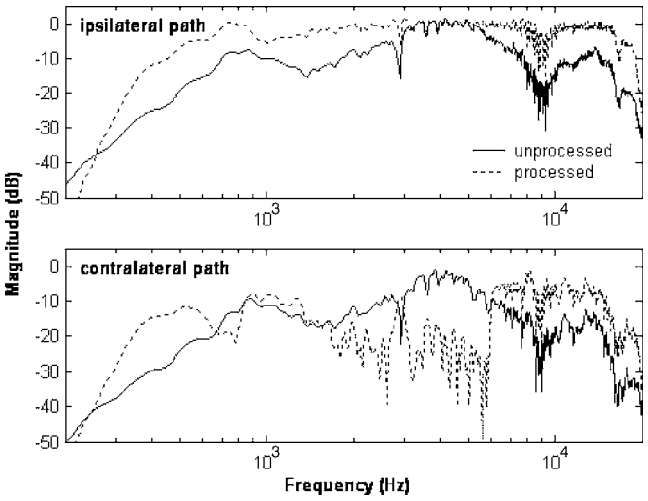


FIG. 11. Comparison between the unprocessed frequency response and that processed by using the band-limited CCS.

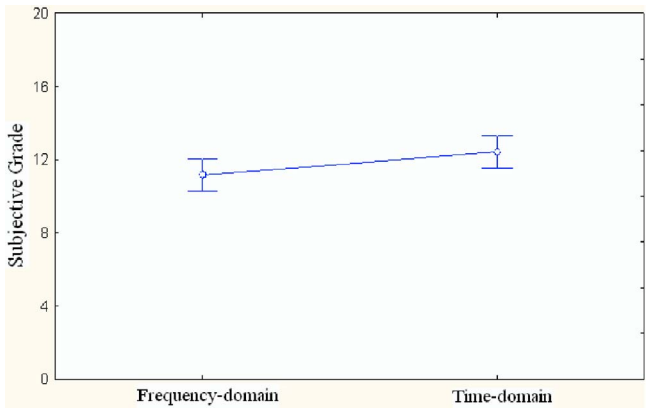


FIG. 12. Total grades summarized for the first listening test in which the frequency-domain and the time-domain deconvolution methods are compared. The significance level, $s=0.041523$, in the MANOVA output.

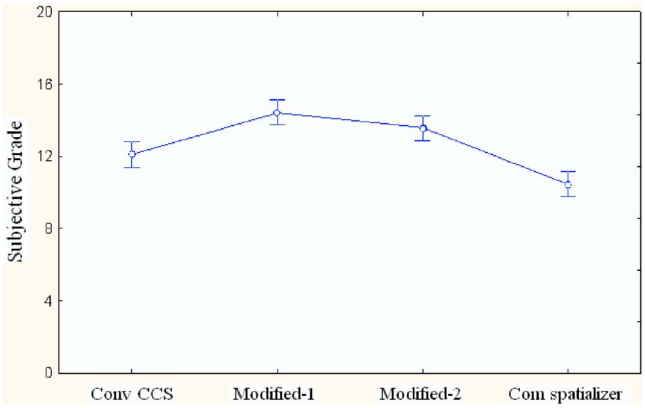


FIG. 13. Total grades summarized for the second listening test in which various CCS approaches including a commercial spatializer are compared. (Conv CCS: conventional CCS with identity matching, modified-1: modified CCS-1, modified-2: modified CCS-2, com spatializer: DiMAGIC VX™ virtual sound imaging system.) The significance level, $s=0.000001$, in the MANOVA output.

V. CONCLUSIONS

A comprehensive study has been undertaken to compare various implementation approaches of audio spatializer for handsets fitted with two closely spaced loudspeakers. The HRTF and the CCS techniques were exploited to implement the audio spatializer. Two deconvolution methods were applied to calculate the inverse filters for the CCS design. Objective and subjective experiments reveal that the time domain approach is superior to the frequency-domain approach when the length of inverse filter is short. An additional benefit of the time-domain method is that it is less liable to preringing artifact that frequently appears in the frequency-domain method.

Different structures of CCS were examined in this study. The experimental results indicate that the direct filtering approaches outperform the filter bank method and the simple lowpass mixing method. In addition, two modified CCS techniques were proposed in the present paper. Unlike the conventional method that tends to over-compensate the ipsilateral responses, the modified methods are capable of delivering better spaciousness without compromising on sound quality. Two additional features of the modified CCS which are attractive in practical application lie in its shorter impulse responses of the inverse filters and the loudspeaker-independent property.

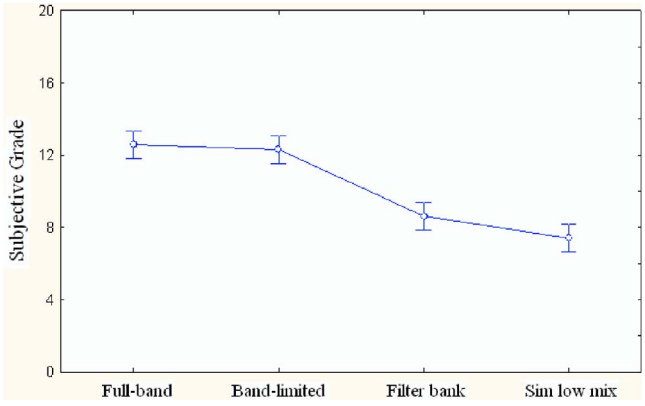


FIG. 14. Total grades summarized for the third listening test in which different structures of CCS implementation are compared. (full-band: full-band CCS, band-limited: band-limited, filter bank: filter bank CCS, sim low mix: simple lowpass mixing CCS.) The significance level, $s=0.019207$, in the MANOVA output.

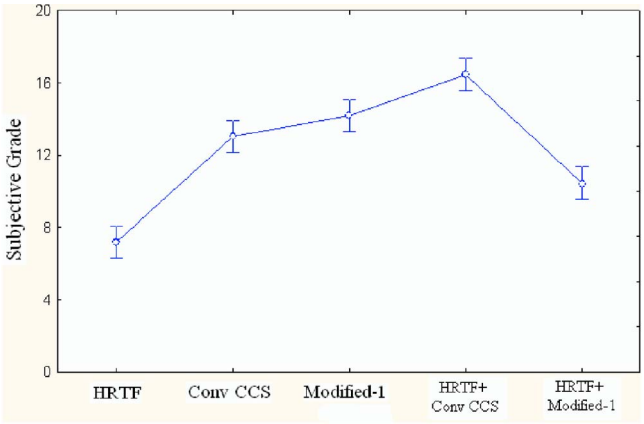


FIG. 15. Total grades summarized for the fourth listening test in which various audio spatializers utilizing the HRTF, the conventional CCS method, the modified CCS-1, and their combinations are compared. (HRTF: HRTF widening, conv CCS: conventional CCS, modified-1: modified CCS-1, HRTF+conv CCS: HRTF combined with conventional CCS, HRTF +modified-1: HRTF combined with modified CCS-1.) The significance level, $s=0.000001$, in the MANOVA output.

Listening tests were also carried out to compare various ways of implementing a spatializer based on HRTF, CCS, and their combination. The experimental results suggest that the widening effect provided by the HRTF solely is insufficient to spatialize the sound image due to the severe crosstalk between the closely spaced loudspeakers. In contrast to the HRTF approach, there is a leap in performance when the CCS is used. In particular, the spatializer combining the HRTF and the conventional CCS method has achieved the best performance in both spatializing performance and sound quality.

ACKNOWLEDGMENT

The work was supported by the National Science Council in Taiwan, Republic of China, under the Project No. NSC94-2212-E-009-019.

[1]B. Gardner and K. Martin, "HRTF measurements of KEMAR dummy-head microphone," MIT Media Lab, 1994; http://sound.media.mit.edu/KEMAR.html. Last accessed 11/10/06.
[2]A. Sibbald, "Transaural acoustical crosstalk cancellation," Sensaura White Paper, 1999, http://www.sensaura.co.uk. Last accessed 11/10/06.
[3]W. G. Gardner, *3-D Audio Using Loudspeakers* (Kluwer Academic, Dordrecht, 1998).
[4]O. Kirkeby, P. A. Nelson, and H. Hamada, "Fast deconvolution of multichannel systems using regularization," IEEE Trans. Speech Audio Process. **6**, 189–195 (1998).
[5]O. Kirkeby and P. A. Nelson, "Digital filter design for inversion problems in sound reproduction," J. Audio Eng. Soc. **47**, 583–595 (1999).
[6]J. F. Claerbout, *Earth Soundings Analysis: Processing Versus Inversion (PVI)*, 1992; http://sep.stanford.edu/sep/prof/toc_html/index.html. Last accessed 11/10/06.
[7]P. D. Hatziantoniou and J. N. Mourjopoulos, "Generalized fractional-octave smoothing of audio and acoustic responses," J. Audio Eng. Soc. **48**, 259–280 (2000).
[8]S. G. Norcross, G. A. Soulodre, and M. C. Lavoie, "Subjective investigations of inverse filtering," J. Audio Eng. Soc. **52**, 1003–1028 (2004).

[9]S. Neely and J. B. Allen, "Invertibility of a room impulse response," J. Acoust. Soc. Am. **66**, 165–169 (1979).
[10]P. M. Clarkson, J. Mourjopoulos, and J. K. Hammond, "Spectral, phase, and transient equalization for audio systems," J. Audio Eng. Soc. **33**, 127–132 (1985).
[11]O. Kirkeby, P. A. Nelson, F. Orduna-Bustamante, and H. Hamada, "Local sound field reproduction using digital signal processing," J. Acoust. Soc. Am. **100**, 1584–1593 (1996).
[12]J. H. Wang and C. S. Pai, "Subjective and objective verifications of the inverse functions of binaural room impulse response," Appl. Acoust. **64**, 1141–1158 (2003).
[13]P. P. Vaidyanathan, *Multirate Systems and Filter Banks* (Prentice-Hall, Englewood Cliffs, NJ, 1993).
[14]S. J. Elliott, P. A. Nelson, and I. M. Stothers, "Sound reproduction systems," U.S. Patent No. 5,727,066 (1998).
[15]G. Keppel and S. Zedeck, *Data Analysis for Research Designs* (W. H. Freeman, New York, 1989).
[16]H. Hamada, "Construction of orthostereophonic system for the purposes of quasiinsitu recording and reproduction," J. Acoust. Soc. Jpn. **39**, 337–348 (1983).
[17]J. Rose, P. A. Nelson, B. Rafaely, and T. Takeuchi, "Sweet spot size of virtual acoustic imaging systems at asymmetric listener locations," J. Acoust. Soc. Am. **112**(5), 1992–2002 (2002).
[18]S. Salamouris, K. Politopoulos, V. Tsakiris, and J. Mourjopoulos, "Digital system for loudspeaker and room equalization," J. Audio Eng. Soc. **43**, 396 (1995).
[19]L. D. Fielder, "Analysis of traditional and reverberation-reducing method of room equalization," J. Audio Eng. Soc. **51**, 3–26 (2003).
[20]ITU-R BS. 1116, "Methods for the subjective assessment of small impairments in audio system including multichannel sound systems," Geneva, Switzerland, 1994.
[21]DiMAGIC, "DiMAGIC VX™ virtual sound imaging system," White Paper, 2000; http://www.dimagic.com/pdf/DiMAGIC_Virtualizer_X_White_paper.pdf. Last accessed 11/10/06.

A neural circuit transforming temporal periodicity information into a rate-based representation in the mammalian auditory system

Ulrike Dicke and Stephan D. Ewert
Carl von Ossietzky Universität Oldenburg, Medizinische Physik, D-26111 Oldenburg, Germany

Torsten Dau[a)]
Centre for Applied Hearing Research, Ørsted•DTU, Technical University of Denmark, DK-2800 Kgs. Lyngby, Denmark

Birger Kollmeier
Carl von Ossietzky Universität Oldenburg, Medizinische Physik, D-26111 Oldenburg, Germany

(Received 26 May 2006; revised 24 October 2006; accepted 26 October 2006)

Periodic amplitude modulations (AMs) of an acoustic stimulus are presumed to be encoded in temporal activity patterns of neurons in the cochlear nucleus. Physiological recordings indicate that this temporal AM code is transformed into a rate-based periodicity code along the ascending auditory pathway. The present study suggests a neural circuit for the transformation from the temporal to the rate-based code. Due to the neural connectivity of the circuit, bandpass shaped rate modulation transfer functions are obtained that correspond to recorded functions of inferior colliculus (IC) neurons. In contrast to previous modeling studies, the present circuit does not employ a continuously changing temporal parameter to obtain different best modulation frequencies (BMFs) of the IC bandpass units. Instead, different BMFs are yielded from varying the number of input units projecting onto different bandpass units. In order to investigate the compatibility of the neural circuit with a linear modulation filterbank analysis as proposed in psychophysical studies, complex stimuli such as tones modulated by the sum of two sinusoids, narrowband noise, and iterated rippled noise were processed by the model. The model accounts for the encoding of AM depth over a large dynamic range and for modulation frequency selective processing of complex sounds. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2400670]

PACS number(s): 43.64.Bt, 43.64.Qh [BLM] Pages: 310–326

I. INTRODUCTION

Amplitude modulations (AMs) are a common physical feature produced by many natural sound sources such as vocal cords, strings, pipes and tubes. Amplitude modulation information is utilized by the auditory system in a variety of complex auditory tasks, including speech perception (e.g., Rosen, 1992), pitch perception (e.g., Schouten *et al.*, 1962; De Boer, 1976) and auditory scene analysis (e.g., Bregman, 1990). However, the mechanisms by which the auditory system extracts, encodes and processes AM information are not yet fully understood (for a review, see Joris *et al.*, 2004). While AM information is encoded in temporal activity patterns of neurons at the early stages of the auditory system, it is generally believed that this temporal periodicity code is transformed into a rate-based code at some higher stage of auditory processing (Rees and Møller, 1983; Schreiner and Langner, 1988; Lorenzi *et al.*, 1995; Schulze and Langner, 1997). Such a rate-based periodicity code also corresponds to the psychophysical concept of amplitude modulation filters that have been suggested to account for the modulation frequency selectivity observed in psychophysical masking experiments (Bacon and Grantham, 1989; Houtgast, 1989; Dau *et al.*, 1997a, b; Ewert and Dau, 2000). However, while the psychophysically derived modulation filters may provide a way of effectively describing perceptual masking data, it remains unclear how such filters might be realized neurally in the auditory system. Likewise, having a nonlinear neural circuit at hand that produces the above transformation using simple stimuli such as sinusoidally modulated tones, it is not clear to what extent the properties of the circuit are compatible with a (linear) modulation filter bank concept motivated by the psychophysical data. Both aspects are addressed in the current study.

Physiological experiments have revealed that the average spike rates of auditory-nerve (AN) fibers are largely independent of stimulus modulation frequency (Rose and Capranica, 1985; Joris and Yin, 1992; Rhode and Greenberg, 1994), indicating that AN fibers are not selective to certain modulation frequencies in terms of their spike rate. The temporal modulation transfer functions (tMTFs) of AN fibers, where modulation gain is computed in response to different modulation frequencies, have a low-pass shape with cutoff frequencies below 1 kHz (Javel, 1980; Palmer, 1982; Kim *et al.*, 1990; Joris and Yin, 1992; Rhode and Greenberg, 1994). Moreover, AM information is inherent in the interspike intervals (ISI) of AN fibers (Rose, 1970; Evans, 1978), indi-

[a)]Author to whom correspondence should be addressed. Electronic mail: tda@oersted.dtu.dk

cating that AMs are encoded in the temporal activity patterns of AN fibers. The majority of neurons in the cochlear nucleus (CN) have been found to display flat or low-pass rate modulation transfer functions (rMTFs), where spike rate is computed in response to different stimuli (Kim *et al.*, 1990; Rhode, 1994; Backoff *et al.*, 1999). Since the temporal synchronization of CN neurons to AM stimuli exceeds that of AN fibers, temporal AM information present in the activity patterns of AN fibers seems to be enhanced by CN neurons (Frisina *et al.*, 1990; Kim *et al.*, 1990; Rhode, 1994; Rhode and Greenberg, 1994; Joris *et al.*, 2004; Palmer, 1995). Thus, since CN neurons are not selective to a certain modulation frequency in terms of their spike rate, AM information can be assumed to be encoded in their temporal activity patterns. While AN fibers and CN neurons have been found to synchronize to modulation frequencies as high as 1 kHz (Javel, 1980; Frisina *et al.*, 1990; Rhode and Smith, 1986; Backoff *et al.*, 1999), the highest modulation frequency for which temporal synchronization has been observed in the auditory midbrain differs from below 120–150 Hz (Rees and Møller, 1983; Krishna and Semple, 2000) up to 1000 Hz (Langner, 1981). Despite this diversity, there is general agreement among authors that the ability of neurons to synchronize to fast amplitude modulations decreases along the ascending auditory pathway (Rees and Møller, 1983; Langner and Schreiner, 1988). Moreover, at the level of the inferior colliculus (IC), neurons have been found that exhibit bandpass shaped rMTFs (Rees and Møller, 1983; Langner and Schreiner, 1988; Rees and Palmer, 1989; Krishna and Semple, 2000). In contrast to CN neurons, these neurons are selective to a certain modulation frequency and respond with a maximum spike rate when stimulated with their so-called best modulation frequency (BMF). Furthermore, rate BMFs have been found to be topographically organized in the central nucleus of the IC covering a BMF range of 10–1000 Hz (Schreiner and Langner, 1988, 1997).

Three biologically motivated models have been presented previously, suggesting different mechanisms of how temporal AM information can be transformed into a rate-based representation between the CN and IC. The neural circuit suggested by Langner (1981) (see also Langner and Schreiner, 1988; Langner, 1997a, b) consists of three functional units at the model CN stage, a trigger unit, a buildup circuit, and an oscillator circuit, with each of these units being associated with a specific neural response type in the CN. At the model IC stage, a coincidence detector unit is located that represents the model output unit. According to Langner (1981), the response types and the neural connectivity of his model units result in a bandpass shaped rMTF of the IC coincidence detector unit. In order to yield a number of bandpass rMTFs with different BMFs, a number of such neural circuits are needed that differ in the integration time constant of the CN buildup circuit. Based on the specific assumptions in his model, Langner (1997a, b) predicts that IC bandpass units can only reveal BMFs that are integer multiples of the stimulus carrier frequency. However, such a relationship between the BMFs of IC neurons and the stimulus carrier frequency could not be confirmed in electrophysiological recordings (Krishna and Semple, 2000). Since the model has not been tested extensively, it is difficult to predict whether a physiologically plausible range of integration time constants used within the buildup circuit of the CN model stage can account for the large range of BMFs observed in the data (Schreiner and Langner, 1988, 1997).

The second biologically motivated model describing the transformation of temporal periodicity information into a rate-based representation was suggested by Hewitt and Meddis (1994). The CN stage of their model employs populations of simulated chopper units (Hewitt *et al.*, 1992) that are morphologically associated with stellate cells in the ventral cochlear nucleus (VCN) (Rhode *et al.*, 1983; Smith and Rhode, 1989). While the chopping frequencies of units within one population are identical, the chopping frequencies among different populations of chopper units differ. Coincidence detector units are located at the model IC stage, with each coincidence detector receiving convergent input from one population of chopper units. If an amplitude modulated stimulus is presented to the model, the chopper units whose chopping frequencies resemble the modulation frequency of the stimulus synchronize best to the stimulus modulation. Thus, this population of chopper units provides a highly synchronized input to its subsequent IC coincidence detector. In contrast, the activity within populations of units whose chopping frequencies differ from the stimulus modulation is less synchronized. The IC coincidence detectors that are driven by these populations receive less synchronized input and respond with a lower spike rate. The model by Hewitt and Meddis (1994) therefore transforms the bandpass shaped tMTFs observed in VCN chopper units into bandpass shaped rMTFs. The different BMFs of different IC coincidence detector units hereby result from different chopping frequencies among the populations of chopper units. The different BMFs are thus obtained from changing a temporal parameter of the simulated chopper units. However, it is not clear whether the range of BMFs (150–700 Hz) found in the tMTFs of VCN chopper units (e.g., Frisina *et al.*, 1990; Rhode and Greenberg, 1994) matches the range of BMFs observed in IC recordings (Schreiner and Langner, 1988, 1997).

The most recent biologically motivated model describing the formation of bandpass shaped rMTFs in IC neurons was suggested by Nelson and Carney (2004). The CN stage of their model employs a VCN bushy cell that receives undelayed excitatory AN input and delayed inhibitory AN input. The model bushy cell projects onto the model IC neuron in the same manner, i.e., it provides undelayed excitatory input and delayed inhibitory input. The interaction of excitatory and inhibitory synaptic inputs results in a bandpass shaped rMTF at the model IC neuron if the excitatory synapses are assumed to be faster and weaker than the inhibitory synapses. Different BMFs within different IC model neurons are obtained by varying the time courses of the synaptic inputs. Thus, similar to the models by Langner (1981) and by Hewitt and Meddis (1994), different BMFs are gained from systematic changes in the temporal model parameters. The IC stage of the model by Nelson and Carney (2004) does not seem to be critical for the generation of bandpass shaped rMTFs since these would already result at the CN model unit

if the same parameters were used for its excitatory and inhibitory synaptic inputs as for the IC model unit.

In contrast to the existing models where temporal model parameters are systematically changed in order to obtain different BMFs within different model units, the modeling approach presented here employs a nontemporal mechanism and therefore differs conceptually from the previous approaches. As in the previous modeling studies, the present model reflects a biologically motivated circuit based on simulated unit responses that can be associated with neural response types. For example, simulated responses of ideal onset (O_I) units, a response type observed in the posteroventral cochlear nucleus (PVCN), are chosen to provide the temporal information in the present model. This is due to the so-called "entrainment" property, i.e., the generation of exactly one spike per stimulus period, as found in O_I units in response to AM stimuli with modulation frequencies below about 800 Hz (e.g., Rhode and Smith, 1986; Rhode, 1994).

Another focus of this study is related to the nonlinear processing inherent in the spike processes of a neural model. In contrast to models that deal with the simulation of continuous, average (population) spike rates or spike probabilities (like, e.g., Nelson and Carney, 2004), the process of spike generation and the simulation of neural responses on a spike level involves strong nonlinearities. The continuous fluctuations in the stimulus waveform or envelope are thereby transformed into a binary pulse or spike code, subjected to a neural circuit and are transformed back to a spike rate reflecting again a continuous quantity. Since the process that establishes the bandpass shaped rMTFs in the present model is realized by a simulation of spikes, it is not clear in advance to what extent the bandpass rMTFs behave like bandpass filters in complex stimulus conditions. A bandpass filtering operation in the envelope domain, as suggested in psychoacoustical studies in humans (e.g., Dau *et al.*, 1997a; Ewert and Dau, 2000) would imply properties of a limited resolution spectral decomposition of the stimulus' envelope.

In the first part of the study, the model is presented and the simulated responses of the individual stages are described and explained, whereby simple stimuli are used, such as high-frequency tone bursts and sinusoidally amplitude-modulated (SAM) tones. It will be shown that the model can account for several key observations in the experimental response properties of the different units, as described in the literature. Specifically, the model accounts for the transformation of temporal AM information into a rate-based respresentation of stimulus periodicity. The model is tested with SAM tones of different stimulation levels and modulation depths. In the second part of the study, the model is tested with complex stimuli, such as pure tones modulated by the sum of two sinusoids, narrowband noises, and iterated rippled noise (IRN). The simulated responses to complex envelope waveforms might reveal important information in two ways. First, since the neural circuit is highly nonlinear, it is not clear if the model is able to separate different components in the envelope spectrum as it would be expected from a linear modulation filterbank analysis. Second, complex envelope waveforms have not yet been systematically investigated in physiological studies. The model thus might be useful to validate hypotheses about the processing of these complex sounds in units that exhibit a bandpass rMTF.

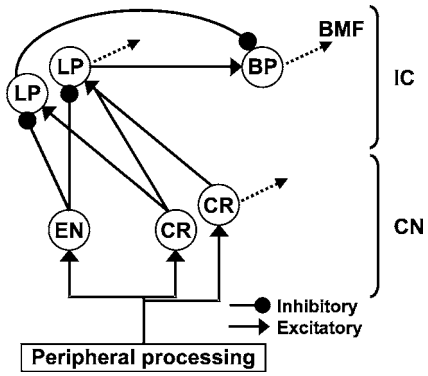


FIG. 1. The architecture of the present model. The model is a cascade of three stages: (i) the peripheral processing stages (bottom), (ii) the CN stage that includes an entrainment (EN) unit and a number of constant-rate (CR) units (middle) and (iii) the IC stage of the model (top) that includes low-pass (LP) and bandpass (BP) units. Excitatory synaptic connections between model units are represented by arrows and inhibitory synapses by filled circles.

II. MODEL

A. General model structure

Here, an overview of the model structure and a description of the function of the circuit are provided. More detailed information about the individual stages of the model are given in Sec. II B. Figure 1 shows a schematic representation of the model of the present study. The model is a cascade of three stages: (i) peripheral processing (bottom), (ii) CN processing (middle), and (iii) IC processing (top). Excitatory synaptic connections are represented by arrows while inhibitory synaptic connections are represented by filled circles. The peripheral processing stage includes basilar-membrane filtering, inner hair-cell transduction and a simulation of AN activity. The CN stage of the schematic model circuit contains three neurons, each of which receives excitatory input from the ascending AN fibers: an entrainment (EN) neuron and two constant-rate (CR) neurons. All model CN units project onto the IC stage of the model that consists of three different neurons, two low-pass (LP) units and one bandpass (BP) unit. Each LP unit of the model IC stage receives inhibitory input from the EN unit and excitatory input provided by a different number of CR units of the CN stage. Hereby, the first LP unit receives input from one CR unit and the second LP unit receives input from both CR units.

In order to illustrate the basic function of the circuit, Fig. 2 shows the expected idealized rMTFs of each model unit in response to 100% SAM tones at a carrier frequency equal to the characteristic frequency of the neural circuit. The idealized rMTF of the EN model unit is shown in the lower left panel of Fig. 2. Here, the spike rate of the EN unit corresponds to the stimulus modulation rate up to 800 Hz. Thus, in response to modulations of rates below 800 Hz, the EN unit entrains to the stimulus modulation, emitting exactly one spike per stimulus period. If the modulation frequency exceeds 800 Hz, the EN unit can no longer follow the fast

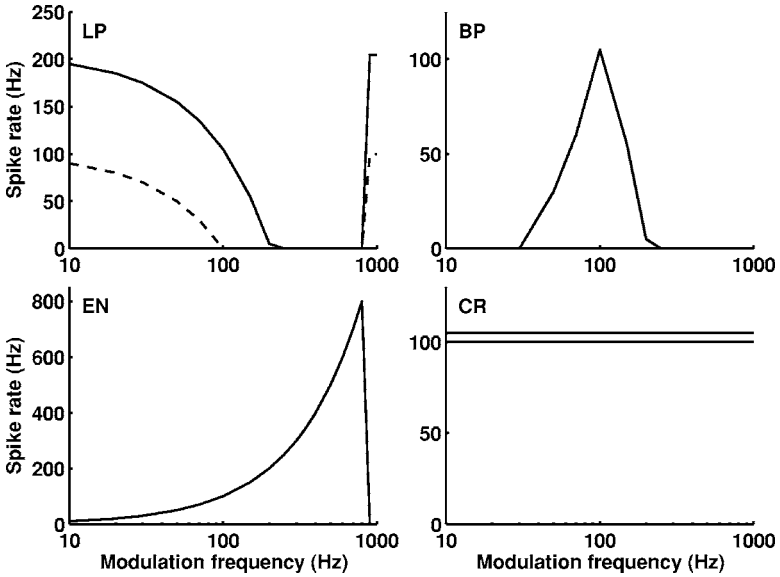


FIG. 2. Schematic rMTFs of the different model units: CN entrainment (EN) unit (lower left panel), CN constant-rate (CR) units (lower right panel), IC low-pass (LP) units (upper left panel) and IC bandpass (BP) unit (upper right panel).

stimulus modulation. In this case, the unit generates only one spike at stimulus onset. The lower right panel of Fig. 2 shows the idealized rMTFs of the two CR units in the CN stage. The two lines in the figure are shown at slightly different levels for visualization purposes. Both CR units exhibit a flat rMTF, i.e., their spike rate (of about 100 Hz in this example) does not depend on the stimulus modulation frequency.

The upper left panel of Fig. 2 illustrates the resulting rMTFs of the two LP units. Within a certain range of modulation frequencies, the activity of both LP units is suppressed, due to the inhibitory contribution provided by the EN unit of the CN stage. As the spike rate of the EN unit reflects the stimulus modulation frequency, it provides little inhibition at low modulation frequencies, and the spike rate of the LP units reflects the excitatory contribution from their CR input units. With increasing modulation rate, the inhibition provided by the EN model unit increases until it exceeds the excitatory contribution from the CR units and suppresses the activity of the LP unit. The activity of the LP unit remains suppressed until the modulation frequency exceeds 800 Hz when the EN unit fails to entrain and instead exhibits an onset response. At such high modulation frequencies, the activity of the LP unit again reflects the excitatory input of its CR input units. The activity of the first LP unit (dashed line), receiving excitatory input from only one CR neuron, becomes suppressed if the stimulus modulation exceeds 100 Hz. In contrast, the second LP unit (solid line), that receives excitatory input from both CR neurons, is not suppressed below a stimulus modulation frequency of 200 Hz.

The model output unit is the BP unit of the IC stage, which receives inhibitory input from the first LP unit (dashed line) and excitatory input from the second LP unit (solid line). Within the model, the inhibitory LP unit input is specified to have a greater synaptic efficiency than the excitatory LP unit input. This results in a bandpass shaped rMTF of the model BP units as shown in the upper right panel of Fig. 2. Since the modulation frequency at which an individual LP unit becomes suppressed depends on how many CR units project onto that unit, different BMFs of the model BP units are obtained from varying the number of CR input units in the present neural circuit. Thus, a rate-based representation of AM is obtained by varying the number of CR units providing input to the different LP model units.

B. Processing stages of the model

1. Auditory periphery

Basilar-membrane filtering is simulated using a bank of fourth-order gammatone filters (Patterson *et al.*, 1988) with center frequencies spaced according to the equivalent rectangular bandwidth scale. A more sophisticated, nonlinear filterbank was not used in the current study since the effect of peripheral nonlinearity on amplitude modulation processing was assumed to be not critical for the functioning of the neural circuit proposed here. The output of each gammatone filter serves as input to Meddis' model of the mechanical-to-neural transduction at the hair-cell-auditory nerve synapse (Meddis, 1986, 1988; Meddis *et al.*, 1990). The output is the instantaneous AN discharge rate function in response to arbitrary stimuli. The discharge rate at a particular point in time can be considered as being proportional to the probability that a spike is generated. Since the frequency range and the number of AN fibers projecting onto the two types of CN model units differ largely, the AN input of these units is computed separately, as described below.

2. General neuron model dynamics

All model units are simulated by a point neuron model with an integrate-to-threshold dynamic (Tuckwell, 1988) using a modified version of the MacGregor point neuron model (MacGregor, 1987; Arle and Kim, 1991). In order to simulate the distinct temporal response properties of the model EN unit, this unit has some specific additional properties which are described in Sec. II B 3. Specific details about the CR unit are given in Sec. II B 5.

In the point neuron model, each presynaptic spike arriving at a model unit causes a postsynaptic current $I(t)$ that contributes to the membrane potential $V_m(t)$ of the cell. For

TABLE I. Simulation parameters of the CR, LP, and BP model units.

Parameter	Description	CR	LP	BP
V_{rest}	Membrane equilibrium potential	−60 mV	−60 mV	−60 mV
τ_m	Membrane time constant	3 ms	10 ms	5 ms
τ_k	Potassium time constant	1.5 ms	1.5 ms	1.5 ms
G	Total resting conductance	50 nS	50 nS	50 nS
b	Potassium conductance change	0.1 mS	50 μS	50 μS
Θ_{act}	Spike threshold	−50 mV	−40 mV	−40 mV
τ (ex)	EPSC time constant	0.5 ms	0.5 ms	0.5 ms
τ (in)	IPSC time constant		1 ms	3 ms
g_0 (ex)	Conductance of an excitatory synapse	0.01 nS	0.06 nS	0.06 nS
g_0 (in)	Conductance of an inhibitory synapse		0.9 nS	1.2 nS

the CR, LP, and BP units, the shape of the postsynaptic current in response to a presynaptic spike is described by an "alpha function"

$$h_\alpha(t) = \begin{cases} g_0 \frac{t}{k} e^{\left(-\frac{t}{\tau}\right)} & \text{for } t > 0 \\ 0 \quad \text{otherwise.} \end{cases} \tag{1}$$

A spike arriving at the unit causes an excitatory postsynaptic current if the synaptic conductance $g_0 > 0$, or an inhibitory postsynaptic current if $g_0 < 0$. The time constant τ determines the duration of the postsynaptic current, and $k = 0.02$ ms represents a normalizing factor. The postsynaptic current $I(t)$ is calculated by a convolution of $h_\alpha(t)$ and a (summed) presynaptic spike activity pattern in units of volts. The voltage change $dV_m(t)/dt$ of the leaky-integrator neuron caused by the flow of synaptic current through the membrane is modeled by a parallel resistor-capacitor electrical circuit with a membrane time constant τ_m:

$$\tau_m \frac{dV_m(t)}{dt} = -V_m(t) + V_k(t) + RI(t) \quad \text{for } t > 0. \tag{2}$$

Here, τ_m equals RC, with R representing the resistance and C the capacitance of the circuit. $V_k(t)$ simulates a dynamically changing potassium conductance, g_k, reflecting the unit's refractoriness, with

$$\tau_k \frac{dg_k(t)}{dt} = -g_k(t) + bs \tag{3}$$

and

$$V_k(t) = -\frac{g_k(t)}{G} V_m(t). \tag{4}$$

Following each spike of the unit represented by its state variable s, the potassium conductance is increased by a constant amount b and decays with the potassium time constant τ_k. $G = 1/R$ is the total resting conductance. The unit's state variable s follows the relation

$$s = \begin{cases} 0 & \text{for } V(t) \leqslant \Theta_{act} \\ 1 & \text{for } V(t) > \Theta_{act}, \end{cases} \tag{5}$$

with $V(t) = V_{rest} + V_m(t)$, where V_{rest} represents the resting potential of the unit. As long as the cell's membrane potential, $V(t)$, is below its spike threshold Θ_{act}, the unit's state variable s is set to zero.

Besides the relative refractory period of the model units, represented by the spike-induced increase in potassium conductance, the unit also incorporates an absolute refractory period of 1.5 ms, during which no further spikes can be emitted by the unit. The specific parameters of the CR, LP and BP units are summarized in Table I.

3. Entrainment units in the CN

The EN unit of the model CN stage provides the temporal periodicity information which is transformed into a rate-based periodicity representation by the present neural circuit. Thus, the quality of the overall model depends crucially on the reliability and temporal precision of the simulated EN unit responses. The EN unit is simulated using a functional point-neuron model similar to the one suggested by Dicke and Dau (2005). In contrast to the other model units described above, this functional model disregards changes in the potassium conductances following a spike. The model describes ideal onset responses as observed in octopus cells in the ventral cochlear nucleus (VCN). Octopus cells have been found to entrain to the stimulus modulation (Rhode and Smith, 1986; Rhode, 1994), when stimulated with SAM pure tones with modulation frequencies <800 Hz. Due to their precisely timed action potentials these cells have been assumed to be the best temporal encoders of AM stimuli among all major cell types in the cochlear nucleus (Frisina *et al.*, 1990; Rhode, 1994; Winter and Palmer, 1995).

In the model, it is assumed that a large number of AN fibers converge on a CN model unit (Dicke and Dau, 2005). Therefore, the deterministic discharge rate function is used as the time-dependent AN input to the EN unit, i.e., no individual spike trains for each AN fiber are computed. In order to account for the wide across-frequency input that octopus cells receive (Oertel *et al.*, 2000), the simulated AN activity is summed across a range of cochlear filters, such that each model unit receives equally weighted input from 11 cochlear filters spanning a frequency range of approximately one octave centered at the unit's CF.

In contrast to the dynamics of the other model units described earlier, the EN unit has a biphasic postsynaptic current instead of the monophasic "alpha function,"

TABLE II. Simulation parameters of the EN model unit.

Parameter	Description	EN
V_{rest}	Membrane equilibrium potential	−60 mV
τ_m	Membrance time constant	1 ms
Θ_{act}	Spike threshold	−45 mV
τ_a	Effective PSC time constant	0.1 ms
τ_b	Effective PSC time constant	0.2 ms
c	Constant PSC factor	0.25
Θ_{rel}	Spike blocking release threshold	−59 mV

$$h_b(t) = \begin{cases} g_0 \frac{t}{k}\left[e^{\left(-\frac{t}{\tau_a}\right)} - ce^{\left(-\frac{t}{\tau_b}\right)}\right] & \text{for } t > 0 \\ 0 \quad \text{otherwise.} \end{cases} \tag{6}$$

As described in Dicke and Dau (2005), the biphasic membrane impulse response realizes a high-pass filtering of the synaptic input, i.e., stimulus transients and fast changes are enhanced relative to sustained stimulus portions. With this feature, the EN model unit effectively acts as a change detector on its synaptic input. The model also uses a spike-blocking mechanism (e.g., Romand, 1978; Dicke and Dau, 2005) that prevents the unit from emitting multiple spikes within one period of a periodic stimulus. Following each spike, the EN unit enters a spike-blocked state during which it cannot exhibit further action potentials. This spike-blocked state lasts until its potential falls below a so-called spike-blocking release threshold Θ_{rel} that releases the cell from spike blocking. Additionally, the EN unit possesses an absolute refractory period of 0.8 ms during which it cannot emit further action potentials. The combination of the biphasic membrane impulse response and the spike-blocking mechanism has been shown to account for the simulation of entrainment to periodic stimuli (Dicke and Dau, 2005). The values of the EN unit parameters are given in Table II.

4. Variation of entrainment unit parameters

With the choice of the EN unit parameters as given in Table II, the units show a step transition from no encoding of modulation depths below 30% to undistinguishly encoding modulation depths above 40%. In order to allow for the encoding of a large dynamic range of modulation depths in the present model, a set of ten different EN units were used that differ in their sensitivity to different modulation depths. This was achieved by varying their spike threshold Θ_{act} and their spike-blocking release threshold Θ_{rel}. While the spike threshold Θ_{act} determines the membrane potential at which the unit is sufficiently depolarized to generate an action potential, the release threshold Θ_{rel} determines the hyperpolarized potential at which the unit is released from spiked blocking (see Fig. 1 in Dicke and Dau, 2005). The threshold potentials chosen for the ten EN units of the present model are given in Table III.

The ten EN units introduce an additional (hidden) layer in the model topology, shown in Fig. 1: The circuit can be considered as being repeated for each of the ten EN units and the final model output is obtained by summing the activity of BP units with identical BMFs along the newly introduced "modulation-depth," or EN, layer. This summed activity is finally normalized in order to obtain the model output henceforth referred to as the relative spike rate of the filters. In principle, the same summation could also be considered for the LP units but is not explicitly addressed here.

TABLE III. Parameter variation of the EN model unit. Pairs of the spike threshold, Θ_{act}, and the spike blocking release threshold, Θ_{rel}, are given. All other parameters were as in Table II.

EN unit	Θ_{act}(mV)	Θ_{rel}(mV)
1	−58.5	−63.5
2	−57.2	−65.0
3	−54.1	−67.5
4	−51.0	−69.5
5	−48.5	−71.0
6	−43.0	−74.2
7	−37.0	−76.5
8	−31.0	−81.5
9	−21.5	−87.0
10	−9.5	−93.0

The introduction of the modulation-depth layer to the model can be viewed as a simplified realization of a population of adaptive EN units that change their parameters in response to the stimulus modulation depth. The rate response of the population would increase monotonically as the modulation depth increases. The simplification of ten EN units with specific choice of parameters reflects a discrete approximation to a neural population. It does not imply that an ordered array of EN units with different thresholds exists. In accordance to the adaptive EN units which provide an increasing population rate response as a function of modulation depth, one would need to assume an increasing number of CR units that are activated with increasing modulation depth of the stimulus.

5. Constant-rate units in the CN

While the EN unit provides the model with the temporal AM information, the CR units provide the mechanism by which the modulation frequency selectivity of the IC units is shaped. The CR units of the present model are simulated using a point-neuron model as described in Sec. II B 2, adjusted to mimic a sustained chopping response as observed in VCN stellate cells. Each CR model unit receives input from 50 AN fibers, all originating from the same peripheral filter channel. The individual spike train of each AN fiber is obtained by comparing the instantaneous hair-cell discharge probability with a random number equally distributed in the range $[0 \ldots 1]$. If the instantaneous spike probability exceeds the random number, a spike is generated and the AN fiber enters a refractory period of 1 ms.

In the following, predictions obtained with the model circuit will be presented using pure-tone and amplitude modulated stimuli.

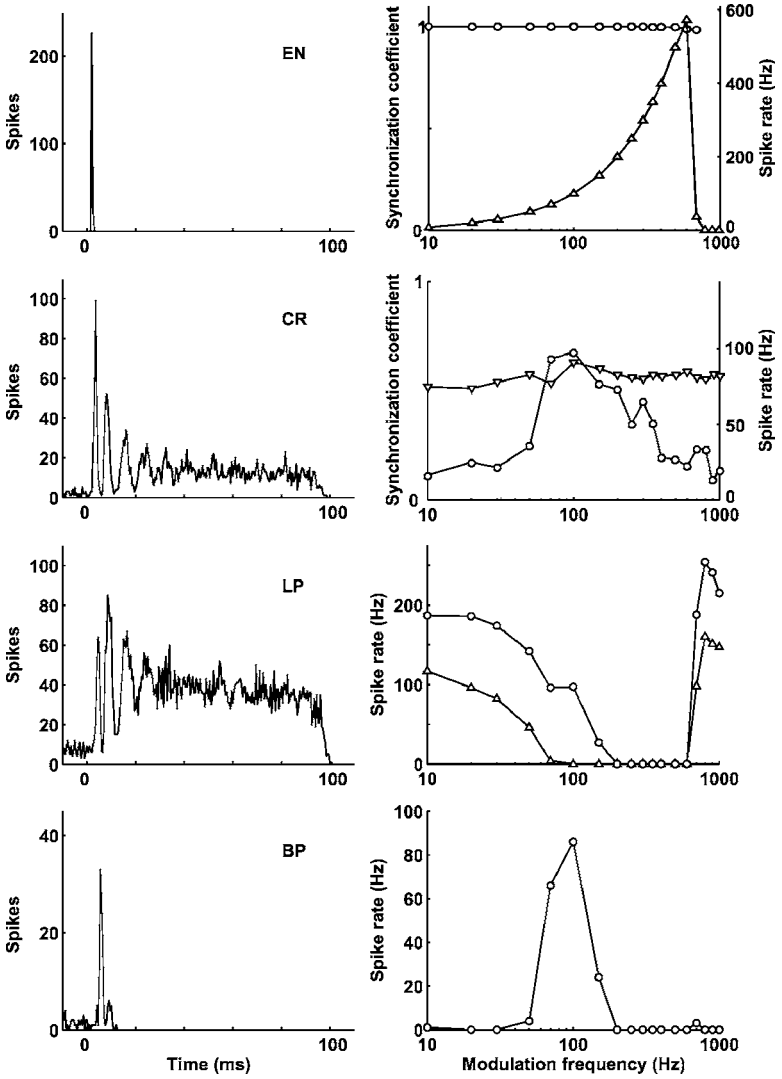


FIG. 3. Left: PSTHs of the different model units in response to a 10 kHz tone burst presented at 60 dB SPL. Right: rMTF and synchronization coefficient (upper two panels) of the different model units (indicated by the abbreviations in the left panels) in response to 100% SAM 10 kHz tones with varying modulation frequencies presented at 60 dB SPL. The upper row shows responses for the EN model unit. The MTFs of the EN unit indicate a nearly perfect entrainment of the unit to the stimulus modulation. The second row shows the CR model unit response. The spike rate of the unit is independent of the modulation frequency of the stimulus. The third row shows the response of the LP model unit receiving input from four CR units. The LP unit reflects the chopping of its input units (left). On the right the rMTF of the LP unit receiving input from eight CR units (circles) and the rMTF of the LP unit receiving input from four CR units (triangles) is shown. The region of suppression occurs at different modulation frequencies in the two units. The bottom row shows the responses of the BP model unit receiving excitatory input from the LP unit with eight CR inputs and inhibitory input from the LP unit with four CR inputs. The unit reveals an onset response (left). However, compared to the EN unit, the BP unit responds with a larger first-spike latency and less temporal precision. The rMTF on the right reveals a BMF of 100 Hz.

III. SIMULATIONS

All simulations were performed within the neural circuit that receives input from the peripheral channel centered at 10 kHz. For the analysis of the different unit types, the neural circuit was tested using sinusoidally amplitude modulated (SAM) 10 kHz tones. The degree of synchronization of the model response to a SAM tone was analyzed using the synchronization coefficient (SC; Goldberg and Brown, 1967), computed as the vector strength of the distribution of spikes. A SC value of 1 indicates perfect synchronization to the stimulus modulation, while SC=0 represents no synchronization at all. Additionally, the rMTF where the spike rate is plotted as a function of the modulation frequency was computed in response to SAM stimuli. SAM stimuli were presented for 1.05 s with the initial 50 ms of a unit's response being excluded from the analysis in order to avoid onset effects. The response types of the different model units are characterized by their peristimulus time histograms (PSTHs) yielded in response to 100 ms duration, 10 kHz tone bursts presented at 60 dB sound pressure level (SPL). PSTHs were gained from 100 stimulus representations of the high-frequency tone burst and analyzed in 0.5 ms bins. All simulations of the present study were performed using discrete time signal processing at a sampling rate of 50 kHz. For the analysis of the BP units, the response to complex sounds (such as two-component AMs, narrowband noise, and iterated rippled noise) as well as different modulation depths were considered. If not stated otherwise, the stimuli were presented at a level of 60 dB SPL.

A. Response characteristics of the different units

First, the output of the different unit types is shown exemplarily for the EN layer with the parameters given in Table II. The upper right panel of Fig. 3 displays the modulation transfer functions of the EN unit in response to a 100% SAM 10 kHz tone. The spike rate (triangles, right scale) of the EN unit reflects the stimulus modulation rate up to 600 Hz. Above 600 Hz, the EN unit cannot follow the fast fluctuations of the modulated stimulus anymore and exhibits purely an onset response to the stimulus. The figure also shows the synchronization coefficient (circles, left scale) of the EN unit that lies between 0.99 and 1 for modulation frequencies below 600 Hz. The high synchronization coefficient, together with a spike rate that reflects the stimulus modulation rate, indicates a nearly perfect entrainment of the unit to the stimulus modulation.

The upper left panel of Fig. 3 displays the PSTH of the EN model unit. The unit exhibits one spike at stimulus onset that is followed by no further activity during the sustained portion of the stimulus due to the spike-blocking mechanism of the model (Dicke and Dau, 2005). The PSTH of the EN model unit can be associated with a cochlear nucleus ideal onset (O_I) unit. These units have been found to respond to pure-tone stimuli ($f>2$ kHz) revealing one sharply timed action potential at stimulus onset followed by little or no (<10 spikes/s) subsequent activity during the sustained portion of the tone (Godfrey *et al.*, 1975; Rhode and Smith, 1986). Thus, in terms of its onset PSTH and its entrainment to SAM stimuli, the response properties of the model EN unit correspond well to the responses observed in PVCN octopus cells.

The right panel in the second row of Fig. 3 shows the modulation transfer functions of a model CR unit in response to a 100% SAM 10-kHz tone. Despite small fluctuations in the spike rate (triangles, right scale), resulting from stochastical fluctuations in the simulated AN activity, the rMTF reveals that the spike rate of the CR neuron (85–95 Hz) does not depend on the modulation frequency of the stimulus. The figure also shows the corresponding synchronization coefficient (circles, left scale). According to its tMTF, the CR unit synchronizes best to a modulation frequency close to 100 Hz, while its synchronization to most other modulation frequencies is poor. The increased synchronization at a modulation frequency of about 100 Hz can be explained from the chopping property of the unit that leads to maximum synchronization when stimulus frequency and intrinsic chopping frequency correspond to each other, while synchronization is strongly decreased if the frequencies do not match.

The corresponding left panel displays the simulated pure-tone PSTH of the CR model unit. It shows that the unit responds with a preferred interspike interval of approximately 10 ms (reflecting the inverse of the chopping rate), which is not related to the periodicity of the 10-kHz tone. The CR unit thus displays a so-called sustained chopping response. The simulated PSTH can be associated with stellate cells in the VCN that have been found to reveal a chopping response when stimulated by a pure tone with an individual chopping frequency that is not related to the stimulus frequency (Rhode *et al.*, 1983; Smith and Rhode, 1989). Moreover, recorded rMTFs of stellate cells have been found to be flat, while their tMTFs are low-pass shaped at low stimulus levels and become increasingly more bandpass-like as the stimulus level increases (Frisina *et al.*, 1990; Kim *et al.*, 1990; Rhode, 1994). Thus, the output of the CR model unit appears to correspond to VCN stellate cell responses in terms of its pure-tone PSTH and its MTFs.

Each LP unit of the model IC stage receives excitatory synaptic input from a different number of CR model units and inhibitory synaptic input from the EN model unit (see Fig. 1). Apart from the different number of CR unit inputs, all LP units of the present model are identical. The right panel in the third row of Fig. 3 shows the rMTFs of a model LP unit that receives 8 CR inputs (circles) in response to 100% SAM 10-kHz tones. This rMTF reveals that the activity of the LP unit is suppressed at modulation frequencies >200 Hz. This is due to the inhibition provided by the EN model unit that increases with increasing modulation frequency (Fig. 3) until it exceeds the excitatory input from the CR units. At modulation frequencies exceeding 800 Hz, where the EN unit exhibits a purely onset response, the activity of the LP unit is no longer suppressed, and it responds with a spike rate that reflects its CR inputs. The rMTF of the model LP unit receiving input from only four CR units (triangles) is also shown for comparison. The activity of this LP unit already becomes suppressed at modulation frequencies between 70 and 100 Hz, due to the smaller number of CR units providing input to this unit.

The corresponding left panel displays the pure-tone PSTH of the LP unit receiving four CR inputs as an example. The LP model unit reveals a sustained PSTH that reflects the chopping found in its CR input, while the EN unit only provides input at the tone onset due to its ideal onset response to high-frequency tones. Electrophysiological recordings from IC neurons using 100% SAM tones have revealed IC rMTFs that partly resemble the simulated LP rMTFs in their distinct shape (Krishna and Semple, 2000). Krishna and Semple (2000) investigated the modulation depth dependency of such IC rMTFs and defined the "region of suppression" as a range of modulation frequencies where the spike rate of a neuron is decreased. IC neurons that show a region of suppression in their rMTFs were found to exhibit sustained pure-tone PSTHs (Krishna and Semple, 2000). The simulations of the pure-tone responses resulting from the model LP unit are consistent with their finding.

Each BP unit of the model IC stage receives weak excitatory synaptic input from one model LP unit and strong inhibitory input from an adjacent model LP unit (see Fig. 1). Compared to the LP unit that provides the excitatory BP input, the LP unit providing the inhibitory BP input receives a smaller number of CR inputs. The bottom right panel of Fig. 3 displays the rMTFs of the model BP unit that receives input from the two LP units shown before, in response to a 100% SAM 10-kHz tone. Weak excitatory input is provided by the LP unit receiving input from eight CR units (third row, circles), while strong inhibitory input is provided by the LP unit receiving input from four CR units (third row, triangles). While the inhibitory LP unit suppresses the activity of the BP unit at low modulation frequencies, it provides no inhibitory contribution at modulation frequencies $\geqslant 100$ Hz. However, the contribution of the excitatory LP unit is not suppressed until the modulation frequency reaches 200 Hz, resulting in a bandpass shaped rMTF of the BP unit. At modulation frequencies >800 Hz, where both LP units are exclusively driven by the CR input, the contribution of the inhibitory LP unit compensates for the excitatory LP unit input such that the BP unit reveals no response.

The bottom left panel of Fig. 3 shows the pure-tone PSTH of the model BP unit. In contrast to the sustained PSTH found in the model LP units (third row, left panel), the model BP unit reveals an onset PSTH. This agrees with data recorded by Krishna and Semple (2000) who found that IC neurons that reveal bandpass shaped rMTFs often possess an onset PSTH.

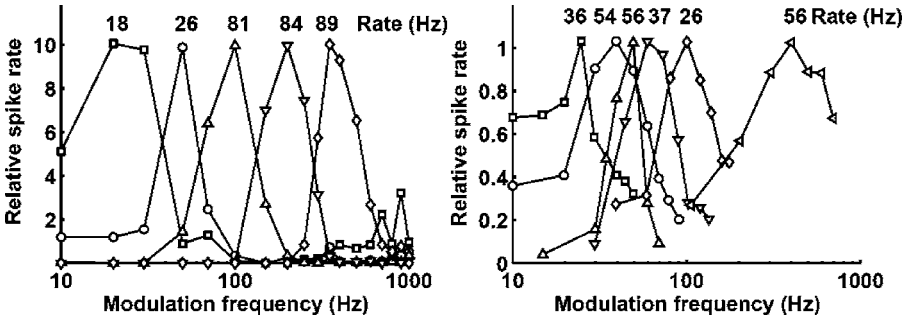


FIG. 4. Left: Simulated rMTFs of the five model BP units in response to 100% SAM 10-kHz tones with varying modulation frequencies. Each rMTF is normalized by its BMF spike rate (given above the rMTFs) that corresponds to the spike rate at the peak of the unit's rMTF. The BMF rate used to gain the relative spike rate from the summed activity of all ten model BP units with identical BMFs is shown above each filter. Right: Recorded IC rMTFs in response to 100% SAM tones with carrier frequencies that correspond to the individual CF of the recorded unit. Stimuli were presented 30–60 dB above the individual response threshold of a unit. Each rMTF is normalized by its BMF spike rate that is given above each rMTF. Replot with permission from Langner and Schreiner (1988).

B. Characteristics of modulation bandpass filters

In order to obtain bandpass shaped rMTFs with BMFs that cover a range of modulation frequencies, $2^n=32$ CR units ($n=5$) were simulated within the same peripheral filter (CF=10 kHz). The 2^n CR units project onto $n+1$ different LP units, each of which receives input from a different number of CR units, ranging from 2^0 to 2^n. In this way, n model BP units can be obtained that reveal bandpass shaped rMTFs with n different BMFs. The BP units analyzed in the following are the result of summation across the EN layer described in Sec. II B 4. The left panel of Fig. 4 displays the simulated bandpass rMTFs of the five BP units in response to 100% SAM 10 kHz tones. Each rMTF is normalized by its individual maximum response rate (i.e., its BMF spike rate). The value of the corresponding BMF spike rate is shown above each rMTF. The right panel of Fig. 4 displays data obtained in IC recordings replotted from Langner and Schreiner (1988). The recorded data were obtained from 100% SAM tones with carrier frequencies that corresponded to the individual CF of the recorded unit, presented between 30 and 60 dB above the unit's individual response threshold.

The quality factor of a bandpass rMTF can be expressed as the BMF divided by the bandwidth of the rMTF 6 dB below its peak ($Q_{6\ \mathrm{dB}}$). The average $Q_{6\ \mathrm{dB}}$ value of 0.8 ± 0.3 reported by Langner and Schreiner (1988) differs from the simulated average $Q_{6\ \mathrm{dB}}$ values given in Table IV. However, the simulated $Q_{6\ \mathrm{dB}}$ values are in the range of the $Q_{6\ \mathrm{dB}}$ values derived from psychophysical modulation masking experiments (Dau *et al.*, 1997a, b; Ewert and Dau, 2000). Due to a lack of recorded data, no average $Q_{3\ \mathrm{dB}}$ value (BMF divided by the bandwidth of the rMTF 3 dB below its peak) can be given for the recorded bandpass MTFs. However, simulated and recorded $Q_{3\ \mathrm{dB}}$ values can be compared using the rMTF with a BMF of 100 Hz as an example. The $Q_{3\ \mathrm{dB}}$ value of about 1.8 of the recorded 100 Hz bandpass lies in the region of the $Q_{3\ \mathrm{dB}}$ value of 2.4 of the simulated 100 Hz bandpass.

TABLE IV. $Q_{6\ \mathrm{dB}}$ values of the model filters.

BMF (Hz)	$Q_{6\ \mathrm{dB}}$
20	0.6
50	2.0
100	1.4
200	1.4
350	1.4

C. Encoding of modulation depth and stimulus level

The upper left panel of Fig. 5 displays the relative spike rates of the four 100-Hz BP units that receive input from the EN model units corresponding to the threshold combinations 1, 2, 5, and 10 (see Table III) as a function of AM modulation depth. The relative spike rates are obtained from a 10 kHz tone modulated at a rate of 100 Hz with different modulation depths. Each of the BP units is shown to encode only a limited dynamic range of modulation depths. Within the corresponding 100 Hz filter of the model, the large dynamic range of modulation depths results from the summed activity of all ten BP units. The upper right panel of Fig. 5 displays the relative spike rate of the 100 Hz filter as a function of the modulation depth of a 10 kHz tone, modulated at a rate of 100 Hz. The relative spike rate is essentially a linear function of the stimulus modulation depth.

The lower left panel of Fig. 5 displays the 100 Hz filter transfer function for a 10 kHz tone amplitude modulated with depths of 20% (circles), 50% (triangles), and 100% (squares). The simulated filter functions reveal that neither the BMF nor the overall bandpass shape of the 100 Hz filter are affected by the modulation depth of the stimulus. Estimated $Q_{6\ \mathrm{dB}}$ values are 1.4 for 100%, 1.6 for 50%, and 1.7 for 20% modulation depth.

The lower right panel of Fig. 5 displays the simulated relative spike rate of the 100 Hz filter as a function of the modulation frequency of a 100% amplitude modulated 10 kHz tone, presented at different stimulus levels. Increasing the stimulus level from 40 to 80 dB yields only small changes in the relative spike rate of the 100 Hz filter indicating that the simulated bandpass filters are independent of stimulus presentation level in the tested range. This differs from most of the recorded responses that depend on the stimulation level, as will be further discussed in Sec. IV B.

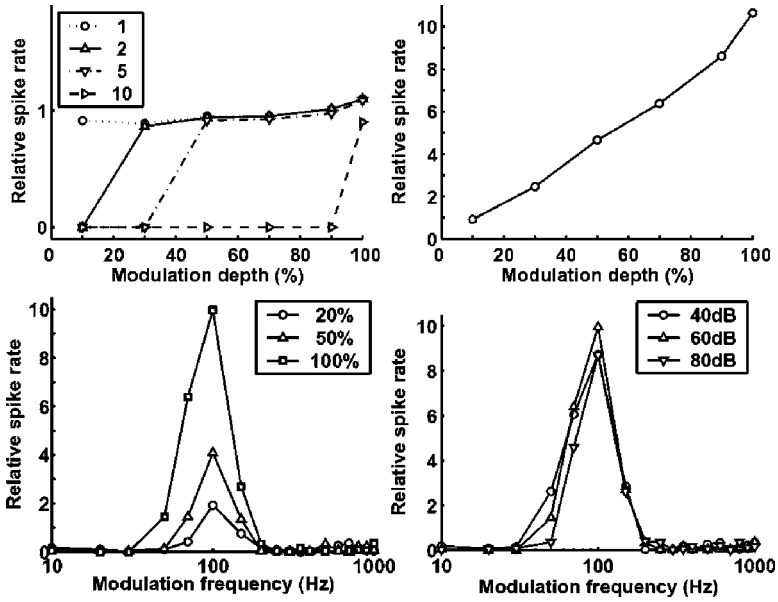


FIG. 5. Upper left panel: Responses of the 100 Hz BP units receiving indirect input from the four EN units with the threshold combinations 1, 2, 5, and 10 (see Table III) as a function of the modulation depth of a 10 kHz tone modulated at a rate of 100 Hz (stimulus presentation level: 60 dB SPL). The simulated spike rate of each BP unit is divided by the BMF rate of the 100 Hz filter. Upper right panel: The relative spike rate of the 100 Hz filter as a function of the modulation depth of the same AM stimulus. Lower left panel: The relative spike rate of the 100 Hz filter in response to amplitude modulated 10 kHz tones with different modulation frequencies and modulation depths (20%, 50%, and 100%). Lower right panel: The relative spike rate of the 100 Hz filter as a function of the modulation frequency of a 100% amplitude modulated 10 kHz tone presented at 40, 60, and 80 dB SPL.

D. Processing of complex envelope waveforms

In the following, it is evaluated to what extent the five modulation filters simulated by the present neural model act like a bank of (linear) filters when applied to complex envelope waveforms that consist of more than a single modulation frequency. Considering the highly nonlinear neural circuit that gives rise to the formation of the bandpass rMTF shape, it cannot be assumed in advance that the model BP units exhibit the frequency resolving properties of a linear bandpass filter when presented with a mixture of amplitude modulation frequencies. However, such a property is commonly associated with modulation filters or the concept of a modulation filter bank.

In order to evaluate to which extent the five modulation filters simulated in the present model can be associated with linear amplitude modulation filters, the simulated relative spike rates obtained in response to complex stimuli were compared to the calculated integrated envelope power at the output of each model filter. This approach is based on the psychophysically motivated envelope-power spectrum model (EPSM, Dau *et al.*, 1999; Ewert and Dau, 2000), which is conceptually similar to the classical power-spectrum model of masking (e.g., Fletcher, 1940; Zwicker *et al.*, 1957) where the signal power at the output of an auditory filter centered at the signal frequency is used to estimate its detectability. In the present study, the amount of the integrated envelope power at the output of a linear filter is calculated (and indicated as "EPSM") and compared to the simulations (indicated as "Sim") obtained with the suggested nonlinear neural circuit.[1] Deviations between the simulations and the EPSM predictions indicate nonlinear effects in the neural circuit.

1. Processing of two-component AM

The amplitude modulated stimuli $s(t)$ used in the simulations consist of a pure-tone carrier with frequency $f_c=10$ kHz that is modulated by the sum of two sinusoids, f_{m1},f_{m2}, with modulation depths $m_1=m_2=0.5$:

$$s(t) = \sin(2\pi f_c t)[1 + m_1 \sin(2\pi f_{m1} t) + m_2 \sin(2\pi f_{m2} t)].$$

Figure 6 displays the temporal envelopes of two complex AM stimuli with $f_{m1}=55$ Hz and $f_{m2}=200$ Hz (upper left panel) and $f_{m1}=18$ Hz and $f_{m2}=200$ Hz (upper right panel). The corresponding envelope spectra, shown in the respective middle panels of the figure, both reveal three spectral components, the dc component (peak at 0 Hz) and two components that correspond to the two imposed modulation frequencies. The amplitudes of the spectral components are given in dB with respect to the dc peak which was chosen to correspond to 0 dB.

The two lower panels of Fig. 6 display the simulated neural activity of the five model filters (circles) in response to the AM stimuli. The output of the EPSM is also shown, represented by the crosses. The simulation results from the neural model largely agree with the EPSM results. Only the result for the 20 Hz filter in response to the AM stimulus with $f_{m1}=55$ Hz and $f_{m2}=200$ Hz (lower left panel) clearly deviates from the EPSM output which will be further addressed in the discussion.

2. Processing of narrowband noise

Narrowband Gaussian noise and low-noise noise of the same bandwidth were used to further characterize the present model. The narrowband Gaussian noise was generated by setting the magnitude of the Fourier components of a broadband Gaussian noise to zero outside the desired passband of 100 Hz. The noise band was centered at 10 kHz. The low-noise noise was generated from a 100-Hz-wide Gaussian noise, centered at 10 kHz, by ten iterations of the following steps: The temporal waveform of the noise was divided by its temporal envelope on a sample-by-sample basis, and the Fourier spectrum of the resulting waveform was restricted to its original bandwidth of 100 Hz. The resulting waveform exhibits a smooth temporal envelope. A more detailed description of this procedure can be found in Kohlrausch *et al.* (1997). The upper row of Fig. 7 displays the temporal envelope of the narrowband Gaussian noise (left) and the corre-

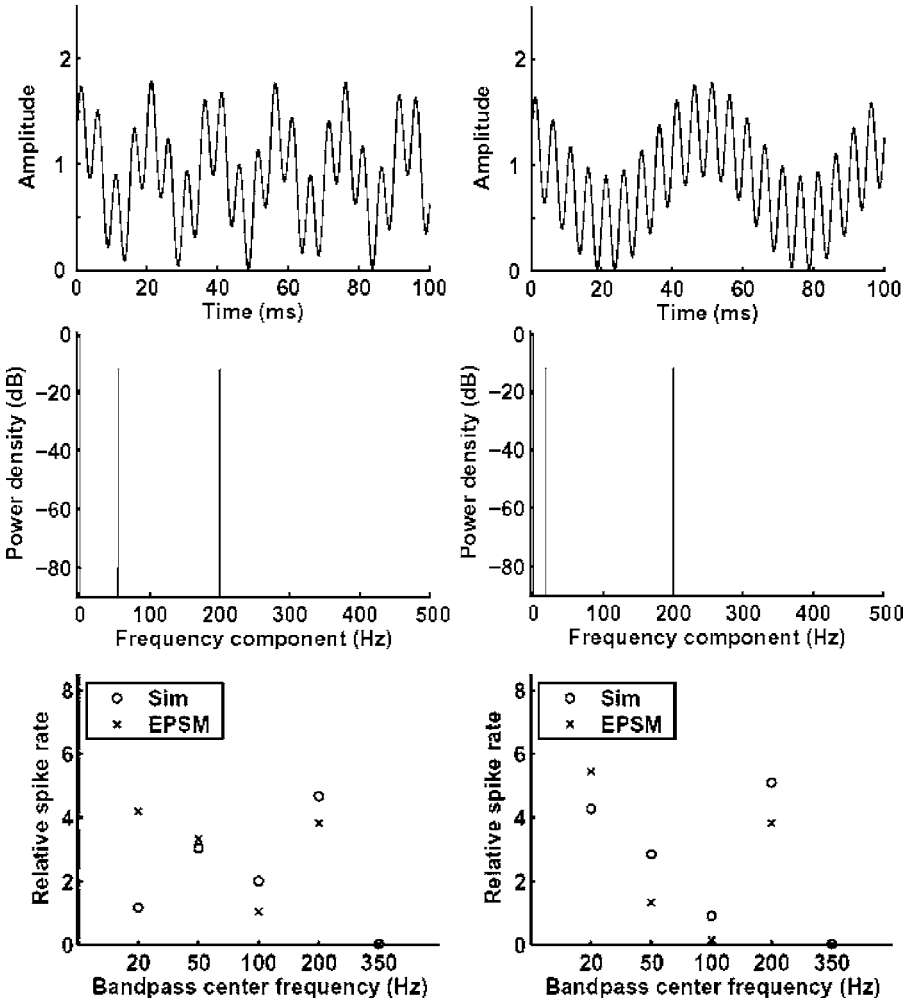


FIG. 6. The upper row shows the temporal waveform of the envelope of a 10 kHz tone modulated by the sum of two sinusoids with modulation frequencies $f_{m1}=55$ Hz and $f_{m2}=200$ Hz (left) and $f_{m1}=18$ Hz and $f_{m2}=200$ Hz (right). The middle row shows the corresponding envelope spectra of the two-component AM stimuli normalized such that the dc peak corresponds to 0 dB. In the bottom row the simulated relative spike rates of the five model filters (circles) and EPSM output (crosses) in response to the stimuli at a presentation level of 60 dB SPL.

sponding low-noise noise (right). The middle row shows the respective envelope spectra, normalized to a dc peak of 0 dB. Except for the dc peak, the Gaussian noise (left) exhibits a triangular continuous envelope spectrum, if plotted on a linear scale, which decreases rapidly for spectral components exceeding the bandwidth of the noise. Despite the identical bandwidths, the envelope spectrum of low-noise noise (right) exhibits a completely different shape. The envelope spectrum of low-noise noise contains much less energy than the envelope spectrum of Gaussian noise (see also Dau *et al.*, 1999).

The bottom row of Fig. 7 shows the simulated neural activity of the five model filters (circles) in response to the Gaussian noise (left) and low-noise noise (right) together with the EPSM results (crosses). In both cases, the simulated activity of the neural model is similar to the output obtained with the EPSM. While the activity at the output of filters with BMFs below the stimulus bandwidth reflects the triangular shape of the envelope spectrum of the Gaussian noise, filters with BMFs that exceed the bandwidth of the stimulus reveal essentially no activity (lower left panel). As expected from comparing the overall amount of energy in the envelope spectra of narrowband Gaussian noise and low-noise noise, in case of the low-noise noise (right bottom panel), the activity of filters with BMFs below the bandwidth of the low-noise noise is generally smaller than those obtained in response to Gaussian noise of the same bandwidth (compare to lower left panel). Filters with BMFs that exceed the bandwidth of the low-noise noise reveal no response, as is also expected from the envelope spectrum of the stimulus.

3. Processing of iterated rippled noise

The iterated rippled noise (IRN) stimulus was generated from a broadband Gaussian noise. The noise was iteratively attenuated and delayed before being added to the original version of the noise. During the iteration process that results in a quasi-periodicity of the stimulus, the output of one iteration stage served as input to the next iteration stage. The process used here followed the "add same" configuration described in Yost (1996). The IRN stimulus was obtained from $N=16$ iterations using a delay $\Delta=20$ ms and a gain $g=1$. While the narrowband stimuli used in the previous sections were not affected by the peripheral filter transfer function at 10 kHz, the broadband IRN stimulus had to be filtered before computing the EPSM output. The temporal envelope of the filtered IRN stimulus is shown in the upper panel of Fig. 8 and the corresponding envelope spectrum is shown in the middle panel. The amplitudes of the spectral envelope components are given in dB with the dc peak corresponding to 0 dB. The magnitude spectrum (not shown) as well as the

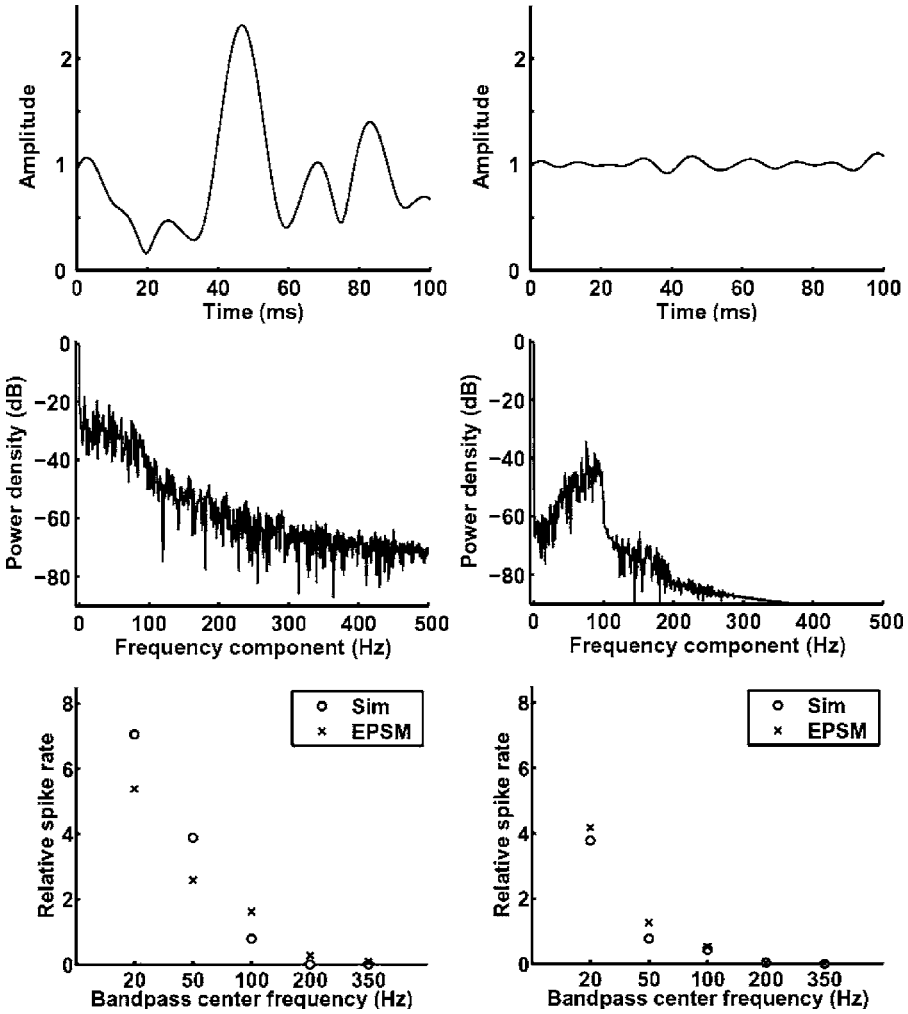


FIG. 7. The upper row shows the temporal envelope of a narrowband Gaussian noise (left) and a low-noise noise (right) centered at 10 kHz with a bandwidth of 100 Hz. The middle row shows the corresponding envelope spectra normalized to a dc peak of 0 dB. In the bottom row, simulated relative spike rates of the five model filters (circles) in response to the respective stimuli presented at a level of 60 dB SPL are shown. The corresponding EPSM output is represented by the crosses.

envelope spectrum of the IRN reveal a rippled structure with peaks at integer multiples of 50 Hz, the reciprocal of the IRN delay $\Delta = 20$ ms.

The bottom panel of Fig. 8 displays the simulated relative spike rates of the five model filters (circles) in response to the IRN stimulus presented at 60 dB SPL. Additionally, the EPSM results are represented by crosses. Although they are in qualitative agreement, the model output activity and the EPSM results differ from one another. For example, the EPSM output for the 20 Hz filter is clearly larger than the simulated spike rate obtained for this filter. The opposite is observed for the 350 Hz filter, where the simulated spike rate exceeds the EPSM output. Possible reasons for this difference between the two models will be addressed below.

IV. DISCUSSION

A. Biological plausibility of the present model

The pure-tone PSTH and the rMTF of each model unit have been compared to data from electrophysiological recordings, such that each unit could be associated with a physiological response type. The simulated responses of the model EN unit, for example, correspond to responses found in PVCN octopus cells. These cells were found to project to the superior paraolivary nucleus (SPN) and the ventral nucleus of the lateral lemniscus (VNLL). The influence of octopus cells on the IC is indirect and inhibitory as neurons in the SPN are GABAergic (Kulesza and Berrebi, 1999), and neurons in the VNLL that are innervated by octopus cells are glycinergic (Saint Marie *et al.*, 1997). An explicit simulation of VNLL cells was not included in the model. It was assumed here that VNLL units show similar response properties as the EN units of the present model. Generally, a large diversity of temporal response types has been observed in the VNLL including a reasonable percentage of onset responses (e.g., Batra and Fitzpatrick, 1999). Recently, Zhang and Kelly (2006) have shown in the rat that VNLL units with onset responses and high precision in timing exist. It has been suggested that the neurons in the VNLL are particularly well suited for preserving and processing temporal information in the auditory system (Oertel and Wickesberg, 2002).

The inhibitory influence of octopus cells on IC neurons therefore seems to agree with the neural connectivity assumed in the present model. Moreover, the ability of octopus cells to entrain to periodic stimuli (Rhode and Smith, 1986; Rhode, 1994; Oertel *et al.*, 2000) has led to the assumption that these cells might play an important role in the processing of amplitude modulation and pitch information (Frisina *et al.*, 1990; Golding *et al.*, 1995; Evans and Zhao, 1997; Cai *et al.*, 2001). The present study specifically assumes that the unit that generates O_I responses in the neural circuit is very

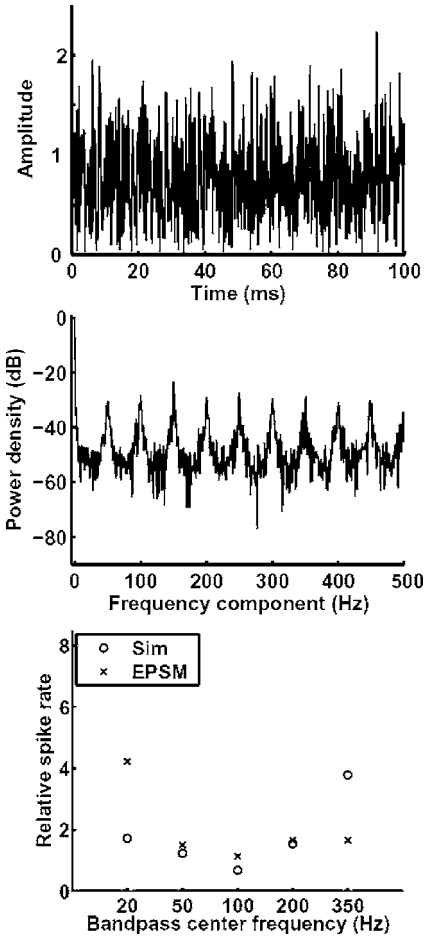


FIG. 8. The upper panel shows the temporal envelope of the IRN stimulus with $\Delta=20$ ms, $g=1$ and $N=16$. The middle panel shows the corresponding envelope spectrum normalized to a dc peak of 0 dB. In the bottom panel, the simulated relative spike rate of the five model filters (circles) in response to the IRN stimulus presented at a level of 60 dB SPL is shown. The corresponding EPSM output is represented by the crosses.

critical for the formation of bandpass shaped rMTFs, and for the processing of complex periodic waveforms.

The CR units of the model circuit can be associated with VCN stellate cells. These units have been found to project to the IC via synaptic terminals that are indicative of an excitatory neural connection (Smith and Rhode, 1989). The excitatory projections of stellate cells onto IC neurons agree with the assumed model connectivity.

Compared to the CN, little is known about cell types and neural connectivity in the IC. However, Krishna and Semple (2000) reported that IC neurons having rMTFs with a region of suppression exhibit a sustained pure-tone PSTH. In contrast, IC neurons that reveal a bandpass shaped rMTF show an onset-type pure-tone PSTH. These findings can be simulated qualitatively by the IC model units. Although it is not known what causes the region of suppression observed in recorded IC rMTFs, the connectivity of the present model might provide a possible mechanism by which the regions of suppression could be shaped.

Little is known about lateral connections between neurons in the IC. It is unclear whether the ordered lateral inhibition among neighboring cells, as assumed in the circuit, is biologically plausible or not. However, the fact that the auditory midbrain appears to be extensively interconnected (Oliver and Huerta, 1991) does at least not seem to argue against the choice of lateral connections in the present model.

The EN units were derived in such a way that they systematically differ in their two threshold potentials, namely the spike and release thresholds. The specific choice of the EN threshold potentials (see Table III) results in a linear relation between the simulated spike rate of a modulation filter and the modulation depth of a SAM stimulus (see upper right panel of Fig. 5). The required differences in the sensitivity of different EN model units to fluctuations in the stimulus envelope might also be obtained by a number of alternative implementations. This could be done, for example, by exclusively varying the spike thresholds of the different EN units, or by exclusively varying their release thresholds or their membrane time constants. Thus, although the way in which the different EN units of the present model are implemented is somewhat arbitrary, the general mechanism assumed here to account for the processing of different modulation depths in SAM stimuli and complex envelope waveforms is not affected by the implementation details of the EN units.

B. Rate normalization of BP units and level dependency

The current model topology directly results in a maximum discharge rate of the BP rMTFs proportional to the BMF of the unit (left panel of Fig. 4). This is not observed in the physiology as seen in the right panel of Fig. 4 which shows data replotted from Langner and Schreiner (1988). A constant maximum discharge rate would be achieved if an increasing number of parallel circuits inversely proportional to the BMF of the respective units was assumed. This solution is comparable to the decreasing number of chopper units as input to the coincidence cells with increasing BMF in the model of Hewitt and Meddis (1994). A higher number of parallel circuits at low BMFs would also increase the fidelity of the coding of these modulation frequencies, an aspect that seems particularly important with respect to the role of low amplitude modulations in speech perception.

Another property of the current model is that the rMTFs of the BP units are essentially level independent between 40 and 80 dB (lower right panel of Fig. 5). The main reason for this behavior is the "cycle count" property of the model EN units that was designed to be level independent over a large range. While psychophysics suggests level independency of AM coding over large level ranges, this appears to be inconsistent with results from physiological studies in detail. However, a diversity of level dependent response functions have been found in the IC. For example, Krishna and Semple (2000) showed responses with either enhancement or with suppression which all showed different amounts of level dependency. The model of Hewitt and Meddis (1994) shows a decrease of the rMTF at BMF with level at levels 20–30 dB above threshold which is caused by the decreasing synchronization of the chopper response to the positive half cycle of

the modulation. Also the Nelson and Carney (2004) model shows some level dependence, with increasing rate up to 30 dB and decreasing rates at 50 and 70 dB.

C. Possible alternative mechanisms

Some of the features of the present model framework might also be realized by alternative units or mechanisms. Especially, the mechanism by which different BMFs are shaped could be realized in several ways within the model framework, as long as two main features are taken into account: First, the spike rates of the CR model units need to be independent of the stimulus modulation frequency, i.e., the CR model units need to produce flat rMTFs. The CR units of the present model exhibit chopper pure-tone PSTHs and thus can be associated with VCN stellate cells. However, as 90% of the CN neurons are found to display flat or low-pass rMTFs in response to AM signals (Backoff *et al.*, 1999), the constant-rate responses of the CR units could be represented by a number of different CN cell types. Second, the mechanism by which the different BMFs of the model BP units are shaped, namely the variation in the number of CR inputs, can be exchanged by other mechanisms that result in varying input rates for different LP units. Such different input rates can, for example, be gained by varying the spike rates among the CR units. Differences in the effectiveness of synapses connecting CR units and LP units are another possible mechanism by which different BMFs could be yielded. Finally, a combination of these three possible mechanisms, i.e., a different number of input units, a variation of spike rates, and different synaptic effectiveness could also produce the formation of different BMFs. It cannot be decided here which CN neuron type might represent the model CR units and which of the suggested mechanisms actually might be realized. However, the suggested mechanism describing the formation of bandpass shaped rMTFs with different BMFs provides a basic mechanism that can be realized in several ways.

D. Nonlinearities in the present model

In order to evaluate the present neural model, simulation results obtained in response to complex stimulus conditions were compared to results of a linear modulation filter bank analysis. In contrast to linear processing of AM information, the simulated spike rates of the biologically motivated model were obtained from a highly nonlinear mechanism, inherently resulting from the nonlinear responses of auditory neurons. The linear modulation filterbank analysis, based on the concept of the envelope-power spectrum model (EPSM), demonstrated that the neural mechanism suggested here actually performs a spectral decomposition of the complex envelope of the stimuli. To our knowledge, no other neural periodicity model has so far been tested with complex envelope waveforms.

However, in some cases, the simulated relative spike rate was found to clearly differ from the EPSM output. Possible explanations for these differences are described in the following. The stochastic fluctuations in the AN activity that serves as input to the units of the model CN stage lead to small deviations of the simulated spike rates and the deterministic EPSM predictions. In addition, small deviations might arise from the way the EN unit threshold potentials are varied: (i) While an EN unit with a low spike threshold exhibits an action potential after only a short integration time, the generation of action potentials in EN units with high spike thresholds requires a longer integration time, and (ii) model EN units with high spike thresholds cannot detect the slowly rising slopes of stimuli modulated at a low frequency. Thus, BP units that receive input from EN units with high spike thresholds do not exhibit the 20 Hz bandpass filter in their rMTFs. Both the increased delay in the EN unit activity and the reduced number of model subcircuits that take part in encoding low modulation frequencies might result in differences between the simulated spike rates and the results from the EPSM in filters with low BMFs. However, this problem could be avoided if the EN units were chosen to have the same low spike threshold and differed only in their release threshold.

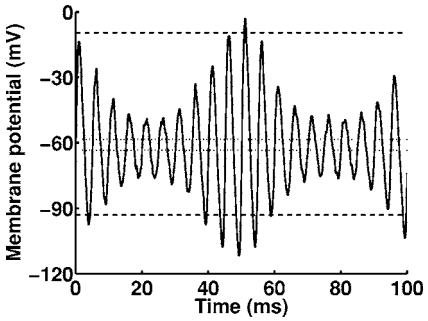


FIG. 9. Membrane potential of the EN model units in response to a 100% amplitude modulated 10 kHz tone modulated at 18 and 200 Hz. The dotted lines indicate the spike threshold ($\Theta_{act}=-58.5$ mV) and the release threshold ($Q_{rel}=-63.5$ mV) of EN unit 1. The dashed lines indicate the spike threshold ($\Theta_{act}=-9.5$ mV) and the release threshold ($\Theta_{rel}=-93$ mV) of EN unit 10.

While the abovementioned EN unit only partly contributes to the deviations between simulated spike rates and the EPSM output, the main reason for the deviations arises from the model mechanism itself, by which temporal periodicity information is mapped onto different BP model units. The temporal periodicity information that is most crucial for the present model is provided by the temporal activity patterns of the EN model units. As has been described before, the membrane potentials of the EN model units integrate over a high-pass filtered version of the temporal stimulus waveform until the unit reaches its spike threshold and exhibits a spike. After a spike, the unit enters its spike-blocking mechanism and is prevented from emitting further spikes until it is released by its membrane potential falling below its release potential. Thus, while the EN model units differ in their spike and their release thresholds, their membrane potentials are identical. Figure 9 displays the membrane potential common to all EN model units in response to the two-component AM stimulus with modulation frequencies of $f_{m1}=18$ Hz and $f_{m2}=200$ Hz. The dotted horizontal lines indicate the spike threshold ($\Theta_{act}=-58.5$ mV) and the release threshold ($\Theta_{rel}=-63.5$ mV) of the EN unit with the smallest threshold potential difference, i.e., EN unit 1 in Table III. The dashed horizontal lines indicate the spike threshold

(Θ_{act}=−9.5 mV) and the release threshold (Θ_{rel}=−93 mV) of the EN unit with the largest threshold potential difference, i.e., EN unit 10 in Table III. While EN unit 1 (dotted lines) entrains to the fast modulation frequency of 200 Hz, EN unit 10 (dashed lines) entrains to the slow modulation frequency of 18 Hz. However, other EN units whose threshold potentials lie between the threshold potentials of EN unit 1 and EN unit 10 might respond with ISIs that do not correspond to one of the stimulus modulation frequencies. These units extract "erratic" periodicities that are not contained in the stimulus envelope waveform itself. The effects of this error are most prominent when two modulation frequencies are applied (e.g, 55 and 200 Hz; lower right panel Fig. 6) where erratic periodicities corresponding to frequencies between the two imposed modulation frequencies occur. In the case of IRN, which has a broadband envelope spectrum with a large extent of short periods (high frequencies), errors also occur increasingly at high frequencies (see Fig. 8).

It is difficult to decide whether these erratic periodicities represent a structural problem of the present model or whether they simply result from the way the EN model units are implemented. However, the recorded synchronization coefficient of an O_1 unit in response to 200% amplitude modulated pure tones with varying modulation frequencies, as shown in the left panel of Fig. 7 in Dicke and Dau (2005), indicates that the latter might be the case. The recorded synchronization coefficient of the O_1 unit is close to 1 for all stimulus modulation frequencies. Although the simulated synchronization coefficient of the EN model unit, displayed in the right panel of the same figure, agrees with the recorded synchronization coefficient for high modulation frequencies, it deviates from the recorded data at low modulation frequencies (e.g., 50 Hz) where the EN model unit also detects the small lobes of the 200% AM stimulus. Thus, the question arises, how O_1 units can selectively follow low modulation frequencies without detecting the small lobes of a 200% AM stimulus while still being sensitive to large stimulus modulation frequencies. If the different EN model units would be assumed to employ more complex, for example, adaptive mechanisms, they might also reveal such a selectivity to the modulation frequency of a 200% AM stimulus within the range of their overall sensitivity (which is determined by their threshold potential difference). If the EN model units would be modeled such that they reproduced these findings, erratic periodicities that are not contained in the stimulus waveform itself could possibly be reduced or even avoided.

E. Implications of the present model

The different BMFs obtained in the BP units of the present model result from varying the input spike rate of the different LP model units. Thus, the formation of different BMFs is yielded by a rate-based mechanism and not by varying a temporal model parameter. This is in contrast to other existing models describing the formation of bandpass shaped rMTFs in the IC based on varying temporal parameters among the model units in order to yield a representation of different BMFs at different IC neurons (Langner, 1981; Langner and Schreiner, 1988; Langner, 1997a, b; Hewitt and Meddis, 1994; Nelson and Carney, 2004). In order to test the biological plausibility of these models, electrophysiological recordings have to reveal whether the range of temporal parameters needed to cover the wide range of modulation frequencies in these models really exists in biological neurons. The present neural circuit describes a fundamentally different mechanism, based on the connectivity of biologically plausible units.

In order to further evaluate the present neural model, the simulated spike rates obtained in response to complex envelope waveforms should be compared to real responses recorded from IC neurons. Since such data are currently not available, it is not clear how IC neurons with bandpass rMTFs respond to complex stimuli such as those used in the present study. Thus, the stimulus driven responses or the nonlinear effects (erratic periodicities) observed in the BP units of the present model could possibly also be observed in recorded responses from auditory neurons. Since the present model depends crucially on the temporal information it receives from the EN model units, it would be interesting to know whether the differences in modulation depth sensitivity suggested in the EN model units could be verified in electrophysiological recordings from onset units in the CN. Moreover, it would be valuable to investigate whether different CN onset units respond differently to complex stimuli such as those used in the present study.

V. SUMMARY

A neural circuit was presented that transforms temporal AM information, as a code of stimulus periodicity at the level of the CN, into a rate-based representation of stimulus periodicity at the level of the IC. Such a rate-based code of stimulus periodicity is assumed to be represented by bandpass shaped rMTFs with different rate BMFs. At the model CN stage, two types of units were simulated: (i) a unit that entrains to the stimulus modulation frequency and thus provides the temporal periodicity information, and (ii) a number of units whose response rates are independent of the stimulus modulation depth and rate. All further response properties, including the bandpass shaped rMTFs at the model IC stage, result from these two CN model units and the model's connectivity. Most of the response properties of each model unit as well as the synaptic connections used in the present model were found to agree with experimental findings. Moreover, the region of suppression observed in recorded IC rMTFs can be explained by the present model, at least qualitatively. Alternative models, describing the formation of bandpass shaped rMTFs between CN and IC, employ systematically varying temporal model parameters in order to cover a large range of BMFs. In contrast, the present model simply varies the number of input units projecting onto different IC model units in order to obtain different BMFs. The present model may provide a biologically plausible and simple mechanism that allows the simulation of the recorded bandpass rMTFs observed in IC recordings. Additionally, the present model allows for encoding a large range of modulation depths and accounts for a spectral decomposition of complex envelope waveforms, e.g., pure tones modulated by the sum of two

sinusoids, narrowband Gaussian and low-noise noise, and iterated rippled noise (IRN). Simulation results in response to these complex envelope waveforms were compared to results according to a linear modulation filter bank analysis. The comparison revealed that the simulated spike rates obtained in different modulation filters of the present nonlinear neural model are broadly consistent with a linear limited-resolution frequency decomposition of the stimulus envelope. The nonlinear neural circuit suggested in the present study might represent a possible neural correlate of a modulation filter bank. The circuit might therefore provide a reasonable network for the processing and encoding of periodicity information in complex stimuli.

ACKNOWLEDGMENTS

We would like to thank our colleagues in the research groups "Medizinische Physik" at Oldenburg University and the "Centre for Applied Hearing Research" at the Technical University of Denmark for their input. We also thank Paul Nelson and two anonymous reviewers for very helpful and constructive remarks and suggestions. This work was supported by the Deutsche Forschungsgemeinschaft (Research Project Nos. GRK 591 and SPP 1046) and the Danish Research Council.

[1]The transfer functions of the modulation filters were obtained from a linear interpolation of the simulated filter data (see left panel of Fig. 4). At low modulation frequencies (<10 Hz), this interpolation was computed from the simulated relative spike rate in response to a 60 dB SPL 10 kHz tone. The EPSM results were obtained from extracting the envelope of the stimulus, computing the envelope power spectrum, and calculating the power of the spectral envelope components in the transfer range of each interpolated modulation filter. The envelope power spectrum, N, was computed from the Fourier transform of the Hilbert envelope of the temporal stimulus waveform. The envelope power spectra of all stimuli were scaled such that their dc peaks were identical. The envelope power, P_{env} (BMF), at the output of a specific modulation filter centered at BMF was calculated by multiplying the envelope power spectrum of the stimulus, N, with the squared transfer function of the filter $W_{\mathrm{BMF}}(f_{\mathrm{env}})$ and integrating across envelope frequencies f_{env}:

$$P_{\mathrm{env}}(\mathrm{BMF}) = \int_{f_{\mathrm{env}}=0}^{\infty} N(f_{\mathrm{env}}) W_{\mathrm{BMF}}(f_{\mathrm{env}}) df_{\mathrm{env}}.$$

The squared transfer functions W_{BMF} all model filters were assumed to be zero for $f_{\mathrm{env}} > 1000$ Hz. As has been shown in Fig. 5, the simulated relative spike rate of a filter in the neural model is essentially a linear function of the modulation depth of an AM pure tone. In order to account for this linear behavior, the square root of the integrated power P_{env}(BMF) at the output of each filter, which corresponds to the rms modulation depth for SAM tones, was compared to the simulated relative spike rates. All EPSM results were scaled by the same constant factor in order to make the two measures directly comparable. This constant was obtained by dividing the sum of simulated relative spike rates of all model filters in all simulations by the corresponding sum of all EPSM results. With the described processing, the EPSM results would perfectly match the predictions of neural circuit in the case of 100% SAM tone stimuli.

Arle, J. E., and Kim, D. O. (**1991**). "Neural modeling of intrinsic and spike-discharge properties of cochlear nucleus neurons," Biol. Cybern. **64**, 273–283.

Backoff, P. M., Palombi, P. S., and Caspary, D. M. (**1999**). "γ-Aminobutyric acidergic and glycinergic inputs shape coding of amplitude modulation in the chinchilla cochlear nucleus," Hear. Res. **134**, 77–88.

Bacon, S. P., and Grantham, D. W. (**1989**). "Modulation masking: Effects of modulation frequency, depth, and phase," J. Acoust. Soc. Am. **85**, 2575–2580.

Batra, R., and Fitzpatrick, D. C. (**1999**). "Discharge patterns of neurons in the ventral nucleus of the lateral lemniscus of the unanesthetized rabbit," J. Neurophysiol. **82**, 1097–1113.

Bregman, A. S. (**1990**). *Auditory Scene Analysis: The Perceptual Organization of Sound* (MIT Press, Cambridge).

Cai, Y., McGee, J., and Walsh, E. J. (**2001**). *Computational Models of Auditory Function*, edited by S. Greenberg and M. Slaney (OS, Amsterdam).

Dau, T., Kollmeier, B., and Kohlrausch, A. (**1997a**). "Modeling auditory processing of amplitude modulation. I. Modulation detection and masking with narrowband carriers," J. Acoust. Soc. Am. **102**, 2892–2905.

Dau, T., Kollmeier, B., and Kohlrausch, A. (**1997b**). "Modeling auditory processing of amplitude modulation. II. Spectral and temporal integration in modulation detection," J. Acoust. Soc. Am. **102**, 2906–2919.

Dau, T., Verhey, J. L., and Kohlrausch, A. (**1999**). "Intrinsic envelope fluctuations and modulation-detection thresholds for narrowband noise carriers," J. Acoust. Soc. Am. **106**, 2752–2760.

De Boer, E. (**1976**). *Handbook of Sensory Physiology*, edited by W. D. Keidel and W. D. Neff (Springer-Verlag, Berlin), Vol. **3**.

Dicke, U., and Dau, T. (**2005**). "A functional point-neuron model simulating cochlear nucleus ideal onset responses," J. Comput. Neurosci. **19**, 239–253.

Evans, E. F., and Zhao, W. (**1997**). "Onset units in guinea pig ventral cochlear nucleus: Neuropharmacological studies," Assoc. Res. Otolaryngol. Abstr. **20**, 116.

Evans, E. F. (**1978**). "Place and time coding of frequency in the peripheral auditory system: Some physiologial pros and cons," Audiology **17**, 369–420.

Ewert, S. D., and Dau, T. (**2000**). "Characterizing frequency selectivity for envelope fluctuations," J. Acoust. Soc. Am. **108**, 1181–1196.

Fletcher, H. (**1940**). "Auditory patterns," Rev. Mod. Phys. **12**, 47–65.

Frisina, R. D., Smith, R. L., and Chamberlain, S. C. (**1990**). "Encoding of amplitude modulation in the gerbil cochlear nucleus: I. A hierarchy of enhancement," Hear. Res. **44**, 99–122.

Godfrey, D. A., Kiang, N. Y. S., and Norris, B. E. (**1975**). "Single unit activity in the posteroventral cochlear nucleus of the cat," J. Comp. Neurol. **162**, 247–268.

Goldberg, J. M., and Brown, J. B. (**1969**). "Responses of binaural neurons of dog superior olivary complex to dichotic stimuli: Some physiological mechanisms of sound localization," J. Neurophysiol. **32**, 613–636.

Golding, N. L., Robertson, D., and Oertel, D. (**1995**). "Recordings from slices indicate that octopus cells of the cochlear nucleus detect coincident firing of auditory nerve fibers with temporal precision," J. Neurosci. **15**, 3138–3153.

Hewitt, M. J., and Meddis, R. (**1994**). "A computer model of amplitude-modulation sensitivity of single units in the inferior colliculus," J. Acoust. Soc. Am. **95**, 2145–2159.

Hewitt, M. J., Meddis, R., and Shackleton, T. M. (**1992**). "A computer model of a cochlear-nucleus stellate cell: Responses to amplitude-modulated and pure-tone stimuli," J. Acoust. Soc. Am. **91**, 2096–2109.

Houtgast, T. (**1989**). "Frequency selectivity in amplitude-modulation detection," J. Acoust. Soc. Am. **85**, 1676–1680.

Javel, E. (**1980**). "Coding of AM tones in the chinchilla auditory nerve: Implications for the pitch of complex tones," J. Acoust. Soc. Am. **68**, 133–146.

Joris, P. X., and Yin, T. C. T. (**1992**). "Responses to amplitude-modulated tones in the auditory nerve of the cat," J. Acoust. Soc. Am. **91**, 215–232.

Joris, P. X., Schreiner, C. E., and Rees, A. (**2004**). "Neural processing of amplitude-modulated sounds," Physiol. Rev. **84**, 541–577.

Kim, D. O., Sirianni, J. G., and Chang, S. O. (**1990**). "Responses of DCN-PVCN neurons and auditory nerve fibers in unanesthetized decerebrate cats to AM and pure tones: Analysis with autocorrelation/power-spectrum," Hear. Res. **45**, 95–113.

Kohlrausch, A., Fassel, R., van der Heijden. M., Kortekaas, S. R., van der Par, S., Oxenham, A., and Püschel, D. (**1997**). "Detection of tones in low-noise noise: Further evidence for the role of envelope fluctuations," Acust. Acta Acust. 83, 659–669.

Krishna, B. S., and Semple, M. N. (**2000**). "Auditory temporal processing: Responses to sinusoidally amplitude-modulated tones in the inferior colliculus," J. Neurophysiol. **84**, 255–273.

Kulesza, R. J., and Berrebi, A. S. (**1999**). "Distribution of GAD isoforms in the superior paraolivary nucleus (SPON) of the rat," Assoc. Res. Otolaryngol Abstr. **22**, 70–71.

Langner, G., and Schreiner, C. E. (**1988**). "Periodicity coding in the inferior colliculus of the cat. I. Neuronal mechanisms," J. Neurophysiol. **60**, 1799–1822.
Langner, G. (**1981**). "Neuronal mechanisms for pitch analysis in the time domain," Exp. Brain Res. **44**, 450–454.
Langner, G. (**1997a**). "Temporal processing of pitch in the auditory system," J. New Music Res. **26**, 116–132.
Langner, G. (**1997b**). "Neural processing and representation of periodicity pitch," Acta Oto-Laryngol. **532**, 68–76.
Lorenzi, C., Soares, C., and Berthommier, F. (**1995**). "Neuronal correlates of perceptual amplitude-modulation detection," Hear. Res. **90**, 219–227.
MacGregor, R. J. (**1987**). *Neural and Brain Modeling* (Academic, San Diego).
Meddis, R., Hewitt, M. J., and Shackleton, T. M. (**l990**). "Implementation details of a computation model of the inner hair-cell/auditory-nerve synapse," J. Acoust. Soc. Am. **87**, 1813–1816.
Meddis, R. (**1986**). "Simulation of mechanical to neural transduction in the auditory receptor," J. Acoust. Soc. Am. **79**, 702–711.
Meddis, R. (**1988**). "Simulation of auditory-neural transduction: Further studies," J. Acoust. Soc. Am. **83**, 1056–1063.
Nelson, P. C., and Carney, L. H. (**2004**). "A phenomenological model of peripheral and central neural responses to amplitude-modulated tones," J. Acoust. Soc. Am. **116**, 2173–2186.
Oertel, D., and Wickesberg, R. E. (**2002**). *Integrative Function in the Mammalian Auditor Pathway*, edited by D. Oertel, R. R. Fay, and A. N. Popper (Springer-Verlag, New York).
Oertel, D., Bal, R., Gardner, S. M., Smith, P. H., and Joris, P. X. (**2000**). "Detection of synchrony in the activity of auditory nerve fibers by octopus cells of the mammalian cochlear nucleus," *PNAS. National Academy of Sciences Colloquium on Auditory Neuroscience: Development, Transduction and Integration*.
Oliver, D. L., and Huerta, M. F. (**1991**). *The Mammalian Auditory Pathway: Neuroanatomy*, edited by D. B. Webster, A. N. Popper, and R. R. Fay (Springer-Verlag, New York).
Palmer, A. R. (**1982**). "Encoding of rapid amplitude fluctuations by cochlear-nerve fibers in the guinea pig," Arch. Oto-Rhino-Laryngol. **236**, 197–202.
Palmer, A. R. (**1995**). *Hearing*, edited by B. Moore (Academic Press, San Diego).
Patterson, R. D., Nimmo-Smith, I., Holdsworth, J., and Rice, P. (**1988**). "An efficient auditory filterbank based on the gammatone function," APU Report No. 2341, Applied Psychology Unit, Cambridge.
Rees, A., and Møller, A. R. (**1983**). "Responses of neurons in the inferior colliculus of the rat to AM and FM tones," Hear. Res. **10**, 301–330.
Rees, A., and Palmer, A. R. (**1989**). "Neuronal responses to amplitude-modulated pure-tone stimuli in the guinea pig inferior colliculus and their modification by broadband noise," J. Acoust. Soc. Am. **85**, 1978–1994.
Rhode, W. S., and Greenberg, S. (**1994**). "Encoding of amplitude modulation in the cochlear nucleus of the cat," J. Neurophysiol. **71**, 1797–1825.
Rhode, W. S., and Smith, P. H. (**1986**). "Encoding timing and intensity in the ventral cochlear nucleus of the cat," J. Neurophysiol. **56**, 261–286.
Rhode, W. S., Oertel, D., and Smith, P. H. (**1983**). "Physiological response properties of cells labeled intracellularly with horseradish perioxidase in the cat ventral cochlear nucleus," J. Comp. Neurol. **213**, 448–463.
Rhode, W. S. (**1994**). "Temporal coding of 200% amplitude modulated signals in the ventral cochlear nucleus of cat," Hear. Res. **77**, 43–68.
Romand, R. (**1978**). "Survey of intracellular recording in the cochlear nucleus of the cat," Brain Res. **148**, 43–65.
Rose, G. J., and Capranica, R. R. (**1985**). "Sensitivity to amplitude modulated sounds in the anuran auditory nervous system," J. Neurophysiol. **53**, 446–465.
Rose, J. E. (**1970**). *Frequency Analysis and Periodicity Detection*, edited by R. Plomp and G F. Smoorenburg (A. W. Sijthoff, Leidon).
Rosen, S. (**1992**). "Temporal information in speech: Acoustic, auditory and linguistic aspects," Philos. Trans. R. Soc. London, Ser. B **336**, 367–373.
Saint Marie, R. L., Shneiderman, A., and Stanforth, D. A. (**1997**). "Patterns of gamma-aminobutyric acid and glycine immunoreactivities reflect structural and functional differences of the cat lateral lemniscal nuclei," J. Comp. Neurol. **389**, 264–276.
Schouten, J. F., Ritsma, R. J., and Cardoza, B. L. (**1962**). "Pitch of the residue," J. Acoust. Soc. Am. **34**, 1418–1424.
Schreiner, C. E., and Langner, G. (**1988**). "Periodicity coding in the inferior colliculus of the cat. II. Topographical organization," J. Neurophysiol. **60**, 1823–1840.
Schreiner, C. E., and Langner, G. (**1997**). "Laminar fine structure of frequency organization in auditory midbrain," Nature (London) **388**, 383–386.
Schulze, H., and Langner, G. (**1997**). "Representation of periodicity pitch in the primary auditory cortex of the Mongolian gerbil," Acta Oto-Laryngol., Suppl. **532**, 89–95.
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.
Tuckwell, H. C. (**1988**). *Introduction to Theoretical Neurobiology* (Cambridge University Press, Cambridge) Vol. **1**.
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.
Yost, W. M. (**1996**). "The pitch of iterated rippled noise," J. Acoust. Soc. Am. **100**, 511–518.
Zhang, H., and Kelly, J. B. (**2006**). "Responses of neurons in the rats ventral nucleus of the lateral lemniscus to monaural and binaural tone bursts," J. Neurophysiol. **95**, 2501–2512.
Zwicker, E., Flottorp, G., and Stevens, S. S. (**1957**). "Critical bandwidth in loudness summation," J. Acoust. Soc. Am. **29**, 548–557.

Distortion product otoacoustic emission of symphony orchestra musicians before and after rehearsal

Karen Reuter[a] and Dorte Hammershøi
Department of Acoustics, Aalborg University, 9220 Aalborg, Denmark

(Received 23 May 2006; revised 17 October 2006; accepted 19 October 2006)

The $2f_1-f_2$ distortion product otoacoustic emission (DPOAE) and hearing levels are obtained for 12 normal-hearing symphony orchestra musicians both before and after their rehearsal. The DPOAE fine structures are determined and analyzed according to the character and prevalence of ripples. Hearing levels, DPOAE levels, and DPOAE fine structures before and after rehearsal are similar, indicating that no or marginal temporary change of the state of hearing were caused by the exposure. The data were further compared to similar data for occupationally nonexposed subjects, one group which was age and gender matched, and other two groups of younger individuals (one group with better hearing levels than the other). The data for the age and gender matched group compared well with the musicians data (and the data for the group of better-hearing younger individuals). In general, the analyses of hearing thresholds and DPOAE data thus lead to the same conclusions concerning the state of hearing. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2395915]

PACS number(s): 43.64.Jb, 43.50.Qp [BLM] Pages: 327–336

I. INTRODUCTION

Otoacoustic emissions (OAE) are sound signals produced by the cochlea as part of the normal-hearing process. They can be recorded in the ear canal and reflect the active process of the outer hair cells (OHCs), which can increase the sensitivity and frequency selectivity for weak sounds. The OHCs are the part of the auditory system that is considered the most vulnerable to loud sounds. Therefore OAEs are considered to be at least as sensitive to detect small noise-induced changes in the auditory system as the hearing thresholds. This could be confirmed by several studies (Attias *et al.*, 1995; Desai *et al.*, 1999 Hotz *et al.*, 1993; Kowalska and Sulkowski, 1997; LePage and Murray, 1998; Lucertini *et al.*, 2002; Murray and LePage, 1998; Sliwinska-Kowalska and Kotylo, 2002), suggesting that OAE might be used as an early warning sign of noise-induced hearing loss. A direct relation between OAE level and the hearing threshold is not fully established (Avan *et al.* (1991); Bonfils and Avan (1992); Gaskill and Brown (1993); Gorga *et al.* (1993)). This may, however, be partly due to the fact that the OAE and hearing threshold inherently reflect two different auditory processes. The hearing threshold also reflects the state of the inner hair cells and depends on the further neuronal processing, including cognitive detection.

Distortion product otoacoustic emission (DPOAE) is the response of the inner ear to two pure-tone stimuli (the primaries at frequencies f_1 and f_2). Because of nonlinear interaction of the two tones in the cochlea, the primaries evoke tones at various frequencies, the most prominent being at the frequency $2f_1-f_2$. DPOAEs are generated by an active system and only exist in the live, healthy cochlea. For the measurement of DPOAE over a particular frequency range the frequencies of the primaries are varied simultaneously while keeping their frequency ratio constant. It is widely accepted (Kalluri and Shera, 2001; Knight and Kemp, 2000; Konrad-Martin *et al.*, 2002; Mauermann *et al.*, 1999a; Talmadge *et al.*, 1998) that the $2f_1-f_2$ DPOAE is generated by two sources, (1) the distortion component generated at the region of primary overlap near f_2 and (2) the reflection component from the distortion product frequency at $2f_1-f_2$.

When the DPOAE is measured with sufficiently high frequency resolution, the DPOAE shows quasi-periodic variations across frequency, called DPOAE fine structure. The DPOAE fine structure is characterized by a periodic structure of maxima and minima with depth of notches of up to 20 dB (Gaskill and Brown, 1990; He and Schmiedt, 1993; Heitmann *et al.*, 1996) and a periodicity of 3/32 octaves (He and Schmiedt, 1993; Mauermann *et al.*, 1997). The DPOAE fine structure has shown to be flattened or reduced, when the f_2 component is isolated by either (1) presenting a suppressor tone close to the $2f_1-f_2$ frequency (Heitmann *et al.*, 1998; Kalluri and Shera, 2001; Talmadge *et al.*, 1999), (2) by spectral smoothing (Kalluri and Shera, 2001; Konrad-Martin *et al.*, 2002), or (3) by using a swept tone stimulus (Long *et al.*, 2004). The results from simulations (Mauermann *et al.*, 1999a) using a nonlinear and active model of the cochlea showed that "the component from the $2f_1-f_2$ site is sensitive to the existence of statistical fluctuations in the mechanical properties along the cochlea partition, i.e., roughness, while the f_2 component is not." From the model point of view, a decrease in DPOAE fine structure is expected, when the $2f_1-f_2$ component is affected by hearing loss, while the f_2 component falls into a healthy region.

Mauermann *et al.* (1999b) determined DPOAE fine structures in subjects with mild to moderate cochlear hearing losses with certain shapes of hearing loss. When the f_2 frequencies were located in the region of normal or near-normal hearing, but the $2f_1-f_2$ frequencies were located in a region of impairment, the distortion product $2f_1-f_2$ was still observable, but the DPOAE fine structure disappeared. When

[a]Author to whom correspondence should be addressed. Electronic mail: kr@acoustics.aau.dk

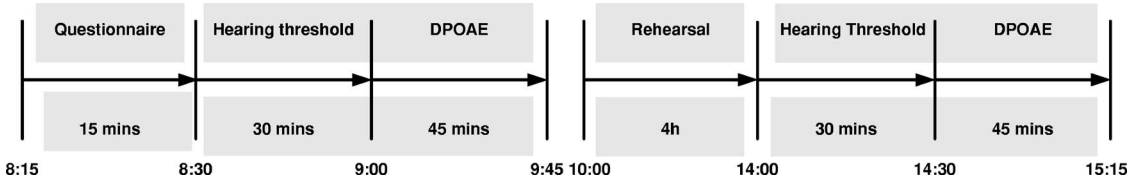


FIG. 1. Time schedule of the experiment. The order of measuring either hearing level or DPOAE is varied, the order is the same for one subject before and after rehearsal.

the $2f_1-f_2$ frequencies fell into a region of normal hearing, fine structure was preserved as long as DPOAEs could be recorded. The results of other studies support the idea that the presence of cochlear fine structure might be a property of the healthy ear. Engdahl and Kemp (1996) found a flattened DPOAE fine structure after an acoustic overexposure, and Rao *et al.* (1996) observed a reduction of DPOAE fine structure after aspirin consumption. Reuter and Hammershøi (2006) determined DPOAE fine structures of 50 normal-hearing humans. The subjects were divided into two groups A and B, subjects of Group A having slightly better hearing levels than subjects of Group B. Subjects of Group A had significantly higher DPOAE levels; the overall prevalence of fine structure ripples did not differ between subjects of the two groups, but ripples were higher and narrower for subjects of Group B than for Group A.

In the present study hearing levels and DPOAE fine structures are determined in a group of symphony orchestra musicians. Musicians are occupationally exposed to relatively high levels of sound and are often overexposed to levels exceeding the "safe" limit of 85 dBA (Early and Horstman, 1996; Royster *et al.*, 1991). These exposures may therefore affect the auditory system in a negative way. Some studies found that threshold levels of musicians are not significantly different than thresholds of nonexposed populations (Karlsson *et al.*, 1983; Kähäri *et al.*, 2001; Obeling and Poulsen, 1999), indicating that the music exposure does not cause severe hearing losses. However, Royster *et al.* (1991) found that 52.5% of individual musicians showed notched audiograms, indicating noise-induced hearing damage.

The aim of the present study is to investigate whether there is a systematic change in the DPOAE fine structure with the state of hearing and whether this change might be more sensitive to detect small changes in the state of hearing than the overall DPOAE level. Two different approaches are used:

1. Since the exposure during the rehearsal might cause temporary changes in the state of hearing, measurements are obtained both before and after the rehearsal to monitor whether the exposure during the rehearsal causes temporary changes in the auditory system.
2. The data of the musicians (Group M) are compared to similar data obtained from three other groups of subjects: a control group of age and gender matched, occupationally nonexposed subjects (Group C), a group of younger subjects with good hearing levels (Group A), and a group of younger subjects with (according to the authors' definition) slightly raised hearing levels (Group B). The data of subjects belonging to groups A and B have been previously reported in Reuter and Hammershøi (2006).

II. MATERIALS AND METHODS

For the musicians (Group M) hearing levels and DPOAE were obtained both before and after the rehearsal. The experiment took place in the facilities of the symphony orchestra in order to start the measurements after the rehearsal as early as possible. During the rehearsal the sound level was monitored. Each session lasted approximately 90 min, 30 min for the pure-tone audiometry and 60 min for the measurement of DPOAE fine structure. Both ears were measured, because there might be small differences between left and right ear for some subjects. Royster *et al.* (1991) showed that violinists and violists showed significantly poorer thresholds at 3–6 kHz in the left ear than in the right ear, consistent with the left ear's exposure from their instruments. The test procedure of this study is illustrated in Fig. 1. Only one subject was tested per rehearsal day. The order of obtaining either hearing level or DPOAE first and testing either right or left ear first was varied. All violin and viola players had their left ear tested first. For one subject the order was always the same in the session before and after rehearsal.

Similar data (DPOAE and hearing levels) were obtained for three other groups of subjects: a control group of age and gender matched, occupationally nonexposed subjects (Group C), a group of younger subjects with good hearing levels (Group A), and a group of younger subjects with (according to the authors' definition) slightly raised hearing levels (Group B). Group C also had both left and right ears tested. For Group A and B subjects one ear was tested, which was randomly chosen.

A. Subjects

The musicians (Group M) are members of the "Aalborg Symfoniorkester." Precondition for the participance in the experiments was that the musicians did not have any known hearing losses. The experiment was performed for 14 subjects. One subject (T2) reported a tinnitus after the rehearsal, one subject (T10) had a moderate hearing loss in one ear. These two subjects T2 and T10 were discarded from the study, because they did not fulfill the requirements. For the DPOAE measurement of subject T9 (left ear after rehearsal) there were problems with the probe fit. This subject is therefore excluded from the first part of the analysis (the comparison of data obtained before and after rehearsal and the comparison of left and right ears). For the comparison of

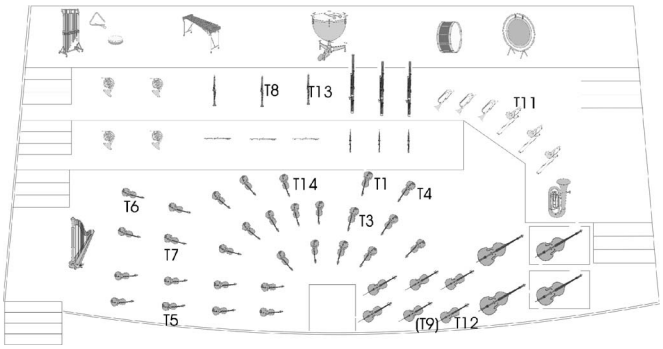


FIG. 2. Sketch of the symphony orchestra, showing the positions of the 12 subjects T1–T14 in the orchestra. Subjects T2 and T10 were excluded from the study and are not shown in the figure, T9 is excluded from parts of the study (see Sec. II A).

Group M data to the other groups of subjects this subject is included. The 12 musicians included in the group analysis are aged between 31 and 53 (mean=39.3) and are nine female and three male. They are from different instrument groups: fours violins, three violas, two celli, two clarinets, and one trombone. On average the subjects have played their instrument for 30 years, have been a member of "Aalborg Symfoniorkester" for 13 years and performed their profession for 16 years. The musicians practice on average 11.3 h per week on their own. The position of the subjects in the orchestra can be seen in Fig. 2. A relatively large group of musicians (T1, T3, T4, and T14) is situated in the area in front of the wood winds and brass section, which is the area in that musicians report high exposure levels.

The control group (Group C) consists of 12 subjects (age and gender matched population to the group of musicians, occupationally nonexposed subjects). Most of the subjects are staff from Aalborg University, who work in small offices. Furthermore, a comparison to the data of 50 university students (previously reported in Reuter and Hammershøi (2006)) is performed. The university students are aged between 20 and 29 (mean=23.6), 24 male, 26 female. All subjects of this group have near-normal hearing with hearing levels below 25 dB. They are further categorized into a group with good hearing levels (Group A) and a group with slightly raised hearing levels (Group B). The subjects of Group B have hearing levels exceeding 10 dB for at least two tested frequencies and have "raised" hearing levels according to the authors' definition. The subjects of all groups reported no known incidents of excessive exposures or known hearing losses.

B. Hearing level

For the hearing level measurement the Madsen Orbiter 922 was used together with Sennheiser HDA 200 headphones. Prior to the experiment the system was calibrated according to ISO 389-8 using the B&K 4153 artificial ear. Pure-tone hearing levels were measured in octaves between 125 Hz and 8 kHz. An ascending method was used, starting at 40 dB hearing level (HL) and decreasing with 10 dB steps. After the first miss the stepsize was changed to 4 dB up and 2 dB down to get a resolution of 2 dB. Group C underwent the same audiometry test as Group M, using the same equipment. For the 50 subjects belonging to groups A and B details of the hearing level measurements are given in Reuter and Hammershøi (2006).

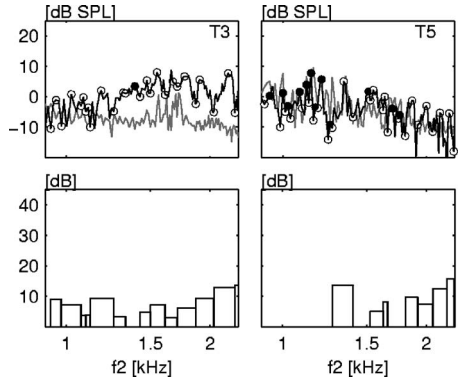


FIG. 3. Derivation of schematic DPOAE fine structure: examples for two musicians. Top panel: DPOAE fine structure (black line), noise floor (dotted gray line), maxima and minima (white circles) and maxima of rejected ripples (black circles). Lower panel: schematic representation of the fine structure ripples.

C. Distortion product otoacoustic emission

The $2f_1-f_2$ DPOAEs were measured using the ILO96 Research system from Otodynamics. DPOAEs were measured in the frequency range of 903 Hz$<f_2<$6201 Hz with $f_2/f_1=1.22$ and fixed primary levels of $L_1/L_2=65/45$ dB. A discussion of the choice of primary levels can be found in Reuter and Hammershøi (2006). The DPOAE fine structure was measured using the frequency resolution "micro." It presents approximately 17 primary tones within 200 Hz intervals for $f_2<3$ kHz and within intervals of 400 Hz for $f_2>3$ kHz. The f_2 frequency increases in 12 Hz steps and

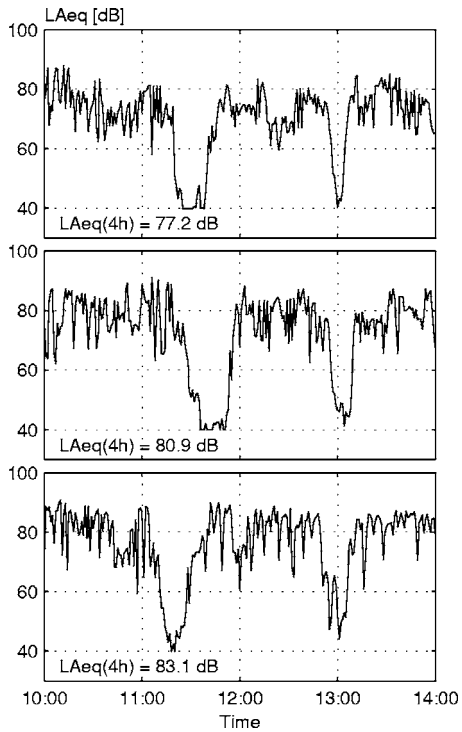


FIG. 4. Monitored sound pressure level during the rehearsal, examples for three rehearsal days. Top panel: rehearsal day with the lowest exposure. Bottom panel: rehearsal day with the highest exposure.

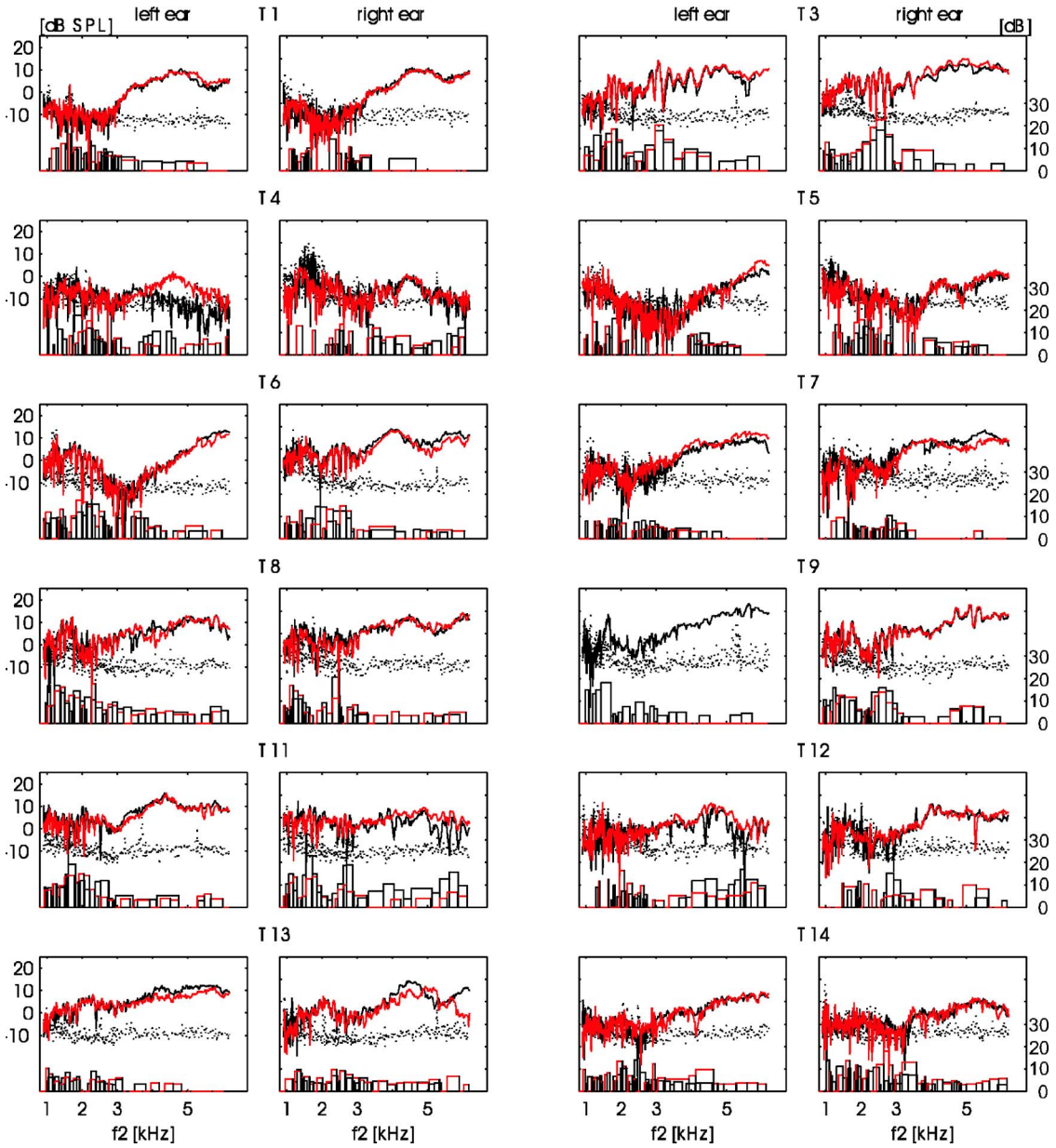


FIG. 5. (Color online) Individual DPOAE fine structure before rehearsal (black line), after rehearsal (red line), noise floor (dotted line) and schematic DPOAE (right ordinate). Measurements of both ears are shown for each subject, left ear (left) and right ear (right). The DPOAE level of subject T9, left ear after exposure, is excluded from the analysis and not shown in the figure.

24 Hz steps, respectively, i.e., a linear frequency spacing is used. Each pair of primary tones was averaged in the time domain (32 subaverages), corresponding to approximately 3 s measurement time for one pair of frequencies. The system uses a constant average time, independent of the signal to noise ratio. To cover the full frequency range, 19 measurements in different frequency ranges were measured and concatenated to one DP gram afterwards. The probe was not removed between these measurements, unless the measurement system indicated that the probe fit was altered. Including the breaks between the measurements that were used to save the data, the measurement of a DPOAE covering the entire frequency range lasted approximately 30 minutes.

Prior to each measurement a checkfit procedure was performed, where two broadband click stimuli are alternately delivered by the two output transducers. The checkfit result is stored in an array and used during data collection to balance and normalize the two stimuli levels. All spectrum analyses are performed by the system. A fast Fourier transform with a frequency resolution of 12.2 Hz is performed. The noise is estimated from the ten Fourier components nearest to but not including the $2f_1-f_2$ frequency. The noise is represented as all levels within two standard deviations of the background noise, i.e., the limits of the 95% confidence region. All measurements are saved as spreadsheet files and further analyzed. The 19 measurements of the different frequency regions are concatenated to a DP gram covering the measured frequency range.

D. Postprocessing of DPOAE fine structure

The DPOAE fine structure of each subject is analyzed by an automatic classification algorithm. The algorithm is

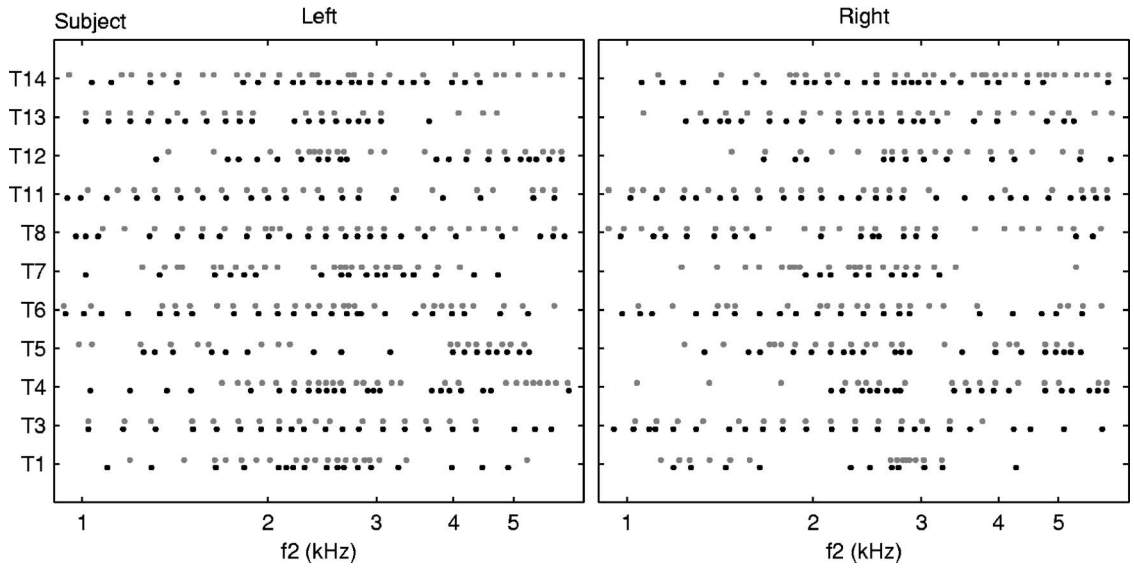


FIG. 6. Individual ripple center frequencies for all individual subjects before rehearsal (black) and after rehearsal (gray) for left ears (left) and right ears (right).

described in detail in Reuter and Hammershøi (2006). It detects ripple maxima and minima and characterizes each ripple by the parameters **ripple center frequency** (frequency centered between two minima on a logarithmic axis), **ripple spacing** (frequency distance between two minima), **ripple height** (level difference between the maximum and the mean of the two minima) and **ripple prevalence** (number of ripples counted in 1/3 octave frequency bands). Ripples are rejected when their height is less than 3 dB and when ripple maxima are less than 3 dB above the noise floor. Ripple height and ripple spacing are calculated in 1/8 octave frequency bands.

For a better illustration of the obtained fine structure characteristics two examples of a "schematic" DPOAE fine structure, derived from the parameters: ripple center frequency, spacing and height, are illustrated in Fig. 3. The top figures illustrate fine structure measurement examples for two subjects of Group M, zoomed into a narrow frequency range. The top figures also show the noise floor and the detected maxima and minima. The bottom figures illustrate the derived schematic fine structure for the same subjects. The subject shown on the left panel has a relatively high-level DPOAE and a pronounced fine structure with ripple heights of up to 15 dB, whereas the subject shown on the right panel has lower-level DPOAE. For the data shown on the right panel many of the fine structure ripple maxima are not 3 dB above the noise floor and are therefore rejected.

E. Monitoring of music sound level

A-weighted, equivalent sound pressure levels (L_{Aeq}) were measured using the B&K 2238 sound level meter. It was calibrated to a free field level of 93.9 dB every day using the B&K 4231 calibrator according to International Electrotechnical Commission standards. The sound level meter was placed in front of the conductor with a distance of 1.3 m at a height of 1.5 m. Figure 4 shows examples for three rehearsal days. Typically the rehearsal starts at 10 am and ends at 2 pm, with two breaks in between. During the breaks the musicians leave the rehearsal room and go to the canteen. The measurement system remained in the rehearsal room during the breaks, and the (low level) measurements are therefore not representative for the exposure during the breaks. The measured L_{Aeq} during the rehearsal varies between 70 and 90 dB. Different pieces of music were rehearsed during the four weeks of experiments. In the first week there was a rehearsal for an opera concert. One subject (Tl) participated in that time and was exposed to an $L_{\mathrm{Aeq,4h}}$ of 80.0 dB. Subjects T3-T11 rehearsed ballet music when they participated in the present experiment. The exposure of these days was 79.7 dB. Subjects T12 and T13 rehearsed a mid summer concert (82.0 dB), and subject T14 rehearsed a popular concert, which was the highest exposure with 83.1 dB. The most quiet day was one day of ballet music with 77.2 dB. All exposure levels are calculated as $L_{\mathrm{Aeq,4h}}$.

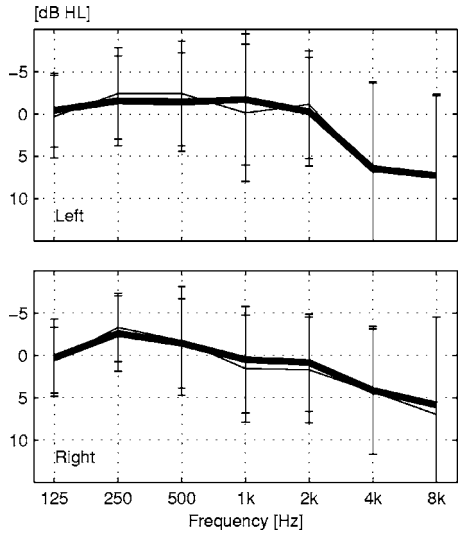


FIG. 7. Hearing level before (thin line) and after (thick line) the rehearsal for left ears (top panel) and right ears (lower panel). Error bars are standard deviations between the 11 subjects.

TABLE I. Balanced three-way-ANOVA p values, [a]$p<0.05$ and [b]$p<0.01$, for musicians data. Parameters: HL (hearing level), DPOAE, ripple spacing, height, and ripple prevalence.

Factor	HL	DPOAE	Spacing	Height	Prevalence
Before/after (rehearsal)	0.9345	0.5283	0.3335	0.2733	0.0942
Left/right (ear)	0.9702	0[b]	0.7833	0.2296	0.1802
Frequency	0[b]	0[b]	0[b]	0[b]	0[b]
Bef/after*Left/right	0.8169	0.0002[b]	0.3974	0.2803	0.9331
Bef/after*Freq	0.989	1	0.4968	0.8628	0.9962
Left/right*Freq	0.6941	1	0.394	0.0211[a]	0.9428

III. RESULTS

A. Individual data

Figure 5 shows the DPOAE fine structure measurements for all individual subjects. For each subject there are two figures: the left panels show left ear data obtained before and after rehearsal, right panels show the corresponding right ear data. Also shown in each figure is the noise floor and a schematic fine structure, which is obtained by calculating the ripple center frequency, ripple spacing, and height of the fine structure ripples. The schematic fine structure depicts the presence of fine structure ripples. None of the subjects have fine structure ripples over the entire measured frequency range. For most subjects the fine structure is prevalent in the frequency range 1–3 kHz. Some subjects have more pronounced fine structure than others, e.g., T3 has a pronounced fine structure, whereas T7 and T13 have very little fine structure, most ripples being below the 3 dB ripple height criterion. When the data from the left and right ear of the musicians are compared, it stands out that DPOAE are similar for the two ears of individual subjects. Also the ripple center frequencies are similar for left and right ears of individual subjects (Fig. 6).

B. Before rehearsal vs after rehearsal

Figure 7 shows the mean hearing levels before and after the rehearsal. Mean hearing levels are around 0 dB between 125 Hz and 2 kHz and up to 8 dB at 4 and 8 kHz. Compared to the 0.50 fractile data for the ISO 7029 (2000) the musicians have slightly better hearing levels. There is no significant difference between the hearing levels taken before and after rehearsal (the results of the analysis of variance (ANOVA) are shown in Table I).

The mean DPOAE levels measured before and after the exposure are plotted in Fig. 8. There is no significant difference in DPOAE level before and after the rehearsal (Table I), i.e., the exposure during rehearsal does not cause a measurable DPOAE shift. The fine structure parameters (ripple spacing, height, and prevalence) were determined according to the automatic procedure described in Sec. II D. The results of this analysis are shown in Fig. 9, showing the parameters before and after rehearsal for the left and right ears. All data are illustrated in dependence of the primary frequency f_2. Mean values and standard deviations are obtained for 11 subjects. The data are presented in 1/8 octave bands for ripple spacing and height and in 1/3 octave bands for ripple prevalence. The ripple widths increase from 100 to 300 Hz with increasing frequency, corresponding to a decrease from 1/8 to 3/32 octaves. Ripple heights of up to 27 dB are found for one individual subject. On average ripple heights are between 4 and 12 dB. The mean ripple prevalence is 2–4 ripples per 1/3 octave band, the prevalence being highest at 2.5 kHz. No statistical significant differences were found for any of the fine structure parameters for the measurements taken before and after the rehearsal (Table I).

TABLE II. Unbalanced two-way-ANOVA p values, [a]$p<0.05$ and [b]$p<0.01$, for four subject groups. Parameters: DPOAE, ripple spacing, height, and ripple prevalence.

Factor	DPOAE	Spacing	Height	Prevalence
Frequency	0[b]	0[b]	0[b]	0[b]
Group	0[b]	0.001[b]	0.0032[b]	0.1353
Freq*Group		0.0485[a]	0.4848	0.2137

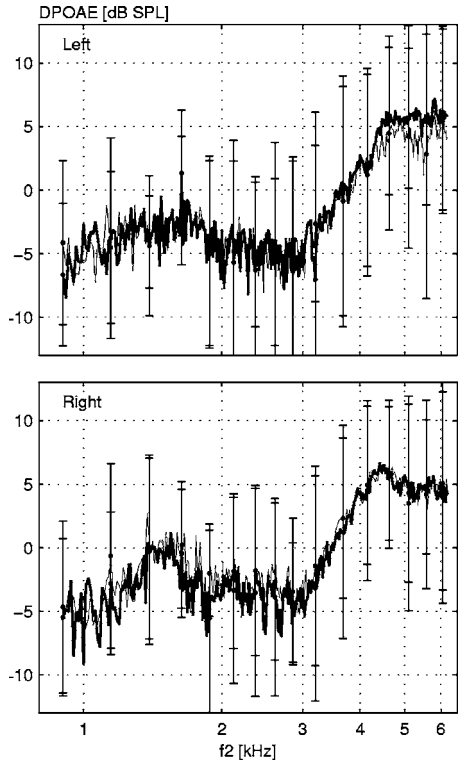


FIG. 8. DPOAE before (thin line) and after (thick line) the rehearsal for left ears (top panel) and right ears (lower panel). Error bars are standard deviations (for every 20th measurement) between the 11 subjects.

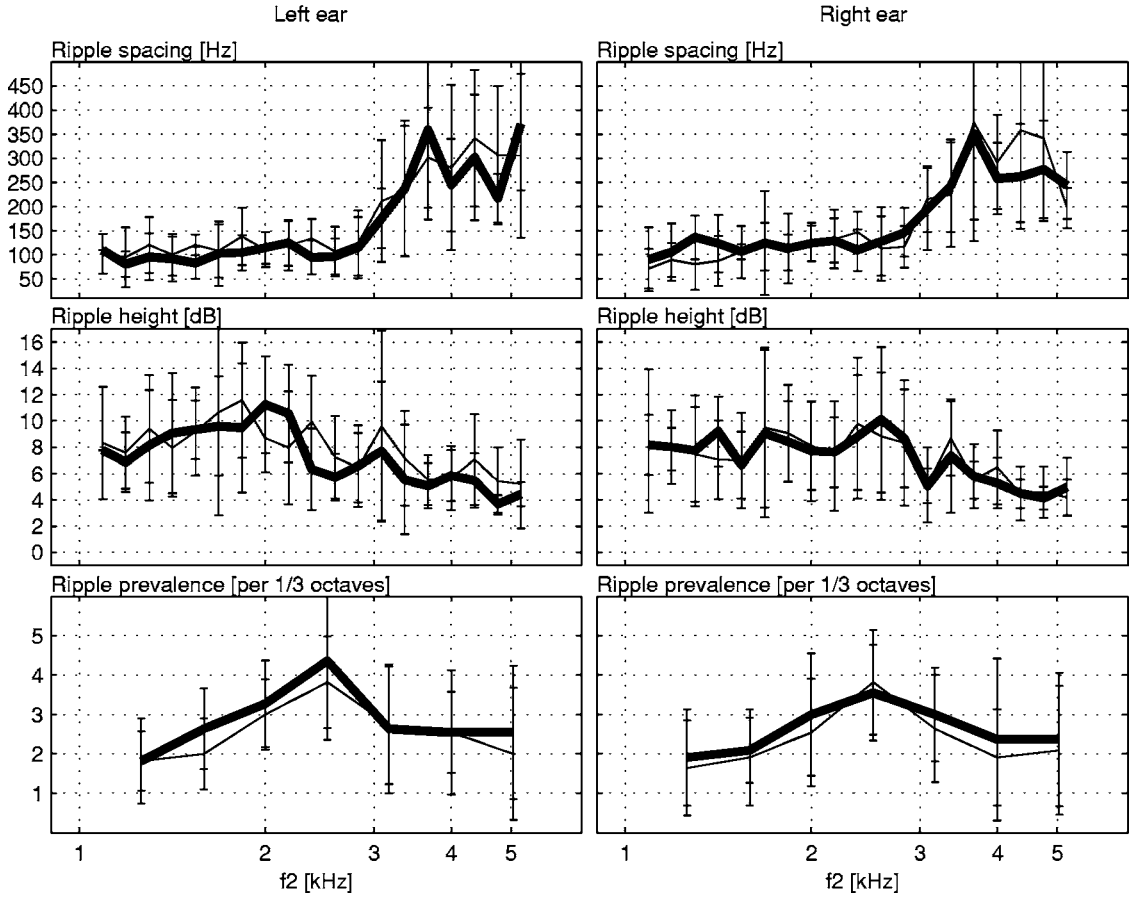


FIG. 9. Ripple parameters **before** rehearsal (thin line) and **after** rehearsal (thick line) for left ears (left panels) and right ears (right panels). Error bars are standard deviations between subjects.

C. Left ear vs right ear

Both hearing levels, DPOAE levels, and fine structure parameters are similar for left and right ears of the musicians. A small difference between left and right ears can be seen at 4 kHz. At that frequency the hearing level is slightly better for right ears (Fig. 7), the mean DPOAE level of the right ears is slightly higher (Fig. 8), and the ripple prevalence is slightly lower for the right ear (Fig. 10). A statistical significant difference between left and right ears could be detected for the DPOAE level only (Table I).

D. Group differences

Hearing levels and DPOAE fine structures of the musicians (Group M) are compared to similar data obtained from three other groups of subjects: a control group of age and gender matched, occupationally nonexposed subjects (group C), a group of 39 university students having very good hearing levels (selected from a group of 50 near-normal hearing students, the hearing level at only one frequency may exceed 10 dB, named Group A), and a group of 11 university students having hearing levels exceeding 10 dB for at least two frequencies, named Group B (data from Groups A and B have been previously reported in Reuter and Hammershøi (2006)). In the group analysis the data obtained for left and right ears of Group M (before rehearsal) and Group C are combined. For Groups A and B the data for one ear only were obtained. The ear was randomly selected. The data were analyzed using an ANOVA test between all groups (Table II) and between two groups at a time (Table III).

The hearing levels of the subjects belonging to the four groups can be seen in Fig. 11. The hearing levels of Groups M and C are rather similar, the hearing levels are slightly raised at the high frequencies. The hearing levels of Groups A and B are flat in the measured frequency range. The mean hearing levels for subjects of Group B are clearly raised compared to the hearing levels of Group A subjects.

The DPOAE overall levels of the four groups are shown in Fig. 12. Group A subjects have significantly higher DPOAE levels than the subjects of the other three groups over the entire measured frequency range (Table III). The overall DPOAE levels of the other three groups are rather similar. At the lowest frequencies, the DPOAE of Group B is higher than the DPOAE of the older subject groups (Groups M and C), whereas at the highest frequencies the DPOAE level of Group B subjects is lowest. The shape of DPOAE levels for Groups M and C is very similar.

Figure 13 shows the obtained DPOAE fine structure parameters for all four groups. Similar results for all four groups are obtained for the ripple prevalence. Significant differences between the groups are obtained for the parameters ripple width and ripple height (Table II). In the lowest frequency range Group A subjects have significantly broader ripples than subjects of the other groups. The ripple height of Group B is significantly higher than the ripple heights of Group A and C (Table III).

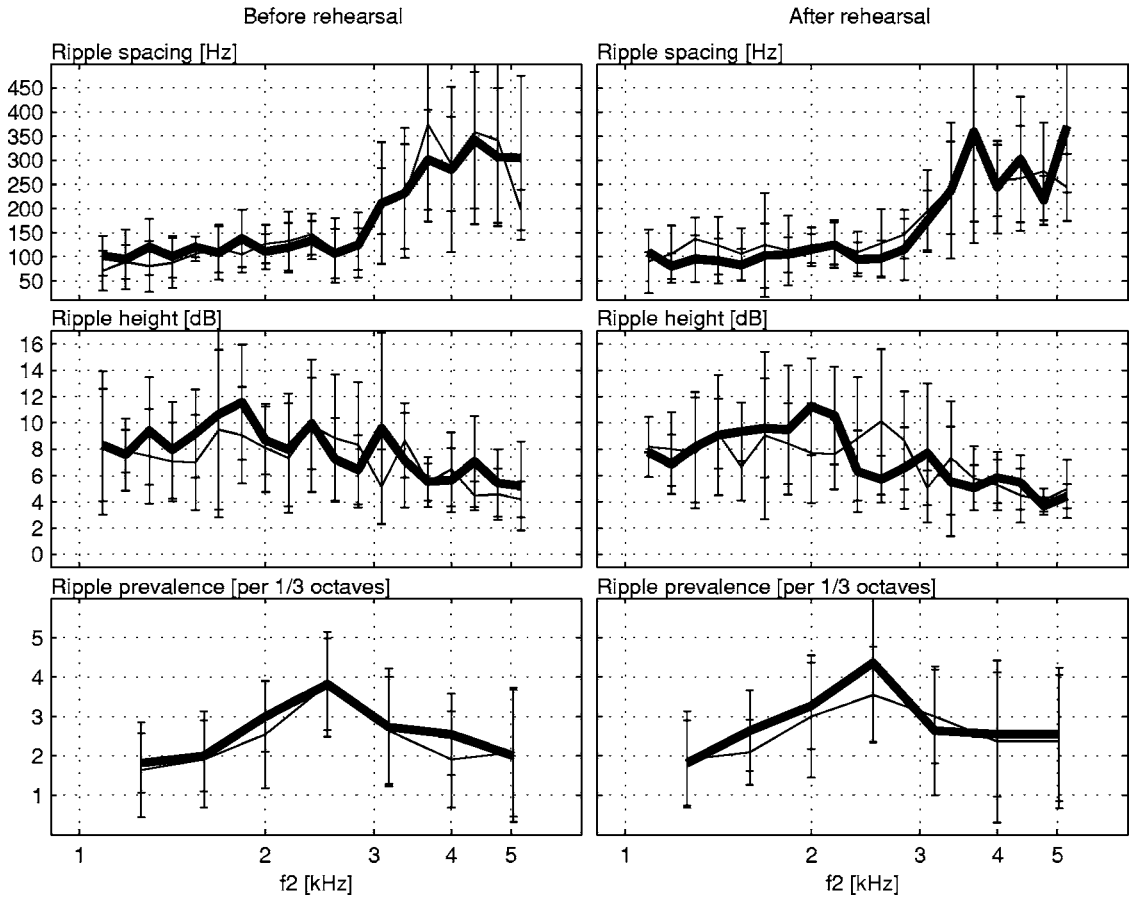


FIG. 10. Ripple parameters **right** ear (thin line) and **left** ear (thick line) before rehearsal (left panels) and after rehearsal (right panels). Error bars are standard deviations between subjects.

IV. DISCUSSION

A. Temporary changes

It was tested in this study, whether the DPOAE fine structure parameters might be sensitive to temporary changes in the auditory system after the exposure of a symphony orchestra rehearsal. None of the measures (hearing levels, DPOAE levels, and DPOAE fine structure characteristics) showed any changes after the rehearsal. This might have different reasons or a combination of these: (1) The exposure levels during the rehearsal are too low to cause any changes in the auditory system. (2) The recovery time of eventual changes is too fast, and the test time of the DPOAE fine structure is too long to detect any changes, or (3) none of the measures is sensitive enough to detect effects of overexposure.

Compared to other studies, which observed temporary threshold shift (TTS) or OAE level shifts after an overexposure, the exposure levels of around $L_{\mathrm{Aeq,4h}}=80$ dB in the present study are rather low. Royster *et al.* (1991) measured mean L_{Aeq} values of 89.9 dB during the Chicago Symphony Orchestra rehearsals, Murray and LePage (1998) measured weekly exposure levels of the Sydney Orchestra musicians in

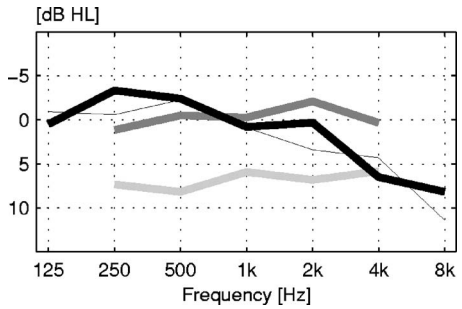


FIG. 11. Hearing levels of four groups of subjects: Group M (12 musicians, black, thick line), Group C (12 controls, black, thin line), Group A (39 students with good hearing levels, gray, thick line), Group B (11 students with slightly raised hearing levels, light gray, thick line).

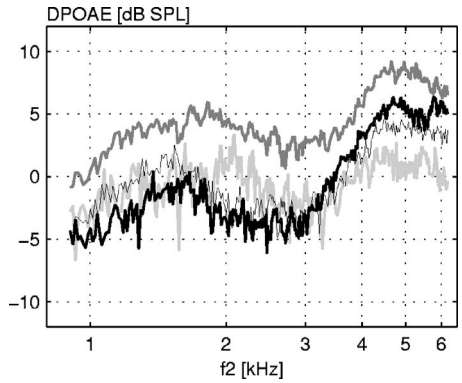


FIG. 12. DPOAE of four groups of subjects: Group M (12 musicians, black, thick line), Group C (12 controls, black, thin line), Group A (39 students with good hearing levels, gray, thick line), Group B (11 students with slightly raised hearing levels, light gray, thick line).

TABLE III. Unbalanced two-way-ANOVA p values, [a]$p<0.05$ and [b]$p<0.01$. Comparison of four subject groups. musicians (Gr M), control group (Gr C), students with good hearing levels (Gr A), and students with slightly raised hearing levels (Gr B). The data of two groups are compared at a time. Parameters: DPOAE, ripple spacing, height, and ripple prevalence.

Group	Factor	DPOAE	Spacing	Height	Prevalence
	Frequency	0[b]	0[b]	0[b]	0[b]
Gr M - Gr C	Group	0.2026	0.8852	0.1481	0.5514
	Freq*Group		0.6012	0.2433	0.9479
	Frequency	0[b]	0[b]	0.0001[b]	0[b]
Gr M - Gr A	Group	0[b]	0.0315[a]	0.2621	0.6156
	Freq*Group		0.1371	0.8844	0.55
	Frequency	0[b]	0[b]	0[b]	0[b]
Gr M - Gr B	Group	0.0804	0.1478	0.048[a]	0.0261[a]
	Freq*Group		0.0019[b]	0.6351	0.2034
	Frequency	0[b]	0[b]	0[b]	0[b]
Gr C - Gr A	Group	0[b]	0.0378[a]	0.739	0.8918
	Freq*Group		0.6122	0.2492	0.8536
	Frequency	0[b]	0[b]	0[b]	0[b]
Gr C - Gr B	Group	0.0052[b]	0.0849	0.0016[b]	0.1039
	Freq*Group		0.3586	0.145	0.1395
	Frequency	0[b]	0[b]	0[b]	0[b]
Gr A - Gr B	Group	0[b]	0.0003[b]	0.0014[b]	0.0444[a]
	Freq*Group		0.029[a]	0.4484	0.0209[a]

the 85–90 dBA range. In the latter study sound levels were monitored at various positions in the orchestra, including the position in front of the conductor. On average L_{Aeq} of 85 dB at the conductor position and average L_{Aeq} of 90 dBA immediately in front of the brass were measured. Since several musicians, that participated in this study, are seated in the area in front of the brass, their exposure level is probably higher than the sound levels monitored in front of the conductor. The purpose of the monitoring of sound pressure level in this study was to observe whether subjects are exposed to similar exposures on the different days, since the hearing of only one subject was monitored at any one specific day. Strasser *et al.* (1999) could observe a TTS in subjects after they have been exposed to classical music in the laboratory. At 2 min after 1 h of exposure to classical music (L_{Aeq}=94 dB) they observed a mean TTS of 11.4 dB, which recovered within 1 h. In the present study the monitoring of both hearing levels and DPOAE fine structures for both ears resulted in a relatively long test time, and possible temporary changes might vanish during the examination duration. However, the exposure duration of 4 h is a rather long exposure time. If the exposure causes any temporary auditory changes, a relatively long recovery time would also be expected.

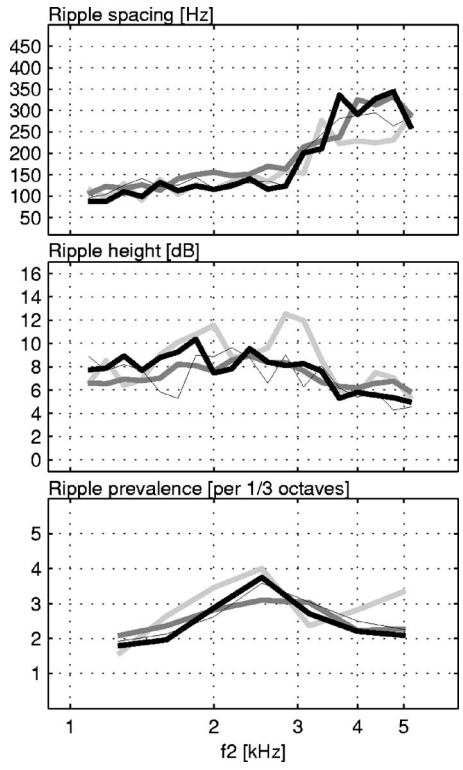


FIG. 13. Ripple parameters of four groups of subjects: Group M (12 musicians, black, thick line), Group C (12 controls, black, thin line), Group A (39 students with good hearing levels, gray, thick line), Group B (11 students with slightly raised hearing levels, light gray, thick line). The data of Groups A and B have previously been presented in Reuter and Hammershøi (2006), Figs. 12–14.

B. Group differences

Hearing levels and DPOAE levels of four groups of subjects were compared in the present study. The groups differ in age and/or exposure history. By comparing the DPOAE data of these different groups, it could be tested whether DPOAE overall levels and DPOAE fine structure characteristics differ from each other. The DPOAE level of Group A subjects is significantly higher than the DPOAE levels of all other groups of subjects over the entire measured frequency range. When comparing the DPOAE fine structure characteristics of Group A subjects to the other groups of subjects, there is a difference in ripple spacing and ripple height between Group A and Group B subjects, but there is no significant difference in fine structure characteristics between

Group A subjects and the other groups of subjects. The DPOAE levels of groups B, C, and M are rather similar. It is not known for certain whether there are differences in the state of hearing between these groups. Therefore it is not possible to conclude whether the DPOAE fine structure is a more sensitive measure to detect early hearing losses than the DPOAE level. A difference in ripple prevalence between groups was noticed around 3 kHz, though. A sharp bend at that frequency can be observed for left ears compared to right ears and left ears after the rehearsal compared to left ears before rehearsal. Group B subjects also show this bend of ripple prevalence around 3 kHZ.

ACKNOWLEDGMENTS

This work was financed by the William Demant Foundation (Oticon) and the Danish Technical Research Council. The authors would like to thank all subjects for participating in this experiment. The cooperation with "Aalborg Symfoniorkester," in particular Palle Kjeldgaard and Sonja Abildgaard, is greatly acknowledged. We would like to thank Rodrigo Ordoñez for assistance in data collection and committed discussion on data analysis.

Attias, J., Furst, M., Furman, V., Reshef, I., Horowitz, G., and Bresloff, I. (**1995**). "Noise-induced otoacoustic emission loss with or without hearing loss," Ear Hear. **16**, 612–618.

Avan, P., Bonfils, P., Loth, D., Narcy, Ph., and Trotoux, J. (**1991**). "Quantitative assessment of human cochlear function by evoked otoacoustic emissions," Hear. Res. **52**, 99–112.

Bonfils, P., and Avan, P. (**1992**). "Distortion-product otoacoustic emissions - Values for clinical use," Arch. Otolaryngol. Head Neck Surg. **118**, 1069–1076.

Desai, A., Reed, D., Cheyne, A., Richards, S., and Prasher, D. (**1999**). "Absence of otoacoustic emissions in subjects with normal audiometric thresholds implies exposure to noise," Noise and Health **2**, 58–65.

Early, K. L., and Horstman, S. W. (**1996**). "Noise exposure to musicians during practice," Appl. Occup. Environ. Hyg. **11**, 1149–1153.

Engdahl, B., and Kemp, D. T. (**1996**). "The effect of noise exposure on the details of distortion product otoacoustic emissions in humans," J. Acoust. Soc. Am. **99**, 1573–1587.

Gaskill, S. A., and Brown, A. M. (**1993**). "Comparing the level of the acoustic distortion product $2f_1-f_2$ with behavioural threshold audiograms from normal-hearing and hearing-impaired ears," Br. J. Audiol. **27**, 397–407.

Gaskill, S. A., and Brown, A. M. (**1990**). "The behavior of the acoustic distortion product, $2f_1-f_2$, from the human ear and its relation to auditory sensitivity," J. Acoust. Soc. Am. **88**, 821–839.

Gorga, M. P., Neely, S. T., Bergman, B., Beauchaine, K. L., Kaminski, J. R., Peters, J., and Jesteadt, W. (**1993**). "Otoacoustic emissions from normal-hearing and hearing-impaired subjects: Distortion product responses," J. Acoust. Soc. Am. **93**, 2050–2060.

He, N., and Schmiedt, R. A. (**1993**). "Fine structure of the $2f_1-f_2$ acoustic distortion product: Changes with primary level," J. Acoust. Soc. Am. **94**, 2659–2669.

Heitmann, J., Waldmann, B., and Plinkert, P. K. (**1996**). "Limitations in the use of distortion product otoacoustic emissions in objective audiometry as the result of fine structure," Eur. Arch. Otorhinolaryngol. **253**, 167–171.

Heitmann, J., Waldmann, B., and Schnitzler, H. (**1998**). "Suppression of distortion product otoacoustic emissions (DPOAE) near $2f_1-f_2$ removes DP-gram fine structure - Evidence for a secondary generator," J. Acoust. Soc. Am. **103**, 1527–1531.

Hotz, M. A., Probst, R., Harris, F. P., and Hauser, R. (**1993**). "Monitoring the effects of noise exposure using transiently evoked otoacoustic emissions," Acta Oto-Laryngol. **113**, 478–482.

Kalluri, R., and Shera, C. A. (**2001**). "Distortion-product source unmixing: A test of the two-mechanism model for DPOAE generation," J. Acoust. Soc. Am. **109**, 622–637.

Karlsson, K., Lundquist, P. G., and Olaussen, T. (**1983**). "The hearing of symphony orchestra musicians," Scand. Audiol. **12**, 257–264.

Kähäri, K. R., Axelsson, A., Hellström, P.-A., and Zachau, G. (**2001**). "Hearing development in classical orchestral musicians. A follow-up study," Scand. Audiol. **30**, 141–149.

Knight, R. D., and Kemp, D. T. (**2000**). "Indications of different distortion product otoacoustic emission mechanisms from a detailed f_1,f_2 area study," J. Acoust. Soc. Am. **107**(1), 457–473.

Konrad-Martin, D., Neely, S. T., Keefe, D. H., Dorn, P. A., Cyr, E., and Gorga, M. P. (**2002**). "Sources of DPOAEs revealed by suppression experiments, inverse fast Fourier transforms, and SFOAEs in impaired ears," J. Acoust. Soc. Am. **111**, 1800–1809.

Kowalska, S., and Sulkowski, W. (**1997**). "Measurements of click-evoked otoacoustic emission in industrial workers with noise-induced hearing loss," Int. J. Occup. Med. Environ. Health **10**, pp. 441–459.

LePage, E. L., and Murray, N. M. (**1998**). "Latent cochlear damage in personal stereo users: A study based on click-evoked otoacoustic emissions," Med. J. Aust. **169**, 588–592.

Long, G. R., Talmadge, C. L., and Lee, J. (**2004**). "Using Sweeping Tones to Evaluate DPOAE Fine Structure," Abstract of the Midwinter Meeting of the Association for Research in Otolaryngology.

Lucertini, M., Moleti, A., and Sisto, R. (**2002**). "On the detection of early cochlear damage by otoacoustic emission analysis," J. Acoust. Soc. Am. **111**, 972–978.

Mauermann, M., Uppenkamp, S., van Hengel, P. W. J., and Kollmeier, B. (**1999a**). "Evidence for the distortion product frequency place as a source of distortion product otoacoustic emission (DPOAE) fine structure in humans. I. Fine structure and higher-order DPOAE as a function of the frequency ratio f_2/f_1," J. Acoust. Soc. Am. **106**, 3473–3483.

Mauermann, M., Uppenkamp, S., van Hengel, P. W. J., and Kollmeier, B. (**1999b**). "Evidence for the distortion product frequency place as a source of distortion product otoacoustic emission (DPOAE) fine structure in humans. II. Fine structure for different shapes of cochlear hearing loss," J. Acoust. Soc. Am. **106**, 3484–3491.

Mauermann, M., Uppenkamp, S., and Kollmeier, B. (**1997**). "Periodizität und Pegelabhängigkeit der spektralen Feinstruktur von Verzerrungsprodukt-Emissionen" ("Periodicity and dependence on level of the distortion product otoacoustic emission spectral fine structure"), Audiol. Akustik **36**, 92–104.

Murray, N., and LePage, E. (**1998**). "Inner ear damage in an opera theatre orchestra as detected by otoacoustic emissions, pure tone audiometry and sound levels," Aust. J. Audiol. **20**, 67–78.

Obeling, L., and Poulsen, T. (**1999**). "Hearing ability in Danish symphony orchestra musicians," Noise Health **1**, 10–27.

Rao, A., Long, G. R., Narayan, S., and Dhar, S. (**1996**). "Changes in the temporal characteristics of TEOAE and the fine structure of DPOAEs with aspirin consumption," *19th ARO Midwinter Research Meeting*, Abstract p. 27.

Reuter, K., and Hammershøi, D. (**2006**). "Distortion product otoacoustic emission fine structure analysis of 50 normal-hearing humans," J. Acoust. Soc. Am. **120**, 270–279.

Royster, J. D., Royster, L. H., and Killion, M. C. (**1991**). "Sound exposures and hearing thresholds of symphony orchestra musicians," J. Acoust. Soc. Am. **89**, 2793–2803.

Sliwinska-Kowalska, M., and Kotylo, P. (**2002**). "Occupational exposure to noise decreases otoacoustic emission efferent suppression," Int. J. Audiology **41**, 113–119.

Strasser, H., Irle, H., and Scholz, R. (**1999**). "Physiological cost of energy-equivalent exposures to white noise, industrial noise, heavy metal music, and classical music," Noise Control Eng. J. **47**, 187–192.

Talmadge, C. L., Tubis, A., Long, G. R., and Piskorski, P. (**1998**). "Modeling otoacoustic emission and hearing threshold fine structures," J. Acoust. Soc. Am. **104**, 1517–1543.

Talmadge, C. L., Long, G. R., Tubis, A., and Dhar, S. (**1999**). "Experimental confirmation of the two-source interference model for the fine structure of distortion product otoacoustic emissions," J. Acoust. Soc. Am. **105**, 275–292.

Properties of distortion product otoacoustic emissions and neural suppression tuning curves attributable to the tectorial membrane resonance

Andrei N. Lukashkin,[a] Judith K. Smith, and Ian J. Russell
School of Life Sciences, University of Sussex, Falmer, Brighton, BN1 9QG, United Kingdom

(Received 12 June 2006; revised 10 October 2006; accepted 16 October 2006)

Mechanically coupled cochlear structures are likely to form a resonator with several degrees of freedom. Consequently one can expect complex, frequency-dependent relative movements between these structures, particularly between the tectorial membrane and reticular lamina. Shearing movement between these two structures excites the cochlear receptors. This excitation should be minimal at the frequency of the hypothesized tectorial membrane resonance. In each preparation, simultaneous masking neural tuning curves and distortion product otoacoustic emissions were recorded. The position of the low-frequency minima in the tuning curves, frequency dependence of the emission bandpass structure, and level-dependent phase reversal were compared to determine if they were generated by a common phenomenon, for example the tectorial membrane resonance. The notch in the masking curves and the phase inversion of the emission growth functions at the auditory thresholds are both situated half an octave below the probe frequency and the high-frequency primary, respectively, and show similar frequency dependence. The emission bandpass structure is, however, likely to be generated by a combination of mechanisms with different ones dominating at different stimulus parameters. © *2007 Acoustical Society of America.*
[DOI: 10.1121/1.2390670]

PACS number(s): 43.64.Jb, 43.64.Kc, 43.64.Ri, 43.64.Nf [BLM] Pages: 337–343

I. INTRODUCTION

Elastically and hydromechanically coupled cochlear structures are likely to form a complex resonator with several degrees of freedom (Zwislocki, 2002). Consequently one can expect quite complex frequency dependence for the relative movements between these structures, for example between the tectorial membrane (TM) and the reticular lamina. Relative movement between the TM and the reticular lamina can have a significant transversal component at low frequencies (Nowotny and Gummer, 2006) as a consequence of the viscoelastic properties of the organ of Corti (Scherer and Gummer, 2004). Shear displacement, however, dominates their relative movement and hence excites the cochlear receptors, the inner and outer hair cells (IHCs and OHCs, respectively), near the characteristic frequency (CF) (Nowotny and Gummer, 2006). In 1980, Allen, and Zwislocki, postulated independently that this complex system possesses at least two resonant frequencies for each tonotopic place along the cochlea.

Allen and Fahey (1993) produced a realistic model of the resonator which is likely to be formed by the TM mass and its elastic attachments to the spiral limbus and the OHC stereocilia (Zwislocki, 1980; Allen, 1980). At the frequency of the resonance, IHC and, consequently, neural excitation, reach a minimum because the TM moves in phase with the reticular lamina, thus resulting in a minimum in the shear displacement between these structures. Allen and Fahey (1993) identify this frequency f_Z with the point at which the tip of the neural tuning curve joins its tail. A secondary minimum in sensitivity is frequently seen at this point (Zwislocki, 1986; Taberner and Liberman, 2006) which is also situated about half an octave below the CF of the auditory fiber. Simultaneous masking neural tuning curves closely resemble the tuning properties of single auditory-nerve fibers (Dallos and Cheatham, 1976). Hence, masking tuning curves might also exhibit this secondary minimum of sensitivity attributed to the TM resonance.

Depending on the relationship between the system parameters, an increase in the basilar membrane (BM) responses could be observed at the TM resonance frequency (Russell and Kössl, 1999; Legan *et al.*, 2000). As a consequence of this complex interaction between the cochlear structures, the TM provides the main drive to the IHCs at the CF, determining the sharpness and sensitivity of the neural tuning curves (Legan *et al.*, 2005).

To date, the TM resonance properties have been observed directly only in preparations of the temporal bone (Gummer *et al.*, 1996; Hemmert *et al.*, 2000). However, the existence of a nonlinear TM resonance in the living, sensitive cochleae has also been inferred from *in vivo* recording of mechanical (Russell and Kössl, 1999; Legan *et al.*, 2000), acoustical (Brown *et al.*, 1992; Allen and Fahey, 1993; Lukashkin and Russell, 2003), and neural (Zwislocki, 1984; Allen and Fahey, 1993) responses. The only known attempt to manipulate the TM mechanical properties *in vivo* (Zwislocki, 1986) produced results consistent with the concept of a radial TM resonance. However, Zwislocki's experiments involved opening the scala vestibuli with the possibility of changing the cochlear sensitivity and its hydromechanical

[a]Electronic-mail: a.lukashkin@sussex.ac.uk

properties (Cooper and Rhode, 1996), including changes of the mechanical responses of the TM and BM.

Distortion product otoacoustic emissions (DPOAEs) are commonly used because they offer a relatively noninvasive means of investigating cochlear function. DPOAEs are thought to originate from the region between the two primary tones where the oscillations from each component combine (Brown and Kemp, 1984). Hence for experimental and clinical purposes it is important that the primary tones are an optimal distance apart, i.e., the f_2/f_1 ratio is optimal. The optimal frequency ratio for the $2f_1$-f_2 DPOAE has been found to fall approximately between 1.1 to 1.4 in different species and different levels of the stimulation. The DPOAE frequency, which corresponds to this optimal ratio, is often described as being about half an octave below f_2 (Brown *et al.*, 1992) and it is similar for the DPOAE components of different order, e.g., $2f_1$-f_2, $3f_1$-$2f_2$, $4f_1$-$3f_2$, etc. Hence, the amplitude maximum is observed at different f_2/f_1 ratios for different DPOAE components. If the f_2/f_1 ratio is too great or too small (f_1 and f_2 are too far apart or too close together) there is a decrease in the amplitude of the DPOAEs (Brown *et al.*, 1992; Allen and Fahey, 1993). Thus DPOAEs exhibit a bandpass structure, the origin of which has been a subject of debate.[1]

This frequency of about half an octave below f_2, which corresponds approximately to the maximum of the DPOAE amplitude, is also characterized by the reversal of the DPOAE phase behavior (Harada *et al.*, 2000; Lukashkin and Russell, 2003) when phase lag turns into phase lead with increasing level of the stimulation. If f_2 is kept constant and f_1 changes, then the DPOAE frequencies will sweep through the frequency region of the hypothetical nonlinear, TM-attributed resonance at the f_2 CF place (Russell and Kössl, 1999; Legan *et al.*, 2000). The DPOAE phase leads with increasing level of stimulation when f_1 is close to f_2 and the level of the low-frequency primary increases. When the f_2/f_1 ratio becomes larger, i.e., when f_1 moves further away from f_2, this phase lead gradually decreases for the same level of the primaries. The level-dependent phase change disappears when the DPOAE frequency becomes equal to the frequency of the hypothesized TM resonance at about a half octave below f_2. The level-dependent phase change inverses and the phase angle lags with increasing the level of stimulation as the DPOAE frequency decreases below that of the resonant frequency. When considering this ratio-dependent phase behavior it is important to remember that the resonant frequency in nonlinear systems depends on the levels of the input signals; thus, phase reversal could occur at different ratios in the case of different primary levels (Lukashkin and Russell, 2003).

In this work, the DPOAEs and simultaneous masking neural tuning curves were recorded from the same preparations. The frequency dependence of the DPOAE maxima and level-dependent phase reversals, and also the position of the secondary minima in the masking tuning curves, were compared to determine whether they are generated by the same phenomenon, for example the secondary, low-frequency resonance attributable to the TM.

II. METHODS

Pigmented guinea pigs (280–390 g) were anesthetized with the neurolept anesthetic technique (0.06 mg/kg body weight atropine sulphate s.c., 30 mg/kg pentobarbitone i.p., 500 μl/kg Hypnorm i.m.). Additional injections of Hypnorm were given every 40 min. Additional doses of pentobarbitone were administered as needed to maintain a nonreflexive state. The heart rate was monitored with a pair of skin electrodes placed on both sides of the thorax. The animals were tracheotomized and artificially respired, and their core temperature was maintained at 38 °C with a heating blanket and head holder. The middle-ear cavity of the ear used for the DPOAE measurements was opened to reveal the round window. Compound action potentials (CAPs) of the auditory nerve were measured from the cochlear bony ridge in the proximity of the round window membrane using Teflon-coated silver wire.

Sound was delivered to the tympanic membrane by a closed acoustic system comprising two Bruel & Kjaer 4134 $\frac{1}{2}$-in. microphones for delivering tones and a single Bruel & Kjaer 4133 $\frac{1}{2}$-in. microphone for monitoring sound pressure at the tympanum. The microphones were coupled to the ear canal via 1-cm-long, 4-mm-diameter tubes to a conical speculum, the 1-mm diameter opening of which was placed about 1 mm from the tympanum. The closed sound system was calibrated *in situ* for frequencies between 1 and 50 kHz. Known sound-pressure levels were expressed in dB SPL *re*: 2×10^{-5} Pa. All sound stimuli in this work were shaped with raised cosines of 0.5 ms duration at the beginning and at the end of stimulation. White noise for acoustical calibration and tone sequences for auditory stimulation were synthesized by a Data Translation 3010 board at 200 kHz and delivered to the microphones through low-pass filters (100-kHz cutoff frequency). Signals from the measuring amplifier were digitized at 200 kHz using the same board and averaged in the time domain. Amplitudes and phase angles of the spectral peaks were obtained by performing an FFT on a time-domain averaged signal, 4096 points in length. Phase data were corrected for the phase angles of the primaries. The maximum level of the system distortion measured with an artificial ear cavity for the highest levels of primaries used in this study ($L_1=L_2=75$ dB SPL) was 70 dB below the primary level. Experimental control, data acquisition, and data analysis were performed using a PC with programs written in TESTPOINT (CEC, MA).

The following experimental procedures were used during DPOAE recording:

(i) DPOAE-grams (f_2 sweeps, f_2/f_1 ratio is constant, L_1 and L_2 are constant, L_2 is 10 dB below L_1) for low-level primaries were recorded on a regular basis during each experiment to confirm stable conditions of the animal. Data collected from an animal were rejected if the DPOAE level changed by more than 5 dB at f_2 frequencies used for the other procedures.
(ii) DPOAE ratio functions measured during f_1 sweeps with constant f_2. L_1 and L_2 were constant during each sweep with L_2 being 10 dB SPL below L_1.
(iii) DPOAE growth functions with increasing L_1 were

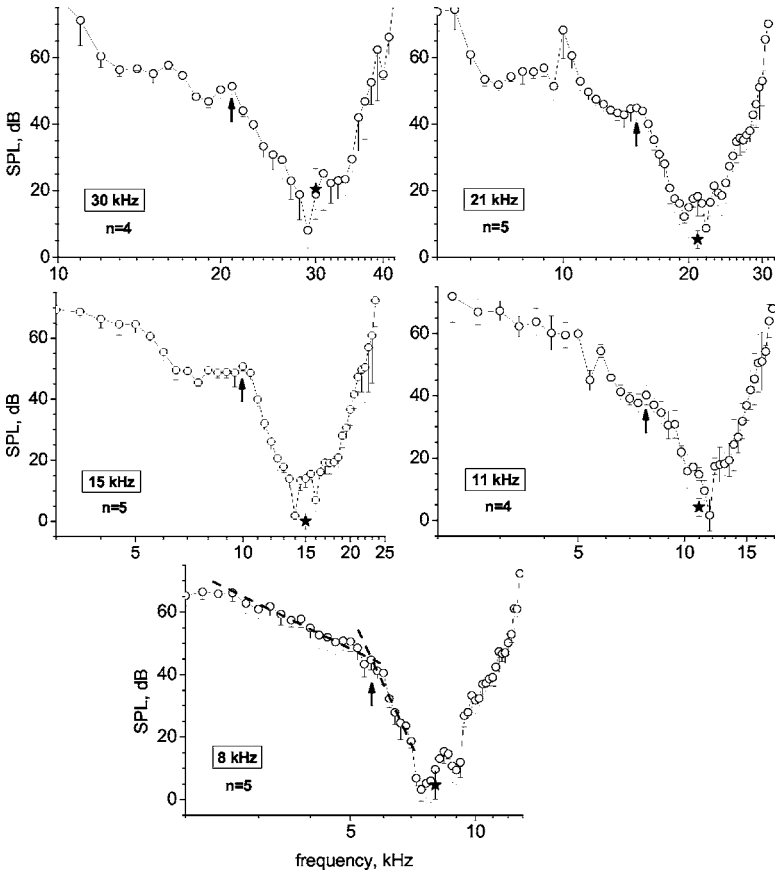


FIG. 1. Simultaneous masking neural tuning curves for different frequencies of the probe tone. The probe tone frequency is indicated within each panel. The probe tone frequencies and the probe tone levels are also indicated by filled stars. Vertical arrows point towards the frequency where the tip and tail of the curves meet. Masking tuning curves show a distinctive notch at these frequencies for the high-frequency probe tones. For the 8-kHz probe tone (bottom panel) this frequency is defined as an intersection of least-square fit lines (dashed lines) to the tail and near tip points of the curve. Number of averages n is indicated for each panel. Error bars are equal to SD.

measured while keeping L_2 and frequencies of the primaries constant. The growth functions were measured at different f_2/f_1 ratios.

CAP suppression tuning curves were derived from simultaneous tone-on-tone masking (Dallos and Cheatham, 1976) using a 10-ms probe tone centered on a 40-ms masker tone. The probe tone of a particular frequency was set to a level where a stable CAP appeared just above the recording noise floor. The frequency of the masker was set and its attenuation was adjusted until the probe tone CAP during every presentation of the masker was suppressed as judged by eye. The masker frequency and level was noted, a new masker frequency was set, and the process repeated. Signals were low-pass filtered at the corner frequency of 1–3 kHz to suppress the cochlear microphonic potentials during CAP measurements.

All procedures involving animals were performed in accordance with UK Home Office regulations.

III. RESULTS

A. Neural masking tuning curves

Allen and Fahey (1993) define a frequency f_Z in single unit neural tuning curves where the tip and tail of the tuning curves meet. Simultaneous masking neural tuning curves (Fig. 1) also have different slopes in the tail and tip regions; hence, a similar frequency f_Z can be determined for them. The f_Z in the masking curves (vertical arrow in Fig. 1) is situated almost exactly a half octave below the probe tone, and f_Z is characterized by a secondary minimum of masking tone efficiency which forms a distinctive notch in the masking curves recorded for the higher frequency probe tones (15–30 kHz in Fig. 1). Because the masking tuning curves closely resemble tuning properties of single auditory-nerve fibers (Dallos and Cheatham, 1976), the masking notch is likely to be analogous to the tuning notch observed in single unit responses (Zwislocki, 1986; Taberner and Liberman, 2006). The depth of the masking notch decreases when the probe tone frequency becomes lower, making it almost undetectable for a probe frequency of 8 kHz (Fig. 1). A similar systematic reduction in the prominence of the notch was also observed for single auditory fibers with lower CF (Zwislocki, 1986; Taberner and Liberman, 2006).

B. Level dependence of the DPOAE

The probe tone level during the masking experiments was set to a minimum level (close to 0 dB SPL for the probe frequencies of 8–21 kHz) where a stable CAP appeared just above the recording noise floor. It was not feasible to employ higher levels of the probe tone because they rapidly elevated the CAP threshold. However, during the DPOAE measurements we had to use levels of the primaries which were higher than the CAP probe tone ($L_1=L_2+10$ dB$=30$ dB SPL) for reliable recording of the higher order emission components, i.e., $3f_1$-$2f_2$, $4f_1$-$3f_2$, well above the noise floor. This difference in the stimulus levels should be taken into account when comparing the hypothesized manifestations of the TM resonance (f_Z and frequencies of the DPOAE maxima and the phase reversals) in neural and acoustical responses presented in this paper because of the level depen-

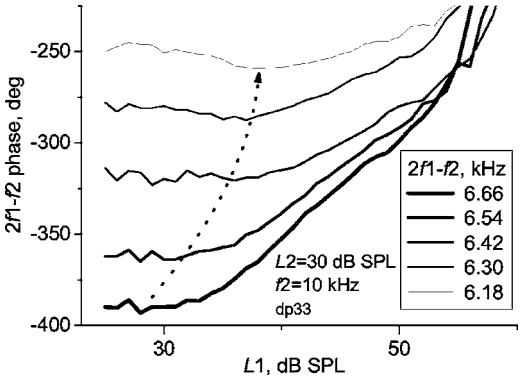


FIG. 2. Level dependence of the phase reversal observed for the $2f_1$-f_2 DPOAE. The $2f_1$-f_2 DPOAE frequency for each curve is shown inside the panel. Dotted arrow indicates a shift of the transition between phase lag and phase lead observed with reduction of the $2f_1$-f_2 frequency. L_2 and f_2 were held constant at 30 dB SPL and 10 kHz, respectively; f_1 was varied to obtain the required $2f_1$-f_2 frequency. Modified from Lukashkin and Russell (2003).

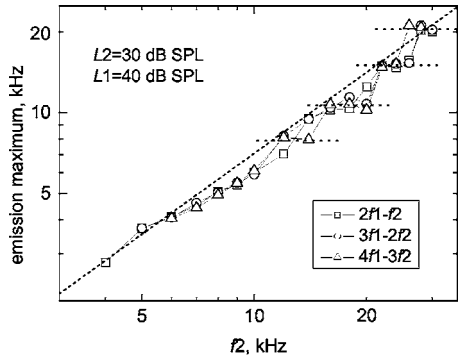


FIG. 4. Frequency dependence of the DPOAE maximum on the f_2 frequency. The DPOAE maxima were determined from records shown in Fig. 5 and similar records for lower f_2. Horizontal dotted lines show the position of plateaus where the frequency of the amplitude maximum is the same for different f_2. These dotted lines are equivalent to the vertical dotted lines in Fig. 5. Diagonal dashed line corresponds to frequencies situated half an octave below f_2.

dence of the DPOAEs. For example, reversal of the DPOAE phase behavior from lag to lead is observed over a finite range of DPOAE frequencies as f_1 is moved further away from f_2 so that the $2f_1$-f_2 frequency decreases (Fig. 2). Initially only low-level responses phase lag, i.e., the curves in Fig. 2, have negative slopes at low levels, and the phase still leads at higher L_1. The phase lag spreads towards progressively higher L_1 with further reduction of the $2f_1$-f_2 frequency (dotted arrow in Fig. 2). This phenomenon is explained by the fact that the resonance frequency of a dynamic nonlinear system depends on the level of the input signals. The resonance frequency may shift up or down depending on the stimulus levels and on the properties of the particular nonlinear system. Accordingly, the phase pattern seen in Fig. 2 for the $2f_1$-f_2 component would be expected if the resonance frequency becomes lower when the level of the stimulus (L_1) is increased.

Similar shifts to lower frequencies with increasing level of stimulation are observed for the bandpass DPOAE structure. For example, the $2f_1$-f_2 DPOAE peaks at a frequency of about 8.1 kHz when L_2=10 dB SPL (Fig. 3). However, when

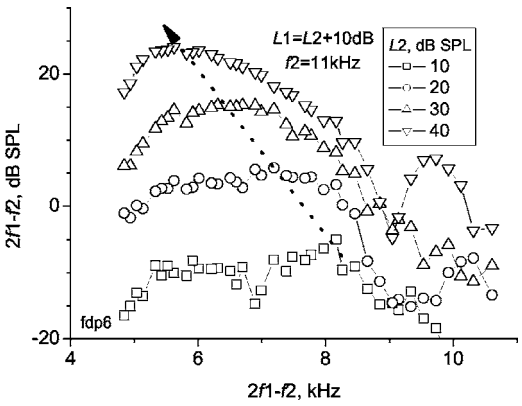


FIG. 3. Shift of the emission amplitude maximum towards lower frequencies at high levels of stimulation. Dotted arrow indicates this shift. f_2 was fixed at 11 kHz; f_1 was varied to obtain the required $2f_1$-f_2 frequency. $L2$ for each curve is indicated within the panel. $L1$ was 10 dB above $L2$. Noise floor for the data presented is below −20 dB SPL.

L_2 approaches 40 dB SPL the frequency of the $2f_1$-f_2 maximum decreases gradually to approximately 5.7 kHz. Because of the level dependence of the DPOAE characteristics, comparisons between the frequency dependencies of the DPOAE bandpass structure and DPOAE phase reversals were made for the same levels of the primaries.

C. Frequency dependence of the DPOAE amplitude maximum and phase reversal

For chosen stimulus levels the DPOAE amplitude maxima are situated exactly half an octave below the high-frequency primary for f_2 smaller than 7 kHz (Fig. 4). However, the maximum's position deviates from this "half an octave below" rule in the midfrequency range of 8–15 kHz. It is worth noting that this frequency range corresponds to the region of the greatest auditory sensitivity in guinea pigs and, hence, to the lowest levels of the probe tone used during masking experiments. For f_2 above about 12 kHz dependence of the DPOAE maxima on f_2 is not gradual and shows horizontal plateaus (Fig. 4). These plateaus are separated almost exactly by half-octave intervals and originate from the multiple peaks of high-frequency DPOAEs (Fig. 5) and the shift in prominence between these peaks as f_2 decreases. As the frequency of f_2 declines the lower frequency peaks become more prominent than the higher frequency peaks. However, this shift in prominence does not occur at the same frequency for each of the different DPOAEs. On closer examination it is obvious that each of the different DPOAE components peaks at different f_2 frequencies. For example, for the same f_2 frequency of 26 kHz (Fig. 5), the $2f_1$-f_2, $3f_1$-$2f_2$ DPOAEs are dominated by the local maximum at 15.2 kHz; however, the $4f_1$-$3f_2$ component peaks at local maximum with half an octave higher frequency of 20.5 kHz.

From comparisons of the frequency dependencies of the positions for the maximum DPOAE amplitude, phase reversal, and low-frequency notch of the masking neural tuning curves, it can be seen that these three dependencies approximately coincide at the highest frequencies (Fig. 6). The frequency dependencies of the amplitude maximum and phase reversal also follow the same trend at the lowest f_2 frequen-

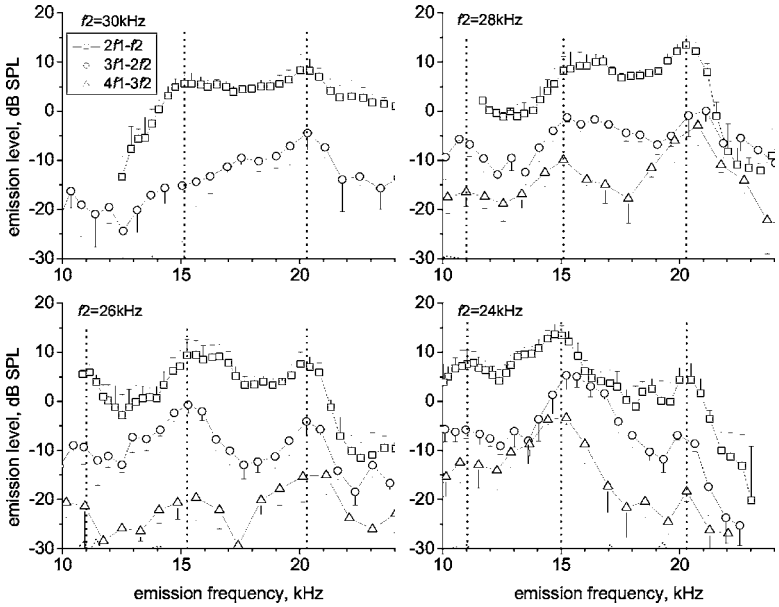


FIG. 5. Multiple peaks of the high-frequency DPOAEs observed at half-octave intervals. Averaged responses for four preparations. Symbol coding for different DPOAEs is shown at the top left panel. Vertical dotted lines are placed near peak frequencies. f_2 is indicated for each panel; f_1 was varied to obtain the required emission frequency. Primary levels were fixed at $L_1 = L_2 + 10$ dB $= 40$ dB SPL. Average noise floor is indicated with dotted lines. It is below -30 dB SPL for almost the entire frequency range presented.

cies (Fig. 6). However, at the lowest f_2 frequencies we cannot make direct comparisons between the DPOAE data and masking notch because masking neural tuning curves could not be obtained with probe tones at frequencies below 8 kHz. For the particular levels of the primaries and levels of the probe tone used in this work, the largest discrepancy between these three phenomena is observed in the region of f_2 and probe tones between 8 and 15 kHz. This is the range of the guinea pig's greatest auditory sensitivity and the range of the largest difference between the levels of the primaries and the probe tones used in these experiments.

IV. DISCUSSION

The frequency f_Z in the neural masking tuning curves, which is frequently characterized by a secondary minimum of sensitivity, is situated almost exactly a half octave below the probe tone for the entire frequency range studied (Fig. 6). Frequencies of the DPOAE maxima and the phase reversals follow the same "half-octave" dependence (in this case half an octave below f_2) for the lowest and the highest frequencies, which may indicate similar mechanisms behind the origin of these phenomena and the neural notch. However, the DPOAE maxima and the reversals are situated significantly lower than half an octave below f_2 in the midfrequency range, which corresponds to the region of the lowest hearing thresholds. Two observations may account for this discrepancy. The probe tone level (0–5 dB SPL) in the midfrequency range is smaller than the fixed L_2 of 30 dB SPL used for the DPOAE measurements. It was not feasible to employ either higher levels of the probe tone (due to rapid elevation of the CAP threshold) or lower L_2 (because the high-order DPOAEs were below the noise floor in this case). However, the level-dependent shifts of the DPOAE phase and magnitude (Figs. 2 and 3) indicate that both the DPOAE maxima and the phase reversal should be observed at frequencies 1.5–2 kHz higher than those recorded if it had been possible to use L_2 as low as 0 dB SPL. With this level and frequency adjustment, the frequencies of the hypothesized manifestations of the TM resonance in the DPOAEs occur a half octave below f_2. This argument is supported by the observation that frequencies of the acoustic and neural phenomena coincide at 30 kHz (Fig. 6) when the probe tone level (20 dB SPL) is closer to the L_2.

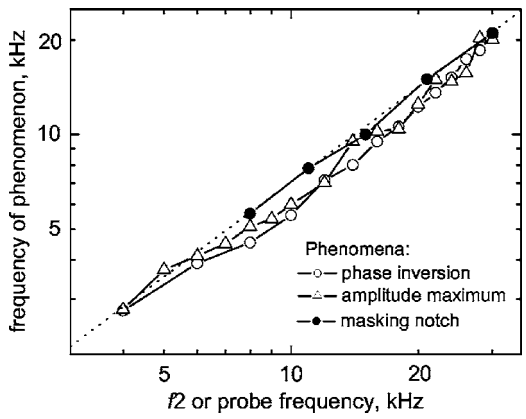


FIG. 6. Dependence of frequency of different manifestations of the hypothesized TM resonance either on the high-frequency primary (for DPOAE, open symbols) or on the probe tone frequency (for neural masking, filled symbols). Dotted line corresponds to frequencies situated half an octave below f_2.

Differences between the positions of the neural notch and the DPOAE maximum are especially prominent at f_2 above 10 kHz, where the position of the DPOAE maximum does not change gradually (Fig. 4) and the dependence of the DPOAE on the primary ratio has several local maxima (Fig. 5). Logarithmic scaling of these frequencies resembles the cochlear logarithmic scaling, which makes it unlikely that the local maxima are artifacts of the acoustic systems used for DPOAE stimulation and recording. The frequencies of the local maxima appear to be due to the intrinsic characteristics of the guinea pig cochlea because these frequencies do not depend on the primary frequencies. They are essentially the same for different f_2 and different DPOAE components (i.e., for different f_2/f_1 ratio) (Fig. 5). The frequency inde-

pendence of the local maxima on the primary frequencies also make it unlikely that these maxima are generated by multiple wave interference, which has been used to explain the frequency periodicity of the evoked emissions recorded from the same species in other experiments (Goodman *et al.*, 2003; Withnell *et al.*, 2003). These local maxima possibly reflect the formation of standing waves in the cochlea (Russell and Kössl, 1999; Shera, 2003b) which affects the DPOAE production either at the f_2 or at the DPOAE CF places.

The difference in frequency dependencies of the neural notch and the DPOAE maximum can be attributable to the complex nature of the DPOAE generation process. There is no consensus about the origin of the DPOAE bandpass dependence on the ratio of the primaries, and several mechanisms have been suggested to explain this dependence (Brown *et al.*, 1992; Allen and Fahey, 1993; Kanis and de Boer, 1997; Stover *et al.*, 1999; Lukashkin and Russell, 2001; Shera, 2003a; Vetešník and Nobili, 2003; de Boer *et al.*, 2005; Fahey *et al.*, 2006). There is direct and indirect evidence for the existence of several mechanisms in the cochlea, including the TM resonance, which will inevitably generate a bandpass DPOAE dependence on the ratio of the primary frequencies with appropriate choice of stimulus parameters. Therefore, we would like to suggest that the large number of hypotheses concerning the origin of the DPOAE bandpass structure has arisen, most likely, because of the simple fact that this structure is generated by a combination of mechanisms with different ones dominating at different combinations of stimulus parameters. The DPOAE bandpass structure recorded from nonmammalian species, where the auditory organs have different organizations, also points towards multiple mechanisms of the bandpass structure generation (Taschenberger *et al.*, 1995; Coro and Kössl, 1998; van Dijk and Manley, 2001).

V. CONCLUSION

The notch in the simultaneous masking neural tuning curves and the phase reversal of the DPOAE growth functions at the auditory thresholds are both situated about half an octave below the probe frequency and f_2, respectively, and show similar dependence on these frequencies. However, the DPOAE bandpass structure is likely to be generated by a combination of mechanisms with different ones dominating at different stimulus parameters.

ACKNOWLEDGMENTS

This work was supported by a grant from Medical Research Council. We thank James Hartley for technical assistance and Markus Drexl for helpful comments on the manuscript.

[1]Mutual suppression between primaries should be observed if their separation is small enough (Kanis and de Boer, 1997; Lukashkin and Russell, 2001), but because the DPOAE spectral components of different orders peak at about the same frequency of a half octave below the f_2, it was concluded that the cochlea bandpass filtered the DPOAE energy (Brown *et al.*, 1992; Allen and Fahey, 1993). This filter was associated with the TM which, it was proposed, is tuned to a frequency about a half octave below the CF of the BM in every cochlear location. In this model, DPOAE maximum is observed at the frequency of the TM resonance where relative displacement between the TM and the reticular lamina is minimal and mechanical energy is shunted away from the hair cells (Zwislocki, 1980; Allen, 1980). Allen and Fahey (1993) define a frequency f_Z in single unit neural tuning curves where the tip and tail of the tuning curves meet. Although Allen and Fahey (1993) found the same dependence of f_Z and DPOAE amplitude maximum on the unit's CF and the f_2, respectively, later publications showed that the concept of a "half octave below" filter for the DPOAE components of different orders could be an oversimplification of the experimental data (Stover *et al.*, 1999). Detailed investigation of the problem (Schneider *et al.*, 2001, 2003) reveals that there is a systematic variation of the DPOAE peak frequency depending on the order of the DPOAE. An alternative explanation (Stover *et al.*, 1999) of the DPOAE dependence on the ratio of the primaries utilizes the idea of multiple source generation for DPOAEs (Furst *et al.*, 1988; Whitehead *et al.*, 1992; Brown *et al.*, 1996; Stover *et al.*, 1996, Moulin, 2000; Fahey *et al.*, 2006). According to this hypothesis, the DPOAE emissions recorded in the ear canal would be the result of the vector summation of emission from the f_2 place and stimulus frequency emission at the DPOAE frequency. Variation in the amplitude and phase of the distortion product from these two sources can give rise to a complex pattern in the vector sum, and, for example, be the origin of its bandpass dependence on the ratio f_2/f_1. A diagonal canyon in the two-dimensional space of the ratio of primaries and their amplitudes, which is an intrinsic property of the DPOAE producing nonlinearity (Lukashkin and Russell, 1998, 1999), also contributes towards the bandpass appearance of the DPOAE amplitude when the ratio of the primaries changes (Lukashkin and Russell, 2001). More recently, it has been suggested that the decrease of the DPOAE when f_2/f_1 approaches unity is due to the wave-interference effect due to the emission generation from an extended cochlear region (Shera, 2003a; Vetešník and Nobili, 2003; de Boer *et al.*, 2005).

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. (**1993**). "A second cochlear frequency map that correlates distortion product and neural tuning measurement," J. Acoust. Soc. Am. **94**, 809–817.

Brown, A. M., and Kemp, D. T. (**1984**). "Suppressibility of the 2*f*1-*f*2 stimulated acoustic emissions in gerbil and man," Hear. Res. **13**, 29–37.

Brown, A. M., Gaskill, S. A., and Williams, D. M. (**1992**). "Mechanical filtering of sound in the inner ear," Proc. R. Soc. London, Ser. B **250**, 29–34.

Brown, A. M., Harris, F. P., and Beveridge, H. A. (**1996**). "Two sources of acoustic distortion products from the human cochlea," J. Acoust. Soc. Am. **100**, 3260–3267.

Cooper, N. P., and Rhode, W. S. (**1996**). "Fast travelling waves, slow travelling waves and their interactions in experimental studies of apical cochlear mechanics," Aud. Neurosci. **2**, 289–299.

Coro, F., and Kössl, M. (**1998**). "Distortion-product otoacoustic emissions from the tympanic organ in two noctuoid moths," J. Comp. Physiol., A **183**, 525–531.

Dallos, P., and Cheatham, M. A. (**1976**). "Compound action potential (AP) tuning curves," J. Acoust. Soc. Am. **59**, 591–597.

de Boer, E., Nuttall, A. L., Hu, N., Zou, Y., and Zheng, J. (**2005**). "The Allen-Fahey experiment extended," J. Acoust. Soc. Am. **117**, 1260–1266.

Fahey, P. F., Stagner, B. B., and Martin, G. K. (**2006**). "Mechanism for bandpass frequency characteristic in distortion product otoacoustic emission generation," J. Acoust. Soc. Am. **119**, 991–996.

Furst, M., Rabinowitz, W. M., and Zurek, P. M. (**1988**). "Ear canal acoustic distortion at $2f_1$-f_2 from human ears: Relation to other emissions and perceived combination tones," J. Acoust. Soc. Am. **84**, 215–221.

Goodman, S. S., Withnell, R. H., and Shera, C. A. (**2003**). "The origin of SFOAE microstructure in the guinea pig," Hear. Res. **183**, 7–17.

Gummer, A. W., Hemmert, W., and Zenner, H. P. (**1996**). "Resonant tectorial membrane motion in the inner ear: Its crucial role in frequency tuning," Proc. Natl. Acad. Sci. U.S.A. **93**, 8727–8732.

Harada, T., Ogawa, K., Inoue, Y., and Kanzaki, J. (**2001**). "Effect of changes in stimulus level on phases of distortion product otoacoustic emissions," Hear. Res. **152**, 152–158.

Hemmert, W., Zenner, H-P., and Gummer, A. W. (**2000**). "Three-dimensional motion of the organ of Corti," Biophys. J. **78**, 2285–2297.

Kanis, L. J., and de Boer, E. (**1997**). "Frequency dependence of acoustic distortion products in a locally active model of the cochlea," J. Acoust.

Soc. Am. **101**, 1527–1531.
Legan, P. K., Lukashkina, V. A., Goodyear, R. J., Kössl, M., Russell, I. J., and Richardson, G. P. (**2000**). "A targeted deletion in α-tectorin reveals that the tectorial membrane is required for the gain and timing of cochlear feedback," Neuron **28**, 273–285.
Legan, P. K., Lukashkina, V. A., Goodyear, R. J., Lukashkin, A. N., Verhoeven, K., Camp, G. V., Russell, I. J., and Richardson, G. P. (**2005**). "A deafness mutation isolates a second role for the tectorial membrane in hearing," Nat. Neurosci. **8**, 1035–1042.
Lukashkin, A. N., and Russell, I. J. (**1998**). "A descriptive model of the receptor potential nonlinearities generated by the hair cell mechanoelectrical transducer," J. Acoust. Soc. Am. **103**, 973–980.
Lukashkin, A. N., and Russell, I. J. (**1999**). "Analysis of the f_2-f_1 and $2f_1$-f_2 distortion components generated by the hair cell mechanoelectrical transducer: Dependence on the amplitudes of the primaries and feedback gain," J. Acoust. Soc. Am. **106**, 2661–2668.
Lukashkin, A. N., and Russell, I. J. (**2001**). "Origin of the bell-like dependence of the DPOAE amplitude on primary frequency ratio," J. Acoust. Soc. Am. **110**, 3097–3106.
Lukashkin, A. N., and Russell, I. J. (**2003**). "A second, low frequency mode of vibration in the intact mammalian cochlea," J. Acoust. Soc. Am. **113**, 1544–1550.
Moulin, A. (**2000**). "Influence of primary frequencies ratio on distortion product otoacoustic emissions amplitude. I. Intersubject variability and consequences on the DPOAE-gram," J. Acoust. Soc. Am. **107**, 1460–1470.
Nowotny, M., and Gummer, A. W. (**2006**). "Nanomechanics of the subtectorial space caused by electromechanics of cochlear outer hair cell," Proc. Natl. Acad. Sci. U.S.A. **103**, 2120–2125.
Russell, I. J., and Kössl, M. (**1999**). "Micromechanical responses to tones in the auditory fovea of the greater mustached bat's cochlea," J. Neurophysiol. **82**, 676–686.
Scherer, M. P., and Gummer, A. W. (**2004**). "Impedance analysis of the organ of Corti with magnetically actuated probes," Biophys. J. **87**, 1378–1391.
Schneider, S., Schoonhoven, R., and Prijs, V. F. (**2001**). "Amplitude of distortion product otoacoustic emissions in the guinea pig in $f1$- and $f2$-sweep paradigms," Hear. Res. **155**, 21–31.
Schneider, S., Prijs, V. F., and Schoonhoven, R. (**2003**). "Amplitude and phase of distortion product otoacoustic emissions in the guinea pig in an ($f1$,$f2$) area study," J. Acoust. Soc. Am. **113**, 3285–3296.
Shera, C. A. (**2003a**). "Wave interference in the generation of reflection- and distortion-source emissions," in *Biophysics of the Cochlea: From Molecules to Models*, edited by A. W. Gummer, E. Dalhoff, M. Nowotny, and M. P. Scherer (World Scientific, Singapore), pp. 439–453.
Shera, C. A. (**2003b**). "Mammalian spontaneous otoacoustic emissions are amplitude-stabilized cochlear standing waves," J. Acoust. Soc. Am. **114**, 244–262.
Stover, L. J., Neely, S. T., and Gorga, M. P. (**1996**). "Latency and multiple sources of distortion product otoacoustic emissions," J. Acoust. Soc. Am. **99**, 1016–1024.
Stover, L. J., Neely, S. T., and Gorga, M. P. (**1999**). "Cochlear generation of intermodulation distortion revealed by DPOAE frequency functions in normal and impaired ears," J. Acoust. Soc. Am. **106**, 2669–2678.
Taberner, A. M., and Liberman, M. C. (**2006**). "Response properties of single auditory nerve fibers in the mouse," J. Neurophysiol. **93**, 557–569.
Taschenberger, G., Gallo, L., and Manley, G. A. (**1995**). "Filtering of distortion-product otoacoustic emissions in the inner ear of birds and lizards," Hear. Res. **91**, 87–92.
van Dijk, P., and Manley, G. A. (**2001**). "Distortion product otoacoustic emissions in the tree frog *Hyla cinerea*," Hear. Res. **153**, 14–22.
Vetešník, A., and Nobili, R. (**2003**). "Phase behaviour of the primaries in distortion product analysis," in *Biophysics of the Cochlea: From Molecules to Models*, edited by A. W. Gummer, E. Dalhoff, M. Nowotny, and M. P. Scherer (World Scientific, Singapore), pp. 557–558.
Whitehead, M. L., Lonsbury-Martin, B. L., and Martin, G. K. (**1992**). "Evidence for two discrete sources of $2f1$-$f2$ distortion—product otoacoustic emission in rabbit. I. Differential dependence on stimulus parameters," J. Acoust. Soc. Am. **91**, 1587–1607.
Withnell, R. H., Shaffer, L. A., and Talmadge, C. L. (**2003**). "Generation of DPOAEs in the guinea pig," Hear. Res. **178**, 106–117.
Zwislocki, J. J. (**1980**). "Five decades of research on cochlear mechanics," J. Acoust. Soc. Am. **67**, 1679–1685.
Zwislocki, J. J. (**1984**). "How OHC lesions can lead to neural cochlear hypersensitivity," Acta Oto-Laryngol. **97**, 529–534.
Zwislocki, J. J. (**1986**). "Analysis of cochlear mechanics," Hear. Res. **22**, 155–169.
Zwislocki, J. J. (**2002**). *Auditory Sound Transmission: An Autobiographical Perspective* (Erlbaum, Mahwah, NJ).

Suppression of distortion product otoacoustic emissions in the anuran ear

Sebastiaan W. F. Meenderink
Department of Physiological Science, University of California, Los Angeles, 621 Charles E. Young Drive South, Los Angeles, California 90095-1606

Peter M. Narins[a)]
Departments of Physiological Science and Ecology & Evolutionary Biology, University of California, Los Angeles, 621 Charles E. Young Drive South, Los Angeles, California 90095-1606

(Received 4 July 2006; revised 3 October 2006; accepted 3 October 2006)

When a two-tone stimulus is presented to the ear, so-called distortion product otoacoustic emissions (DPOAEs) are evoked. Adding an interference tone (IT) to these two DPOAE-evoking primaries affects normal DPOAE generation. The "effectiveness" of interference depends on the frequency of the IT in relation to the primary frequencies and this provides clues about the locus of emission generation within the inner ear. Here results are presented on the effects of ITs on DPOAEs thought to originate from the basilar papilla (BP) of a frog species. It is found that the IT always resulted in a reduction of the recorded DPOAE amplitude: DPOAE enhancement was not observed. Furthermore, iso-suppression curves (ISCs) exhibited two relative minima suggesting that the DPOAEs arise at different loci in the inner ear. These minima occurred at fixed frequencies, which coincided with those primary frequencies that resulted in maxima in DPOAE audiograms. The occurrence of two minima suggests that DPOAEs, which are presumed to originate exclusively from the BP, partially arise from the amphibian papilla as well. Finally, the finding that the minima in the ISCs are independent of the primary or DPOAE frequencies provides support for the notion that the BP functions as a single auditory filter.
[DOI: 10.1121/1.2382458]

PACS number(s): 43.64.Jb, 43.64.Tk, 43.80.Lb [BLM] Pages: 344–351

I. INTRODUCTION

Upon stimulation with two appropriately chosen pure tones, the inner ear produces weak sounds at frequencies other than those present in the stimulus. These so-called distortion product otoacoustic emissions (DPOAEs) were first discovered in humans (Kemp, 1979), but were subsequently found in all other classes of tetrapods (i.e., birds (Taschenberger and Manley, 1998), reptiles (Manley *et al.*, 1993) and amphibians (Van Dijk and Manley, 2001). This wide distribution of DPOAEs (and otoacoustic emissions in general) among tetrapods suggests that otoacoustic emissions are the result of a fundamental property of the sensory hair cells/epithelium within the inner ear (Köppl and Manley, 1993). After the initial discovery of DPOAEs it was recognized that their amplitudes could be affected by the introduction of a third tone (henceforth termed "interference tone" or "IT"). If a specific frequency region of the auditory epithelium is important for DPOAE generation, then an IT at this frequency should be most effective in reducing the DPOAE level. As such, the effect of the IT is frequency dependent and the "effectiveness" of suppression can be used to infer the various loci of DPOAE generation along the sensory epithelium.

The frog inner ear is different from that of mammals in that it does not have a cochlea to detect airborne sound. Rather, there are two auditory end organs (called papillae) that are separated both anatomically as well as in their frequency ranges of highest sensitivity. Both papillae share (1) the absence of a basilar membrane, or analogous structure (hair cells are directly imbedded in the cartilaginous wall of the papillar recesses), (2) the absence of distinct groups of hair cells (i.e., no inner or outer hair cells) and (3) the presence of a tectorial membrane that covers the sensory hair cells and into which the hair bundles protrude. A detailed anatomical and physiological description of the anuran ear, including a discussion of the differences and similarities with the mammalian ear, can be found in Lewis and Narins (1999). The sensory epithelium of the amphibian papilla (AP), which is most sensitive to the low and mid frequencies of the frog's hearing range, consists of an elongated strip of hair cells that is organized in a tonotopic fashion (i.e., hair cells exhibit best frequencies that increase systematically along the sensory epithelium). Hair cells receive both afferent and efferent nerve fibers, with the latter having tuning characteristics similar to those found in mammals at similar characteristic frequencies (Lewis, 1992). Finally, the papilla is thought to generate both spontaneous and evoked (e.g., distortion product) OAEs (Van Dijk *et al.*, 1996; Van Dijk and Manley, 2001). The other hearing organ, the basilar papilla (BP), is much simpler in structure and function. It consists of a small epithelial patch containing relatively few hair cells [between 50 and 100 in *R. catesbeiana* (Lewis, 1978)]. In each individual, afferent nerve fibers innervating the BP

[a)]Author to whom correspondence should be addressed. Electronic mail: pnarins@ucla.edu

are all tuned to the same frequency (Ronken, 1990) while efferent innervation is absent [except in the frog species *X. laevis* (Hellman and Fritzsch, 1996)]. Presumably this papilla only generates evoked OAEs; spontaneous OAEs seem absent (Van Dijk *et al.*, 1996; Van Dijk and Manley, 2001). It is this relative simplicity in anatomy and (electro-) physiology which makes the frog basilar papilla an important model system for the study of auditory function. Essentially, the BP seems to behave as a single auditory filter with tuning presumably deriving from a passive, mechanical filtering mechanism.

These properties of the BP are reflected in a simple model for this hearing organ, which was used to qualitatively describe the dependence of DPOAE generation on stimulus tone frequencies (Van Dijk and Manley, 2001; Meenderink *et al.*, 2005a) and the relation between DPOAE- and neural tuning (Meenderink *et al.*, 2005b). The presence of only a single auditory filter fixes the location of emission generation by necessity; there is no sensory epithelium in the BP that is tuned to a different frequency to which the location of emission generation can shift with changing stimulus and/or DPOAE frequencies. Here, we focus on the effects of a third tone on the generation of DPOAEs thought to arise in the BP of the Northern leopard frog *Rana pipiens pipiens*. In light of the described properties of the BP it is expected that the effect of the IT is largest when its frequency coincides with the frequency that results in the largest emission level in the DPOAE audiogram. Moreover, given the fixed location of DPOAE generation, it is hypothesized that variation in the frequencies of the primaries should not affect the frequency of the IT that produces the largest effect on DPOAE amplitude.

II. MATERIALS AND METHODS

DPOAEs were recorded from the right ear of seven Northern leopard frogs, *Rana pipiens pipiens* (body weight: 21.2–42.4 g). Each animal was immobilized with an intramuscular injection of pentobarbital sodium solution in the left hind limb (Nembutal: effective dosage 55–60 mg/kg body weight) and after approximately 30 min was transferred to a sound-attenuating chamber. Here it was wrapped in soaked gauze to facilitate cutaneous respiration and placed in the recording setup on a vibration-isolation table.

To record DPOAEs, a small plastic tube was sealed to the skin surrounding the frog's tympanic membrane using high-vacuum silicone grease (Dow Corning Corp., Midland, MI). The tube contained an ER-10A microphone system (Etymotic Research, Elk Grove Village, IL) for sound pressure recording, the output of which was amplified by the built-in amplifier (amplification=20 dB), digitized (RP2: Tucker-Davis Technologies, Alachua, FL) and stored on computer disk for offline analysis. DPOAEs were evoked by two primaries (with frequencies f_1 and f_2, where $f_2>f_1$) that were delivered to the ear via separate tube phones (ER-2: Etymotic Research). These were connected to the microphone system using short plastic tubings. Each primary was generated from a separate digital/analog (D/A) channel (RP2: TDT) and attenuated (PA5: TDT) to set the desired level (L_1 and L_2, respectively).

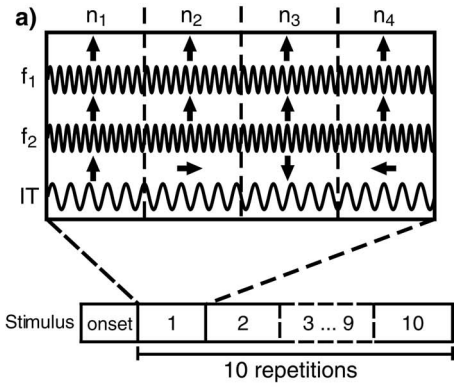
In order to add the interference tone to the stimulus, an additional speaker (DT48: Beyerdynamic, Heilbronn, Germany) was connected to the recording probe. This speaker was mounted in a conical brass housing that opened directly in the small plastic probe tube. Like the two primaries, the IT was generated from a separate D/A channel (RP2: TDT) and set to the desired level using a programmable attenuator (PA5: TDT). Customized software, written in MATLAB (The Mathworks, Natick, MA) and RPvds (TDT) were used to control the generation and the attenuation of the three tones as well as the recording of the microphone signal. Both D/A and A/D conversion was done at a frequency of 12.2 kHz.

All DPOAE recordings were made with equal-level primaries $L_1=L_2$, either at 81 dB sound pressure level (SPL) or 76 dB SPL. The frequency f_1 was chosen such that N sample points (where N is an integer) held exactly 40 periods of this tone and varied between 1323 and 2026 Hz across frogs. Depending on the ratio f_2/f_1 (either 1.1 or 1.2), this block of sample points also held an integer number of periods for f_2 (44 or 48 periods, respectively). As a result, the DPOAEs evoked by these two tones are also exactly periodic over the same number of sample points.

For each primary combination, the interference tone was presented at different frequencies (f_3) and levels (L_3) that made up a 40×18 matrix. For each f_3, the level of the interference tone was systematically decreased from 20 dB above L_1 to 12 dB below L_1 in 2 dB steps (17 presentations). After this, one additional recording was made in the absence of the IT. This last recording was used as a reference to calculate the amount of suppression at that particular f_3, while the variation of the unsuppressed DPOAE levels across f_3s was used to assess whether the frog had moved during the recordings.

The frequency of the interference tone was chosen such that its phase rotated 90° for consecutive blocks of N sample points (see Fig. 1(a)) and was systematically increased from low to high frequencies over a range and in steps that were relative to the frequency f_1. That is, f_3 is given by the fraction $40\times f_3/f_1=(m+1/4)$, where m is an integer ranging from 11 to 89 in steps of 2 (resulting in the 40 different f_3s for each f_1,f_2 combination). This particular choice for the f_3s ensures that the IT phase rotates in steps of 90° for consecutive blocks of N sample points, while at the same time results in a stimulus (consisting of the two primaries and the IT) that is exactly repeated every $4N$ sample points. The stimulus was presented continuously for $(N+10^{*}4N)$ sample points (which equals 0.8–1.2 s depending on f_1). The first N sample points allowed for the onset of the stimulus (and DPOAEs) and these were removed from the recorded microphone signal prior to analysis.

In the analysis (offline), the recorded signal was averaged in two steps. First, each block of $4N$ sample points was subjected to a level-crossing, artifact-rejection routine (i.e., for each block, the absolute maximum voltage was compared to the entire signal's mean voltage. When this extreme exceeded the mean by more than four standard deviations, the



b)

	n_1	n_2	n_3	n_4	present in $(n_1+n_2+n_3+n_4)/4$
f_1	0	0	0	0	yes
f_2	0	0	0	0	yes
f_3	0	90	180	270	no
$2f_3$-f_1	0	180	0	180	no
$3f_3$-$2f_1$	0	270	180	90	no
$4f_3$-$3f_1$	0	0	0	0	yes
$2f_1$-f_2	0	0	0	0	yes
$2f_1$-f_3	0	270	180	90	no
$3f_1$-$2f_3$	0	180	0	180	no
$2f_2$-f_1	0	0	0	0	yes

FIG. 1. (a) Stylized example of the three-tone stimulus used in the experiments. The stimulus consisted of the two primaries and the interference tone (IT) and was presented for $(N+10^{*}4N)$ sample points. The first N sample points in the stimulus allowed for the onset of the three tones (and DPOAEs); they were removed from the recorded signal prior to analysis. The frequency f_1 was chosen such that N sample points included 40 periods for this primary. For frequency f_2, the N sample points included either 44 or 48 periods depending on the ratio f_2/f_1 (either 1.1 or 1.2). As a consequence, after each N sample points (indicated by the arrows), the pattern of the two primaries repeats exactly. Note that for clarity, only a quarter of the periods for the primaries are drawn (10 and 11 periods, respectively). The frequency of the IT was chosen such that it completed an integer number of periods over $4N$ sample points, but its phase was rotated by 90° in each consecutive block of N sample points (illustrated by the 90° rotation of the consecutive arrows for this signal). This particular construct of the three-tone stimulus results in cancellation of the IT and several of the DPOAEs it evokes when the four blocks $n_{1\ldots4}$ (of N sample points) are averaged together. At the same time, neither the two primaries nor the DPOAEs they evoke are cancelled from the averaged signal. (b) Table that gives the starting phase (in degrees) of several frequency components in each of the blocks $n_{1\ldots4}$. Whether the component cancels in the described average is also given.

block was considered to contain an artifact and therefore was excluded from the analysis). Artifact-free blocks were subaveraged in two buffers A and B. The mean of the two buffers, (A+B)/2, can be used to estimate the DPOAE amplitudes, while their difference, A—B, provides an estimate for the noise levels at the DPOAE frequencies. This first averaging step results in two signals, each with a length of $4N$ sample points. Next, each of these signals was further averaged as $(n_1+n_2+n_3+n_4)/4$, where $n_{1\ldots4}$ indicate the four consecutive blocks of N sample points within the signal. The periodicity of the two primaries, as well as the DPOAEs they evoke, over N sample points, ensures that their levels are unaffected by this averaging step. Since the IT phase rotates 90° for the consecutive blocks $n_{1\ldots4}$, this tone is canceled from the signal when these blocks are added together. In addition, the majority of DPOAE components (with the exception of terms with integer multiples of $4f_3$) that may arise from the interaction between the IT and either one of the primaries are also cancelled in the final average due to the particular phase variation of the interference tone (see Fig. 1(b)). Potentially, the introduction of a third tone introduces energy at frequencies that coincide with the DPOAE frequencies ($2f_1$-f_2 and $2f_2$-f_1) and thus may bias the estimate of the IT effect on DPOAE generation. With the recording method used here, such spurious loading of "unwanted" frequency bins is much reduced. Finally, the levels of the $2f_1$-f_2 and $2f_2$-f_1 DPOAEs and the related noise floors were calculated from the final averages using Fourier analysis.

Throughout this study, the "Principles of Animal Care" (NIH-publication 85-23, revised 1985) and USA regulations were followed; all protocols were approved by the University of California Animal Research Committee.

III. RESULTS

The matrix of (f_3, L_3)-combinations presented for a fixed primary combination was used to calculate iso-suppression contourlines (ISCs). These are curves that show the level of the interference tone (L_3) necessary to decrease the DPOAE level by a given criterion amount as a function of interference tone frequency (f_3). Representative sets of ISCs for $2f_1$-f_2 and $2f_2$-f_1 are given in Fig. 2. These data were obtained in one frog with f_1=1725 Hz, f_2=1896 Hz (f_2/f_1 =1.1) and $L_1=L_2$=81 dB SPL. This figure illustrates several of the properties that were generally observed in our experiments. First, if the IT produced any consistent effect, it was always a reduction of the DPOAE amplitude (i.e., DPOAE suppression). For none of the (f_3, L_3)-combinations tested did the DPOAE levels increase with the introduction of the IT; DPOAE enhancement was not observed. Second, for a fixed f_3, increasing L_3 invariably resulted in a larger reduction of the DPOAE level. The deviation from this seen in Fig. 2(a) (when $f_3 \approx f_1$ and $L_3 > L_1$) is most likely caused by unwanted energy that is added to the DPOAE-frequency bin (directly from the IT or from higher-order distortion products that arise via interaction between the IT and f_1) despite the recording method used. As such, this small (f_3, L_3)-area with relatively little suppression is an artifact and does not reflect specific properties of $2f_1$-f_2 generation. Third, for a fixed L_3 the amount of DPOAE suppression varied in a frequency-dependent manner. This frequency dependence is also reflected by the ISCs, which are more or less W shaped. In the example shown, the absolute minimum of the different contourlines coincides in frequency with f_1 for both distortion products, while a secondary minimum is present at a somewhat lower IT frequency ($f_3 \approx 1$ kHz) that did not coincide with either the primary frequencies or the DPOAE fre-

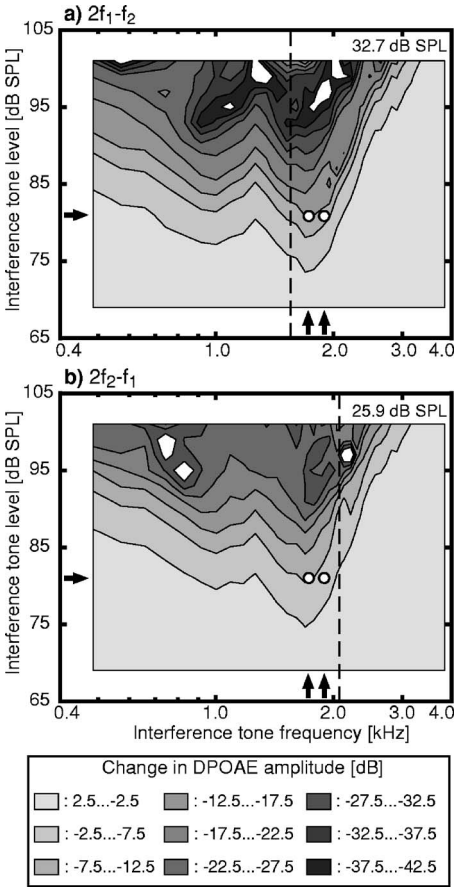


FIG. 2. Iso-suppression contourplots illustrating the effect of the IT on the amplitudes of the DPOAE at (a) $2f_1$-f_2 and (b) $2f_2$-f_1 that were generated by a fixed-primary combination (f_1=1725 Hz, f_2/f_1=1.1, L_1=L_2=81 dB SPL) in one frog. In each panel, the contour lines are drawn at 5 dB intervals, with the appropriate changes in DPOAE level (relative to the unsuppressed DPOAE level) given in the key. The arrows on the abscissa and ordinate, as well as the white circles within the plot, give the frequencies and levels of the two primary tones. The vertical dashed line indicates the DPOAE frequency, while the number in the upper-right corner is the mean unsuppressed DPOAE level.

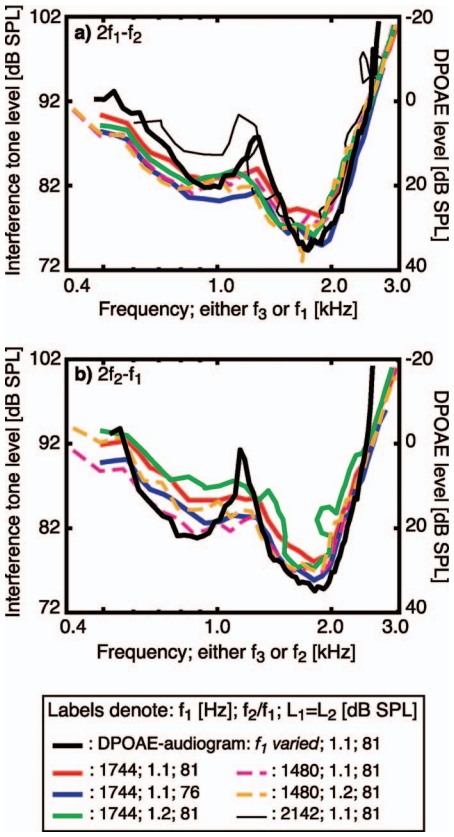


FIG. 3. Several iso-suppression contour lines (ISCs) at 6 dB suppression for DPOAEs at (a) $2f_1$-f_2 and (b) $2f_2$-f_1 obtained in one frog. In each panel, each contour line was obtained with a different parameter combination for the primaries, as given in the key. The shape and the position of the contour lines within the (f_3, L_3) area is independent of these parameters; all contour lines are more or less superimposed. For f_1=2142 Hz, only the DPOAE at $2f_1$-f_2 was of high enough level to calculate the corresponding ISC; the line is therefore absent in panel (b). Each panel also shows the corresponding DPOAE audiogram (i.e., DPOAE amplitude as a function of primary frequency for fixed primary levels) recorded in the same animal. To facilitate graphical comparison, the DPOAE audiograms are plotted with a reversed ordinate (the right y axes give DPOAE level). Notice that the range of the ordinate for DPOAE levels is twice that of the IT-level range (left y axes) and that for $2f_1$-f_2 the DPOAE audiogram is plotted vs f_1, while for $2f_2$-f_1 it is plotted as a function of f_2. In effect, the IT resulted in largest DPOAE suppression when it coincided in frequency with the primary frequency (f_1 for $2f_1$-f_2 and f_2 for $2f_2$-f_1) that corresponded to the maximum in the DPOAE audiogram. The "tip" of the ISCs did not relate to either one of the primary frequencies or the DPOAE frequency.

quency. In all cases, the absolute minimum occurred for a higher IT frequency than the secondary minimum. For clarity, we will refer to these two minima in the ISCs as high-frequency minimum (HFM) and low-frequency minimum (LFM), respectively.

The observation from Fig. 2 that the HFM occurs when f_3=f_1 is not generally true. This is illustrated in Fig. 3, which shows several ISCs at 6 dB suppression, all obtained in one frog. Here, each curve was obtained for a different combination of primary frequencies. It can be seen that the position of these curves within the (f_3, L_3)-area did not depend on the primary-tone parameters. Each panel in Fig. 3 also shows a DPOAE audiogram (i.e., DPOAE amplitude, obtained by sweeping the primary frequency f_1 while keeping f_2/f_1 and L_1=L_2 constant, plotted as a function of frequency) that was recorded from the same ear. To facilitate graphical comparison with the ISCs, these audiograms have been plotted with a reversed ordinate. As typically found for anuran DPOAE audiograms, it exhibits two relative maxima which are thought to reflect emission generation in the AP and the BP (e.g., Van Dijk and Manley, 2001; Meenderink *et al.*, 2005a). Clearly, the ISCs and the reversed DPOAE audiogram largely overlap. However, notice that the vertical (level) range for the DPOAE audiograms is twice the range used for the ISCs.

The similarity between the different ISCs indicates that for each f_3 the L_3 required to achieve the criterion amount of suppression did not change with varying primary and DPOAE frequencies or levels. In addition, the observation that the two minima in the different ISCs and the two maxima in the DPOAE audiogram occurred at similar frequencies suggests that the origin/loci of the suppressed DPOAEs is closely associated with the primary frequencies that resulted in these latter maxima. Apparently, the loci of DPOAE generation did not vary with varying primary frequencies. Rather, they seemed to be at a fixed location along

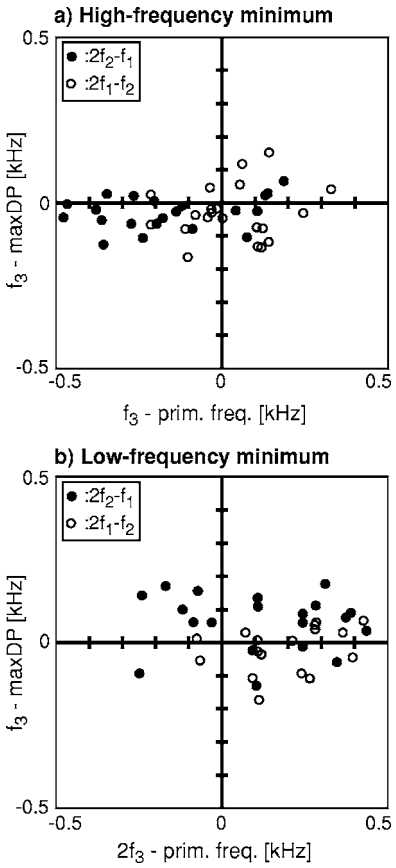


FIG. 4. For each (f_3, L_3) matrix we determined the IT frequency at the low-frequency and high-frequency minimum in the ISCs (averaged over ISCs at −2, −4, −6, −8, and −10 dB suppression) for $2f_1$-f_2 and $2f_2$-f_1. In (a) these frequencies at the high-frequency minima were compared to the primary frequencies (f_1 for $2f_1$-f_2 and f_2 for $2f_2$-f_1) used to evoke the suppressed emissions and to the primary frequencies (again f_1 for $2f_1$-f_2 and f_2 for $2f_2$-f_1) that resulted in the high-frequency maxima in the corresponding DPOAE audiograms. In (b), a similar comparison is made for the low-frequency minima, but here the primary frequencies that were used to evoke the suppressed emissions were compared to the first harmonic of the IT frequency. This was done to assess whether the low-frequency minimum reflected suppression of the DPOAE by this harmonic and as such constituted an artifact rather than revealing a second DPOAE source.

the inner ear's sensory epithelium and their locations were at those primary frequencies that resulted in the two relative maxima in the DPOAE audiogram.

To further illustrate these observations, the f_3 at HFM and LFM (taken as the average for ISCs at −2, −4, −6, −8, and −10 dB) were determined in each (f_3, L_3)-matrix. In addition, we obtained the primary frequencies (f_1s for $2f_1$-f_2 and f_2s for $2f_2$-f_1) that resulted in the two relative maxima in the corresponding DPOAE audiograms (i.e., same frog, $L_1 = L_2$ and f_2/f_1). In Fig. 4(a), a comparison is made between the primary frequencies (relative to the f_3s at the HFM) used to evoke the suppressed DPOAEs and those corresponding to the high-frequency maximum in the DPOAE audiograms. The data in this plot are scattered around the abscissa, indicating that maximum suppression occurred when the IT was close in frequency to the primary frequency that resulted in the high-frequency maximum in the DPOAE audiogram. At the same time, the absence of a clear correlation between the data and the ordinate indicates that the former frequency is not dependent on the primary frequency that evoked the suppressed emission.

A similar plot is shown for the LFM in Fig. 4(b). However, in this plot the primary frequency that evoked the DPOAE is compared to the first harmonic frequency of the IT, rather than f_3 itself. This was done to assess whether the LFM reflected suppression of the DPOAE by this harmonic and as such constituted an artifact rather than revealing a second DPOAE source. As in Fig. 4(a), the data indicate that the LFM occurs when f_3 was similar to the primary frequency that resulted in the low-frequency maximum in the DPOAE audiogram. Here, the absence of clear correlation between the data and the ordinate indicates that suppression is not induced by the first harmonic of the IT around the primary frequency used to evoke the DPOAE. This latter is further supported by the notion that the level of the first harmonic of the IT is at least 40 dB below the IT level and as such is too low to induce significant suppression at any frequency.

IV. DISCUSSION

To evoke DPOAEs, a stimulus consisting of two primaries is presented to the ear. When an appropriately chosen third tone is added to this stimulus, the primary-evoked DPOAEs are affected. Several terms are in use to denote this third tone. Initially, it was called a "suppressor tone" since it was believed that such a tone reduces/attenuates the amplitude of the generated emission (Brown and Kemp, 1984). However, with the finding that the amplitude of the recorded DPOAE can also increase with the introduction of a third tone, the more general term "interference tone" (IT), which can indicate both suppression and enhancement, was suggested (Martin *et al.*, 2003). For consistency, this second term has been used throughout this manuscript, although no enhancement was observed for anuran BP-DPOAEs. It should be noted that the term "bias tone" is sometimes used to indicate the third tone (Bian *et al.*, 2002; 2004). In these cases, the interest is not in the "averaged" effect of the third tone on the DPOAE. Rather, the third tone is presented at a relatively low frequency and its effect on the time course of the generated DPOAE is mapped as a function of the bias-tone phase. The idea is that the third tone biases the operating point of the nonlinearity underlying DPOAE generation, thus resulting in DPOAEs that are different from the one being evoked with the nonlinearities operating point "at rest." Since the operating point is systematically dependent on the phase of the bias tone, it is possible to derive the nonlinearity involved in the DPOAE generation from these recordings.

The effect of an IT on the generation of DPOAEs has been documented for various classes of tetrapods (Köppl and Manley, 1993; Taschenberger and Manley, 1998; Frank and Kössl, 1995; Martin *et al.*, 1999), including humans (Brown and Kemp, 1984). The results from these experiments have been used to assess the characteristics of mechanical tuning and to derive the different locations of DPOAE generation. In mammals, iso-suppression contour lines (ISCs) exhibit a

minimum around f_2, indicating that this frequency region in the cochlea is important in lower-sideband[1] DPOAE generation. Further support for the origin of DPOAEs from this location comes from cochlear-model considerations (Talmadge *et al.*, 1998) and the very small DPOAE group delays that are observed for fixed-f_2 recording paradigms (Schneider *et al.*, 2003). Similarly, a second source for these emissions has been identified at the tonotopic location of the DPOAE frequency. Moreover, another potential cochlear location for DPOAE generation has been identified using interference tones (Martin *et al.*, 1999). This source approximately coincides with the tonotopic location of the first harmonic of f_2 and as such is considerably higher in frequency than either primary or the generated DPOAE.

In the frog BP, the origin of DPOAEs is expected to be less complex. The presence of several DPOAE sources that are at different locations *and* at different characteristic frequencies is unlikely, given the simple anatomy of the papilla and the observation that it functions as a single auditory filter. More precisely, we hypothesized that the unique properties of the BP result in maximum DPOAE suppression at a fixed IT frequency. This frequency is "preordained" by the mechanical tuning properties of the papilla and is given by the primary frequency that resulted in a maximum in the DPOAE audiogram. As such, it is independent of the absolute (f_1,f_2) and the relative (f_2/f_1) primary frequencies that are used to evoke the emissions. As illustrated in Figs. 3 and 4(a), these predictions are indeed observed, thus providing additional evidence for the unique tuning properties of the basilar papilla. However, the ISCs exhibit a secondary minimum at those frequencies that coincide with the low-frequency maximum in the corresponding DPOAE audiogram (Figs. 3 and 4(b)).

Anuran DPOAE audiograms generally have a bimodal shape. That is, they consist of two frequency ranges for which emission levels are relatively high. A more or less distinct frequency range (the extent of which seems species dependent) exhibiting reduced emission levels separates the former two. Across species, it has been found that the low-frequency DPOAE range closely resembles the frequency range for which the AP has the highest neural sensitivity (Van Dijk and Manley, 2001; Vassilakis *et al.*, 2005). A similar correspondence is observed for the high-frequency DPOAE-range and the highest neural sensitivity in the BP. Based on this, it has been hypothesized that the low-frequency DPOAEs arise from the AP, while high-frequency DPOAEs are generated in the BP. In addition, the findings that DPOAEs from the low-frequency and high-frequency range exhibit several differences (in the growth rate of input/output curves [Meenderink and Van Dijk, 2004], in their responses to anoxic conditions [Van Dijk *et al.*, 2003] and in their dependence on body temperature [Meenderink and Van Dijk, 2006]) further favors the idea that separate anatomical/functional structures (i.e., the AP and the BP) are involved in their generation. Based on these observations it has been assumed that DPOAEs arise from one of the two papillae, and that the primary frequencies determine whether they arise from the AP ($f<1250$ Hz in *Rana pipiens pipiens*) or the BP ($f>1250$ Hz); the nomenclature AP-DPOAEs and BP-DPOAEs have been employed to indicate these two conditions.

In the suppression experiments presented here, only primary frequencies that were thought to evoke BP-DPOAEs were used. Given the postulated strict separation between the generation of AP-DPOAEs and BP-DPOAEs, it is to be expected that ISCs exhibit only a single (high-frequency) minimum. However, we found that ISCs exhibited a low-frequency minimum as well. As shown in Fig. 4(b), this minimum is closely associated with DPOAEs that apparently arise from the AP. It seems that a combination of primary tones that, so far, was thought to exclusively generate DPOAEs from the BP also evokes emissions from the AP and that these two components combine to give the recorded DPOAE. The contribution of these two components is further illustrated in Fig. 5. These data were obtained in a frog that exhibited an unusual DPOAE audiogram in that the low-frequency DPOAEs were relatively weak (Figs. 5(c) and 5(d)). That is, the DPOAE levels at the low-frequency maxima were 25.3 dB ($2f_1$-f_2) and 21.2 dB ($2f_2$-f_1) below those of the high-frequency maxima. In similar DPOAE grams (same $L_1=L_2$ and f_2/f_1) recorded from other frogs, these differences were only 6.9 dB (±4.5 dB, standard deviation) and 7.2 dB (±3.5 dB), respectively. In conjunction with this, ISCs did not exhibit a clear low-frequency minimum (Figs. 5(a) and 5(b)), indicating the absence of significant low-frequency components in the recorded DPOAEs.

To date, the terms AP-DPOAEs and BP-DPOAEs have generally been used to indicate emission generation originating solely from the AP and the BP, respectively. At least for "BP-DPOAEs" we find that there is a considerable component that arises from the AP, thus lowering the utility of this term.

The finding by Martin *et al.* (1999; 2003) that in mammals significant suppression and/or enhancement can occur for ITs at one octave above f_2 motivated us to extend the IT frequency range accordingly. As may be expected based on the known anuran inner ear physiology, we did not find any contribution to the generated DPOAEs from such high frequencies. As far as is known, the frog inner ear holds no sensory epithelium that is tuned to higher frequencies than those transduced in the BP[2] and that may generate these high-frequency DPOAE components. It has been suggested that the high-frequency DPOAE component may be the basis of the notches that are commonly observed in mammalian DPOAE input/output (I/O) curves, this in contrast to the idea that these notches result from interference between a low-level and a high-level component.[3] Anuran I/O curves also may exhibit notches, while high-frequency suppression and/or enhancement is absent. It is therefore unlikely that the proposed mechanism of interference between DPOAE components from different tonotopic locations gives rise to the notches in anuran I/O curves.

In contrast to our findings in the frog, the introduction of an IT may result in increased DPOAE amplitudes in other tetrapods. DPOAE enhancement seems to contradict with the idea that the IT reduces the amplitude of the DPOAE generated by the sensory epithelium at the IT's tonotopic location. However, such enhancement can occur when the DPOAE

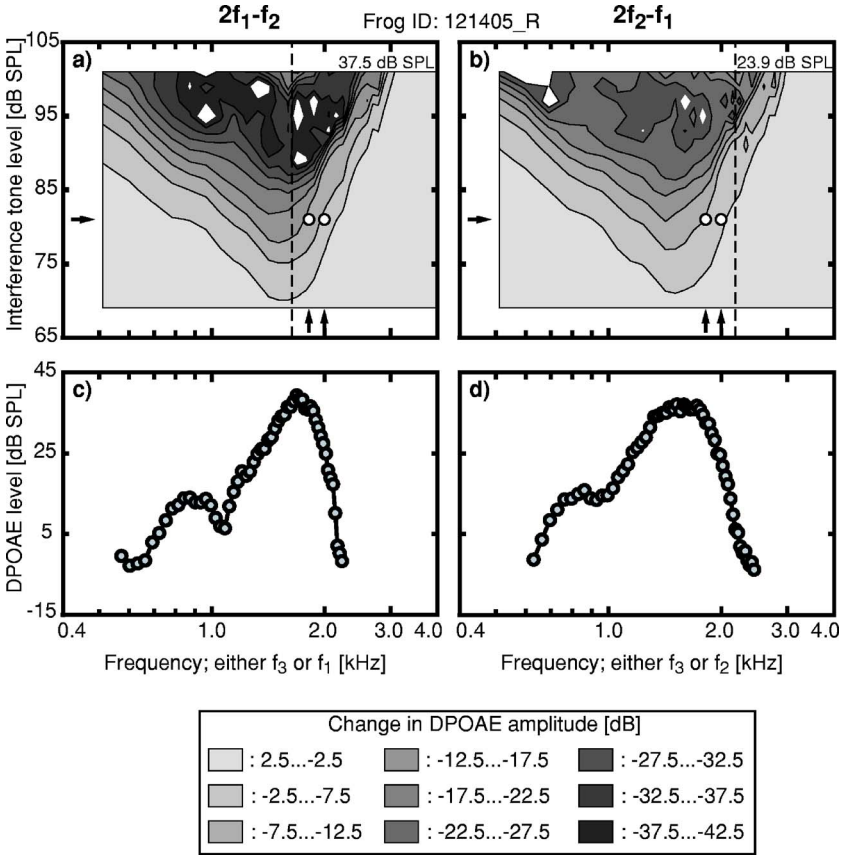


FIG. 5. Iso-suppression contour plots obtained in one frog for (a) $2f_1$-f_2 and (b) $2f_2$-f_1. The layout of these panels is the same as in Fig. 3. That is, the contour lines are drawn at 5 dB intervals with the corresponding levels given in the key. The parameters of the primaries are indicated in each panel by the arrows on the axes and the filled circles (f_1=1815 Hz; f_2/f_1=1.1; L_1=L_2=81 dB SPL), the vertical dashed line indicates the DPOAE frequency and the unsuppressed DPOAE amplitude is given in the upper right corner. The iso-suppression contour lines obtained in this frog are unique in that they are V-shaped rather than W-shaped; the second minimum at lower frequencies is absent. It seems that there is no contribution from the amphibian papilla to the recorded DPOAEs. This notion is supported by the shape of the DPOAE audiograms obtained in this frog [panels (c) and (d) respectively], which show an absence of strong AP-DPOAEs. Apparently the amphibian papilla in this frog did not generate "normal" DPOAEs, resulting in both the deviant shape of the DPOAE audiograms and the ISCs.

recorded in the ear canal is the result of two (or more) DPOAE components that are generated at different locations along the sensory epithelium. Then, the IT may reduce only one of these components while the other component(s) are unaffected. If the various DPOAE components interfere destructively (this depends on the relative phase between and the relative amplitude of the various DPOAE components), attenuation of one component can result in an increase in amplitude of the DPOAE that is recorded in the ear canal. The finding that in the frog BP no DPOAE enhancement occurs could thus indicate two possibilities: the recorded DPOAE (1) consists of only one component or (2) it consists of two or more components, but these do not interfere destructively with each other. As argued above, the recorded anuran DPOAEs seem to be the composite of two components. Apparently the phase relation between these components is always (at least at these high primary levels) such that they interfere constructively. It has been established that DPOAE group delays increase with decreasing primary levels (Meenderink and Van Dijk, 2006); perhaps the phase difference between the proposed AP- and the BP-component can be such that DPOAE enhancement occurs at lower primary levels, but this was not explored in the present study.

The present study focused on the effect of interference tones on high-frequency DPOAEs (BP-DPOAEs) only; their effect on low-frequency DPOAEs (AP-DPOAEs) was not investigated. This was primarily due to the limitations of our setup, which did not allow broadcasting of the high-level, low-frequency ITs necessary to suppress these latter DPOAEs. The presented suppression characteristics for BP-DPOAEs deviated on several accounts from those observed in other tetrapods. Instead of a frequency-dependent location of DPOAE generation that shifts along the epithelial tonotopic axis with the primaries, the origin of anuran BP-DPOAEs was at a fixed locus that seemed to be "preordained" by the mechanical tuning properties of the papilla. These findings are in agreement with the notion that the basilar papilla functions as a single auditory filter. Notwithstanding, an additional low-frequency component is present in the recorded BP-DPOAEs. The correlation of this minimum with the low-frequency maximum in the corresponding DPOAE audiograms favors the idea that this component originates from the low-frequency amphibian papilla, thus requiring a revision of the meaning of the term BP-DPOAEs.

ACKNOWLEDGMENTS

This work was supported by NIH grant No. DC-00222 to P.M.N.

[1]Lower sideband DPOAEs are emissions that are lower in frequency than both primary frequencies (e.g., $2f_1$-f_2, $3f_1$-$2f_2$).

[2]Recently, the transduction of ultrasounds has been shown to occur in a ranid frog species, but at present it is not known which inner ear end organ is involved (Feng *et al.*, 2006).

[3]Notches in DPOAE I/O curves may also be explained without the assumption of interference between two components. See, for instance, Lukashkin *et al.* (2002).

Bian, L., Chertoff, M. E., and Miller, E. (**2002**). "Deriving a cochlear transducer function from low-frequency modulation of distortion product otoacoustic emissions," J. Acoust. Soc. Am. **112**, 198–210.

Bian, L., Linhardt, E. E., and Chertoff, M. E. (**2004**). "Cochlear hysteresis: Observation with low-frequency modulated distortion product otoacoustic emissions," J. Acoust. Soc. Am. **115**, 2159–2172.

Brown, A. M., and Kemp, D. T. (**1984**). "Suppressibility of the $2f_1$-f_2 distortion product otoacoustic emission in gerbil and man," Hear. Res. **13**, 29–37.

Feng, A. S., Narins, P. M., Xu, C. H. *et al.* (**2006**). "Ultrasonic communication in frogs," Nature (London) **440**, 333–336.

Frank, G., and Kössl, M. (**1995**). "The shape of $2f_1$-f_2 suppression tuning curves reflects basilar membrane specialization in the mustached bat, *Pteronotus parnellii*," Hear. Res. **83**, 151–160.

Hellman, B., and Fritzsch, B. (**1996**). "Neuroanatomical and histochemical evidence for the presence of common lateral line and inner ear efferents and of efferents to the basilar papilla in a frog, *Xenopus laevis*," Brain Behav. Evol. **47**, 185–194.

Kemp, D. T. (**1979**). "Evidence of mechanical nonlinearity and frequency selective wave amplification in the cochlea," Arch. Oto-Rhino-Laryngol. **224**, 37–45.

Köppl, C., and Manley, G. A. (**1993**). "Distortion-product otoacoustic emissions in the bobtail lizard. II. Suppression tuning characteristics," J. Acoust. Soc. Am. **93**, 2834–2844.

Lewis, E. R. (**1978**). "Comparative studies of the anuran auditory papillae," Scanning Microsc. **II**, 633–642.

Lewis, E. R. (**1992**). "Convergence of design in vertebrate acoustic sensors," in *The Evolutionary Biology of Hearing*, edited by D. B. Webster, R. R. Fay, and A. N. Popper (Springer-Verlag, New York), pp. 163–184.

Lewis, E. R., and Narins, P. M. (**1999**). "The acoustic periphery of amphibians," in *Comparative Hearing: Fish and Amphibians*, edited by R. R. Fay and A. N. Popper (Springer-Verlag, New York), pp. 101–154.

Lukashkin, A. N., Lukashkina, V. A., and Russell, I. J. (**2002**). "One source for distortion product otoacoustic emissions generated by low- and high-level primaries," J. Acoust. Soc. Am. **111**, 2740–2748.

Manley, G. A., Köppl, C., and Johnstone, B. M. (**1993**). "Distortion-product otoacoustic emissions in the bobtail lizard. I. General characteristics," J. Acoust. Soc. Am. **93**, 2820–2833.

Martin, G. K., Stagner, B. B., Jassir, D., Telischi, F. F., and Lonsbury-Martin, B. L. (**1999**). "Suppression and enhancement of distortion product otoacoustic emissions by interference tones above f_2. I. Basic findings in rabbits," J. Acoust. Soc. Am. **136**, 105–123.

Martin, G. K., Villasuso, E. I., Stagner, B. B., and Lonsbury-Martin, B. L. (**2003**). "Suppression and enhancement of distortion product otoacoustic emissions by interference tones above f_2. II. Findings in humans," Hear. Res. **177**, 1111–1122.

Meenderink, S. W. F., Narins, P. M., and Van Dijk, P. (**2005a**). "Detailed f_1,f_2 area study of distortion product otoacoustic emissions in the frog," J. Assoc. Res. Otolaryngol. **6**, 37–47.

Meenderink, S. W. F., and Van Dijk, P. (**2004**). "Level dependence of distortion product otoacoustic emissions in the leopard frog, *Rana pipiens pipiens*," Hear. Res. **192**, 107–118.

Meenderink, S. W. F., and Van Dijk, P. (**2006**). "Temperature dependence of anuran distortion product otoacoustic emissions," J. Assoc. Res. Otolaryngol. **7**, 246–252.

Meenderink, S. W. F., Van Dijk, P., and Narins, P. M. (**2005b**). "Comparison between distortion product otoacoustic emissions and nerve fiber responses from the basilar papilla of the frog," J. Acoust. Soc. Am. **117**, 3165–3173.

Ronken, D. A. (**1990**). "Basic properties of auditory-nerve responses from a 'simple' ear: The basilar papilla of the frog," Hear. Res. **47**, 63–82.

Schneider, S., Prijs, V. F., and Schoonhoven, R. (**2003**). "Amplitude and phase of distortion product otoacoustic emissions in the guinea pig in an (f_1,f_2) area study," J. Acoust. Soc. Am. **113**, 3285–3296.

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.

Taschenberger, G., and Manley, G. A. (**1998**). "General characteristics and suppression tuning properties of the distortion-product otoacoustic emission $2f_1$-f_2 in the barn owl," Hear. Res. **123**, 183–200.

Van Dijk, P., and Manley, G. A. (**2001**). "Distortion product otoacoustic emissions in the tree frog, *Hyla cinerea*," Hear. Res. **153**, 14–22.

Van Dijk, P., Narins, P. M., and Mason, M. J. (**2003**). "Physiological vulnerability of distortion product otoacoustic emissions from the amphibian ear," J. Acoust. Soc. Am. **114**, 2044–2048.

Van Dijk, P., Narins, P. M., and Wang, J. (**1996**). "Spontaneous otoacoustic emissions in seven frog species," Hear. Res. **101**, 102–112.

Vassilakis, P. N., Meenderink, S. W. F., and Narins, P. M. (**2005**). "Distortion product otoacoustic emissions provide clues to hearing mechanisms in the frog ear," J. Acoust. Soc. Am. **116**, 3713–3726.

Wave propagation patterns in a "classical" three-dimensional model of the cochlea

Egbert de Boer
Room D2-226, Academic Medical Center, University of Amsterdam, Meibergdreef 9, 1105 AZ, Amsterdam, The Netherlands

Alfred L. Nuttall
Oregon Hearing Research Center, NRC04, Oregon Health & Science University, 3181 SW Sam Jackson Park Road, Portland, Oregon 97239-3098, and Kresge Hearing Research Institute, University of Michigan, 1301 E. Ann Street, Ann Arbor, Michigan 48109-0506

Christopher A. Shera
Eaton-Peabody Laboratory, Massachusetts Eye and Ear Infirmary, 243 Charles Street, Boston Massachusetts 02114

(Received 29 June 2006; revised 6 October 2006; accepted 11 October 2006)

The generation mechanisms of cochlear waves, in particular those that give rise to otoacoustic emissions (OAEs), are often complex. This makes it difficult to analyze wave propagation. In this paper two unusual excitation methods are applied to a three-dimensional stylized classical nonlinear model of the cochlea. The model used is constructed on the basis of data from an experimental animal selected to yield a smooth basilar-membrane impedance function. Waves going in two directions can be elicited by exciting the model locally instead of via the stapes. Production of DPOAEs was simulated by presenting the model with two relatively strong primary tones, with frequencies f_1 and f_2, estimating the driving pressure for the distortion product (DP) with frequency $2f_1-f_2$, and computing the resulting DP response pattern – as a function of distance along the basilar membrane. For wide as well as narrow frequency separations the resulting DP wave pattern in the model invariably showed that a reverse wave is dominant in nearly the entire region from the peak of the f_2-tone to the stapes. The computed DP wave pattern was further analyzed as to its constituent components with the aim to isolate their properties. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2385068]

PACS number(s): 43.64.Kc, 43.64.Bt [WPS] Pages: 352–362

I. INTRODUCTION

Experimental data on the responses of animal cochleae to complex stimuli often present a bewildering picture. One common example is the situation where two pure tones are presented simultaneously, and the response of the cochlea is investigated as to the properties of distortion products (DPs). The response can be observed in the form of oscillations of the basilar membrane (BM), in the form of the acoustic pressure or as otoacoustic emissions (for instance, DPOAEs), and analyzed in the time or the frequency domain. Depending on the experimental technique employed, and the parameters of the stimuli, a number of different phenomena can be discerned in these responses: frequency selectivity, distortion effects, multiple reflections, linear and nonlinear wave interference, effects of wave travel (dispersion), suppression, compression, and so on.

Properties associated with cochlear reflections have been combined into the theory of "coherent reflection" (Zweig and Shera, 1995). In two previous papers (Nuttall *et al.*, 2004; de Boer and Nuttall, 2006) we have described how data obtained from an animal with a spontaneous oscillation of the BM and a concomitant otoacoustic emission, can be analyzed and interpreted to show that coherent reflection constitutes a good explanation of the spectral irregularities experimentally observed. The second report is a good example of the use of a model of the cochlea as a device for obtaining deeper insight. It should be stressed at this point that the particular model was constructed directly *from the measured response of a particular animal*. In the past we have often utilized this method. For that special purpose the experimental data were acquired by presenting wide-band periodic noise signals to the animal, recording the waveforms of the responses, and analyzing those waveforms with cross-correlation functions (de Boer and Nuttall, 1997, 1999, 2000a, b). The underlying theory, the EQ-NL theorem, was described by the first author in 1997.

A brief description of this theorem follows. The cochlea is modeled as a nonlinear system of which the elements are smoothly distributed along the length of the basilar membrane (BM). For excitation with a noise stimulus, the frequency components of the signals inside the model are so close together that distortion products (combination tones) are swamped, so that only overall nonlinearity (in particular, compression) remains. The nonlinear model is compared to a linear one, in which the elements are given parameters that are modified by compression. It is shown that, with the proper choice of the latter elements and for the same stimulus signal used, the input-output cross-correlation functions are the same for the two models. Using this theorem to de-

scribe cochlear responses allows us to use the linear version of the model, for forward and inverse solutions.

That linear model is constructed in such a way that it reproduces or simulates the data obtained in a particular animal. Stimulation with signals of different levels results in different parameters: different BM impedance functions. For another animal the same type of model can be used, but it has to be given different parameters. With this technique it has proven possible to explain global and refined properties of linear as well as nonlinear effects observed in the cochlea; see the aforementioned reports on effects of stimulus intensity on the response and de Boer and Nuttall (2002) on tone versus noise responses. A different, equally universal but more conventional use of a model of the cochlea is amply illustrated by the work of Talmadge *et al.* (1998a, b) and Tubis *et al.* (2000a, b). In this case the model used is to be considered as a *generic* one, capable of explaining the most general properties of responses on the basis of mathematical elaborations.

In a cochlear model two types of waves propagating along the basilar membrane are, in general, possible: one in the direction of the apex (the forward-traveling wave) and one in the direction of the stapes (the backward-traveling or reverse wave). In a "classical" model (this term is defined in the Appendix) these two waves are equivalent in their modes of propagation. Whether or not both waves are observable depends on the circumstances (type of stimulation, local mechanical properties, and presence of internal reflections). Ever since otoacoustic emissions were detected (Kemp, 1978), attention has been given to reverse waves, their origin, and their properties, because otoacoustic emissions are generally assumed to arise from reverse waves reaching the stapes.

In particular, the question has been raised whether the total delay (latency) of an otoacoustic emission signal can be decomposed into a forward-wave delay and a reverse-wave delay. In the simplest case, that of a stimulus-frequency emission (SFOAE), the two delays would be, in a crude approximation, the same. However, the source of the forward wave is a simple, localized, one—the stapes—but in contrast, the source of the reverse wave is distributed over a certain range—the region of "activity." Therefore, even in this relatively simple case the interpretation is not simple. For a distortion-product emission (DPOAE) three frequencies are involved, and it is much more difficult to unravel the actual delay of the emission signal. Many authors have held the view that indeed the round-trip delay is composed of a forward and a backward delay, and can approximately be decomposed in this way (among others, Schneider *et al.*, 1999; Prijs *et al.*, 2000; Goodman *et al.*, 2003, 2004; Shera and Guinan, 2003).

Recently, findings by Tianing Ren (2004a,b, Ren *et al.*, 2006) have cast doubt on this opinion. Ren performed measurements in the gerbil at several locations along the basal part of the basilar membrane. In analyzing evidence of distortion products (DPs), he did not detect a reverse DP wave in situations where the source of the DP was located apically from the measurement location. And a new interpretation of the data (e.g., Ren, op. cit.; Ruggero, 2004; Siegel *et al.*, 2005; Ren and Nuttall, 2006) has appeared.

The present paper has been written with the intention of analyzing wave-traveling problems into some more detail and generality. As earlier in our work, a (classical) model of the cochlea is constructed on the basis of data obtained from one particular experiment. That model is subsequently analyzed with respect to several possible subtle effects associated with wave propagation. To wit, the model is stimulated with various types of stimulus and analyzed as to its frequency and space-domain responses. In Sec. II briefly we describe the technique used in obtaining and analyzing the original data, and the genesis of the model on that basis. In Sec. III describe specific operations carried out with the model, including the use of artificial forms of stimulation. In Sec. IV, a fundamental—and successful—test of the extended computation method is described. In Sec. V it is attempted to simulate excitation as it occurs in the genesis of Distortion Products (DPs) and DPOAEs. It is shown that invariably two waves are generated: one traveling in the direction of the apex and one in the direction of the base. In these model exercises, the reverse wave is *always* found to be dominant in the region near the stapes. In Sec. VI a method is developed to approximately separate these forward and reverse waves into two component waves. The result shows that the two component waves are of a different nature, but that their amplitudes are never more than 20 dB different. In a further test, "basis waves" for the model were created, and it was shown that the component waves behave essentially as these basis waves. In Sec. VII, finally, we summarize and discuss the results.

II. DATA AND MODEL

We first describe the acquisition of experimental data in general terms. Data on movements of the basilar membrane (BM) were collected from the basal turn of the guinea-pig cochlea with a laser velocimeter, for details see the papers cited earlier. The present study, on deeply anesthetized animals, was consistent with NIH guidelines for humane treatment of animals and was approved by the Oregon Health & Science Committee on the Use and Care of Animals. The best frequency (BF) at the recording site was between 15 and 18 kHz. As judged by the Compound Action Potential (CAP), the hearing loss due to surgery was between 7 and 25 dB at 18 kHz.

We have collected responses to sounds consisting of *wide bands of flat-spectrum pseudorandom noise*. The velocity of the BM was measured as a function of time, and input-output cross-correlation functions (ccfs) were computed from stimulus and response signals (see de Boer and Nuttall, 1997). Finally, the ccf spectrum was derived from the ccf waveform. For greater accuracy we constructed the "composite ccf spectrum" from parts of the frequency range measured with different bandwidths and stimulus levels—for low-frequency bands we used data with stimuli presented at higher levels—see de Boer and Nuttall (1999) for more details. Such "composite ccf spectra" will convey all the infor-

mation needed to develop a model of the cochlea adapted to a particular experiment. The last-cited paper also gives details on how the measured ccf spectrum is converted into a *cochlear response pattern*, a complex function of location variable x. The so-derived response pattern will be used as a "template."

Given a certain cochlear pattern, BM velocity as a function of distance x along the length of the BM (the template), we can apply the inverse-solution method. With that method the *BM impedance*—which again is a function of x—is determined in such a way that a cochlear model with this impedance produces a cochlear-pattern response that is (almost) identical to the original pattern. In view of the aforementioned EQ-NL theorem, this model is a linear model. The model has been made three dimensional so that it can accommodate long as well as short waves. A set of convenient approximate solution methods for this model has been published by de Boer (1997), but for the present work the Green's function method—as used in earlier work—was selected. The advantage is greater accuracy for uncommon situations (such as we will employ here). It should be noted at this point that the Green's function describes only the hydrodynamics of the fluid, constrained as it is by the geometry of the model. Details of the inverse-solution technique are found in de Boer and Nuttall, 1999 and details on the variations of that technique used for the present paper are given in the Appendix.

The EQ-NL theorem helps to interpret results where the original stimulus conditions force the cochlea into nonlinearity. In point of fact, only in the case of production of distortion products (DPs) will we use strong (tonal) stimuli and adapt our parameters appropriately, but the propagation of DP waves will be treated by linear perturbation theory. In all other cases treated in this paper we consider only linear systems.

Figure 1 illustrates the "template response" on which our model will be based. The particular experiment, 19922, was selected from our database because it yielded a BM impedance function that is fairly smooth. The upper panel shows the original response pattern $\nu_{\text{tem}}(x)$, plotted over a length of 6 mm (sufficient to accommodate the range of frequencies around 17 kHz—the best frequency for the location at which we made the original recording), in the form of a thick continuous curve for the amplitude and a somewhat thinner dotted curve for the phase. The response shown is derived from the measured response of the cochlea for a noise stimulus presented at 20 dB SPL per octave, and has been converted from the frequency to the location (x) domain. The curves shown should be interpreted as the basilar membrane velocity pattern, amplitude, and phase, for the frequency of 17 kHz. Although the figure shows response and impedance data over an interval of 6 mm, the actual model had a length of 12 mm, and that length was subdivided into 1024 points. The thin continuous line in this panel shows the amplitude of the response of the same animal recorded post-mortem. The difference between the maximal values of the two amplitude curves (approximately 30 dB) illustrates the signal amplification that takes place in the viable cochlea. The very thin dotted line shows the post-mortem phase. The

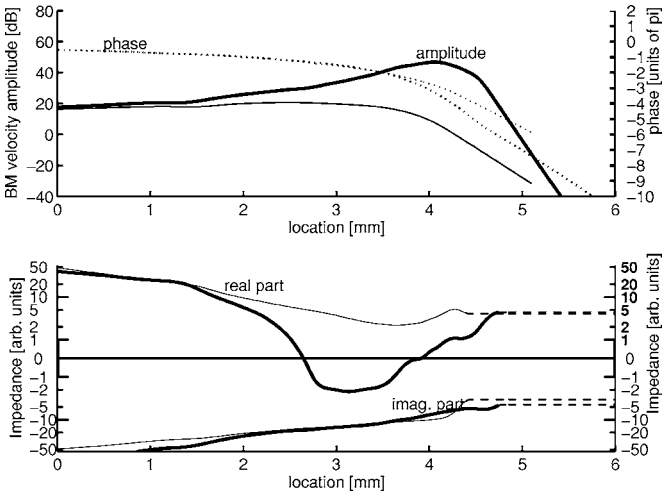


FIG. 1. The template used for the computations, derived from experiment 19922. *Upper panel*: measured response, BM velocity, converted from the frequency domain (f) to the location domain (x). Amplitude in dB re the stapes velocity. The phase in radians re stapes. Thick curves: viable cochlea. Stimulus intensity: 20 dB SPL per octave. Thin curves: response of the same animal recorded post-mortem. *Lower panel*: BM impedance, real and imaginary parts. The abscissa scale is the same as in the upper panel. The ordinate scale is nonlinear to accommodate large as well as small values. It is linear between −1 and 1 and compressed outside this range. The unit corresponds to 2 or 2000 kg m^{-2} s^{-1}. Dashed lines indicate how the impedance function is artificially extended beyond the range where the computation can be regarded as accurate. Thick curves: viable cochlea. Thin curves: impedance of the same animal's cochlea post-mortem.

curves have been drawn up to the point where the accuracy of the original, nonsmoothed, data becomes insufficient in our eyes. The lower panel of Fig. 1 shows the BM impedance in the two cases, on the same abscissa scale and a nonlinear ordinate scale, described in the legend to the figure. The impedance was derived with the inverse-solution method. The response of the model with this BM impedance has been computed (this is called the "resynthesized response"), and hasbeen found to be indistinguishable from the original responses, in both cases. In order to be able to carry out resynthesis the impedance functions are artificially extended (see the coarsely dashed lines on the right) beyond the points where the response functions are truncated. It is the model based on the viable low-level response of Fig. 1 that is used in the computations on wave propagation described further on. Parameters in the computation will be modified wherever the stimulus signal drives the cochlea into nonlinearity.

III. A POINT SOURCE

In the following step of our modeling exercise we strive toward extreme simplification of stimulus presentation. When we want to optimally consider forward and reverse waves, we should apply the stimulus signal in a spatially restricted form, not at the stapes but somewhere along the length of the cochlear pattern. The most abstract form of stimulus is then a point-like pressure source placed at a location along the length of the template. We refer to this stimulus type as "injection." The model response to such a stimulus follows from a straightforward solution to the model equation (see the Appendix for the technique applied

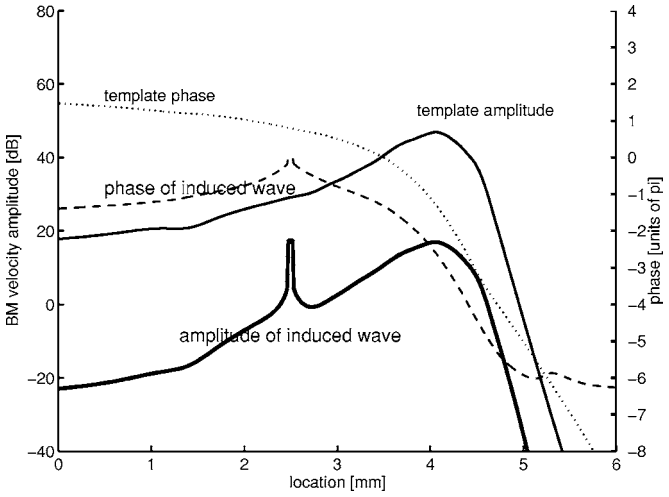


FIG. 2. Composite wave elicited by a spatially confined pressure source, no stapes stimulation. See the text. All functions shown are BM velocities. The template response is the same as in Fig. 1. "Injection" occurs in a small region around 2.5 mm. The phase curve indicates a reverse wave (phase angle decreasing with negative-going x) to the left of the injection region as well as a forward wave (phase angle decreasing with positive-going x) to the right of that region.

in this case). To reduce reflection at the stapes, which would complicate the figure, we have modified the model equation, by a procedure that is also described in the Appendix . As regards interpretation, we must take into account the artificiality of the situation. In addition, we depend on the accuracy of the Green's function solution method for this unusual stimulus situation. Figures 2 and 3 present the results of the computations for two placements of the source, and these figures clearly show the behavior of the generated waves in these two cases.

Figure 2 shows the wave pattern produced when a constant pressure (operating at 17 kHz) is prescribed ("injected") at five consecutive points centered at location x =2.5 (mm), i.e., near the left end of the "active" region. There is no stimulation elsewhere nor at the stapes, and reflection at the stapes is minimized, as stated above. It is clear from the phase curve that from the "injection" region a wave travels to the left (toward the stapes) and that another wave travels to the right (in the direction of the apex). On its path the amplitude of the reverse wave decreases gradually. This wave is not or barely amplified because its source lies at the left-hand edge of the active region (compare Fig. 1, lower panel). In contrast, the wave to the right (the forward wave) acquires and demonstrates the amplification that is associated with the template response.

Figure 3 shows the corresponding situation when the same type of "injection" occurs just to the right of the response peak, at x=4.3 (mm). Remember that in Figs. 2 and 3 the plotted patterns are solely due to "injection;" no stimulus is applied to the stapes, and stapes reflection is minimized. In Fig. 3 the left-going wave is more extended and shows the effect of local amplification occurring along its course. In contrast, the rightgoing wave covers only a small region of the x axis before it starts to behave like an evanescent wave—showing almost constant phase. Note, also, that the phase slopes near the injection region are steeper in Fig. 3 than in Fig. 2: the waves originate closer to the peak of the template response. Both Figs. 2 and 3 demonstrate that forward and reverse waves are created together, starting from a space-confined region. That region seems wider than the actual injection region, because cochlear activity contributes to local stimulation and is not spatially confined. The interference pattern shown in Fig. 3 in the 4 mm region may be due to interference of the waves in this extended region. There may be a contribution from interference of the reverse wave and a forward wave caused by residual reflection at the stapes. Finally, there may also be an influence of irregularity of the BM impedance in this region—remember that both forward and reverse waves will undergo amplification in the 4 mm region—and this makes it difficult to analyze the situation.

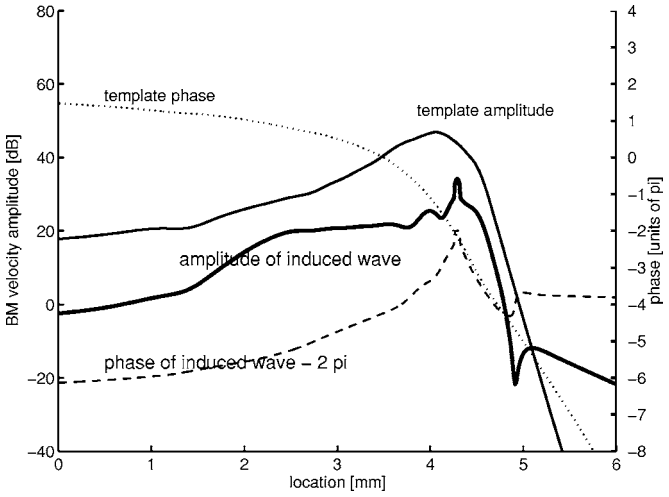


FIG. 3. The same as Fig. 2, but "injection" occurs here in the region around 4.3 mm, beyond the peak of the response. Again two waves are observed, but the forward-going one is soon quenched in the region where the BM impedance has become constant. The reverse wave undergoes amplification in the same region where the template response increases in power. Internal interaction and residual reflection from the stapes cause small amplitude and phase undulations to the left of the injection region.

IV. A DISTRIBUTED SOURCE

In the next step we critically test the extended solution procedure by approaching an actual and realistic situation very closely. We stimulate the cochlea with a pure tone, via the stapes. Using the terminology defined in the Appendix , we decompose the BM impedance—as determined from the inverse solution—as the sum of $Z_{\mathrm{BM}}^{\mathrm{pass}}$, the impedance associated with the response measured post-mortem, and $Z_{\mathrm{BM}}^{\mathrm{act}}$, the impedance specifically associated with the cochlear amplification system. Clearly, $Z_{\mathrm{BM}}^{\mathrm{act}}$ is determined by subtracting $Z_{\mathrm{BM}}^{\mathrm{pass}}$ from Z_{BM}:

$$Z_{\mathrm{BM}}^{\mathrm{act}} = Z_{\mathrm{BM}} - Z_{\mathrm{BM}}^{\mathrm{pass}}. \tag{1}$$

Note that that both terms in the right member of this equation are found from experiments (in the same animal). Then, the pressure p_{act} defined by

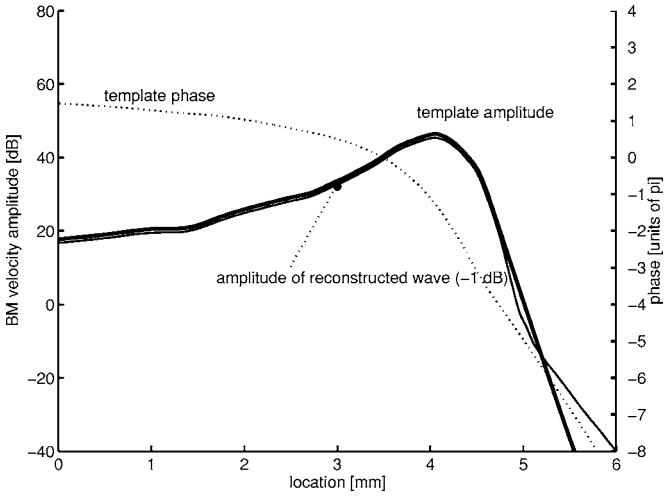


FIG. 4. Distributed "injection." Resynthesis with a "passive" model and injection with the "active" pressure. See the text. The resynthesized response is plotted 1 dB lower than it is computed, and as a thinner curve. The associated phase curve is so similar to that of the template that it has been omitted from the figure. This figure demonstrates the accuracy of the computation procedure.

$$p_{\mathrm{act}} = -1/2\,\nu_{\mathrm{BM}} \cdot Z_{\mathrm{BM}}^{\mathrm{act}}, \tag{2}$$

is the equivalent pressure generated by the Outer Hair Cells (OHCs), responsible for cochlear amplification. The minus sign and the factor $\frac{1}{2}$ are due to sign conventions, see the Appendix .

We next construct a model with the "*passive*" BM impedance $Z_{\mathrm{BM}}^{\mathrm{pass}}$, and stimulate it with both a given velocity of the stapes and the distributed pressure p_{act} from Eq. (2). Figure 4 shows the result. The so-computed response is virtually identical to the original response of the model (i.e., the template response) computed with the full BM impedance Z_{BM}. In order to show the individual curves, the amplitude curves have been displaced by 1 dB with respect to one another. It is seen that the correspondence extends to a point where the response is approximately 40 dB below its peak. The phase curves are indistinguishable over this range, and only the original function has been displayed. The conclusion is obvious: with a distributed form of (pressure) excitation the computation technique is virtually flawless.

One remark needs to be made with respect to Figs. 2–4. In the former two figures reflection at the stapes has been minimized with the aim of showing the reverse wave in its purest form. In Fig. 4 the stapes boundary condition has been returned to the original form—with a prescribed stapes velocity—in order to create a condition with distributed injection that is equivalent to the one used for the template response.

V. SIMULATION OF A DP CONDITION

We now attempt to realize a condition that is comparable to the one where two stimulus tones ("primaries"), with frequencies f_1 and f_2, are given and the principal DP is generated. We first need to simulate the response patterns of these primaries. For that purpose we construct a second template, from data obtained at a *stronger* stimulation (the primary stimuli are generally presented at a fairly high level) from the same animal, but still corresponding to the original frequency. As a result of nonlinearity, this template will, in general, have a lower and wider peak than the template used earlier. Next, two copies of that second template are made, these are displaced with respect to the original template and are made to correspond to the two primary frequencies f_1 and f_2. In doing this, we keep the DP frequency f_{DP}, given by

$$f_{\mathrm{DP}} = 2f_1 - f_2, \tag{3}$$

constant (equal to 17 kHz in this case), and assume the ratio f_2/f_1 to be given. Call the respective template functions $\nu_{\mathrm{prim1}}\,(x;f_1)$ and $\nu_{\mathrm{prim2}}\,(x;f_2)$, keeping in mind their respective frequencies, f_1 and f_2. These shifted templates represent the excitation provoked by the two primary tones. Note that they reflect the compression that the primary tones undergo.

We now need to obtain an estimate of how the OHCs are stimulated by them. Consider the presentation of one of these primaries, say, the first. Let $Z_{\mathrm{BM}}^{(1)}(x;f_1)$ be the BM impedance corresponding to this second template response, displaced according to the frequency f_1. Then, let $Z_{\mathrm{BM}}^{\mathrm{act1}}(x;f_1)$ be the active component of this impedance [cf. Eq. (1)]. As a result of nonlinearity the function $Z_{\mathrm{BM}}^{\mathrm{act1}}(x;f_1)$ differs from $Z_{\mathrm{BM}}^{\mathrm{act}}$ of Eq. (1), evaluated for the frequency f_1, especially in its real part. The model thus reflects the nonlinearity that this primary tone undergoes. The active component $p_{\mathrm{act1}}(x;f_1)$ of the pressure generated by the OHCs is given by

$$p_{\mathrm{act1}}(x;f_1) = -1/2\,\nu_{\mathrm{prim1}}(x;f_1) \cdot Z_{\mathrm{BM}}^{\mathrm{act1}}(x;f_1). \tag{4}$$

In our simple conception of the OHCs, we have assumed that these cells are simple no-memory transducers, all frequency dependence is absorbed in the complex nature of $Z_{\mathrm{BM}}^{\mathrm{act1}}(x;f_1)$ (see Fig. 1 and related description in de Boer and Nuttall, 2002). This means that $p_{\mathrm{act1}}(x;f_1)$ in Eq. (4) expresses, apart from a factor, the excitation of the OHCs at the frequency f_1, let us call that excitation $\psi_1(x;f_1)$. Similarly, $Z_{\mathrm{BM}}^{\mathrm{act2}}(x;f_2)$ is the active component of the BM impedance valid for the frequency f_2 for the second primary tone. The model thus also reflects the nonlinearity that the second primary tone undergoes. The active component $p_{\mathrm{act2}}(x;f_2)$ of the OHC pressure at the frequency f_2 is given by

$$p_{\mathrm{act2}}(x;f_2) = -1/2\,\nu_{\mathrm{prim2}}(x;f_2) \cdot Z_{\mathrm{BM}}^{\mathrm{act2}}(x;f_2). \tag{5}$$

The excitation associated with $p_{\mathrm{act2}}(x;f_2)$—which also is a complex function of x—will be called $\psi_2(x;f_2)$. Now, let both primary tones be presented together. Inside the hair cells the sum signal is subjected to nonlinear distortion. We assume that the third-order term in the polynomial expansion of this distortion term is the most important one. Expanding the expression

$$[\mathrm{Re}(\psi_1(x;f_1)e^{2\pi i f_1 t} + \psi_2(x;f_2)e^{2\pi i f_2 t})]^3, \tag{6}$$

in terms with various frequencies, gives as the principal DP with the frequency $2f_1 - f_2$, a term of which the effective (complex) amplitude $A_{\mathrm{DP}}(x)$ is given by

$$A_{\mathrm{DP}}(x) = 3/4A\,\psi_1^2(x;f_1) \cdot \psi_2^*(x;f_2). \tag{7}$$

The square arises because the DP frequency f_{DP} contains the term $2f_1$, the asterisk (*) denotes the complex conjugate

and is needed because the DP frequency contains the term $-f_2$. We have included a constant of proportionality A that at the moment is left unspecified, we will come back to this factor presently. A pressure proportional to $A_{DP}(x)$, having the frequency f_{DP} (equal to 17 kHz), will be active as the injection source for the DP, it will be denoted by $p_{act}(x)$. The pressure $p_{act}(x)$ or, abbreviated, p_{act}, can be included in the model computation in the same sense as in the preceding sections; see more later. In evaluating expression (7), the actual template functions $\nu_{prim1}(x;f_1)$ and $\nu_{prim2}(x;f_2)$ and the associated impedance functions $Z_{BM}^{act1}(x;f_1)$ and $Z_{BM}^{act2}(x;f_2)$ have to be substituted. In summary, this implies that nonlinearity of the response to the primary tones is properly taken care of. Although in experimental practice the f_1 tone is generally presented at a 10 dB higher level than the f_2 tone, we have neglected this difference and have used one level (60 dB in this case) for the template functions of both primary tones. The reason is that the difference in the shape of the response patterns would be small and that the actual stimulus amplitude would only appear as a constant factor in Eqs. (4) and (7). In addition, we omit suppression of the propagating DP wave by the primary tones.

We now solve the model equation for the wave with the DP frequency. In this solution the model is equipped with the BM impedance function that corresponds to the low-level template of Fig. 1 because we consider propagation of the DP as a linear perturbation. As in Sec. III, we remove excitation with the DP frequency at the stapes in this case because no DP component is present in the actual stimulus signal at the stapes. Furthermore, we minimize the reflection of reverse waves at the stapes to simplify interpretation. The insertion of Eq. (7) in the model equation (A9) yields the DP component of the BM velocity, to be called $\nu_{DP}(x)$. The proportionality factor A in Eq. (7) has been chosen to make the amplitude of $\nu_{DP}(x)$ at the location $x=4$ (mm) equal for the three figures to follow, in this way following the Allen-Fahey paradigm (cf. de Boer *et al.*, 2005a). Figure 5 shows the result for the frequency ratio f_2/f_1 equal to 1.05. In this case the overlap in the excitation patterns is quite large, as the function labeled p_{act} in the figure demonstrates. The DP velocity has a large and pronounced peak in its amplitude, somewhat more apically than the overlap region (thickest curve).

Two waves emerge from this peak: one going to the left and one going to the right. Both waves undergo amplification by the cochlear amplification mechanism, an effect that is included in our computation paradigm. Note that in this case the reverse wave is relatively small in its amplitude. This is caused by wave interference in the sense described by Shera (2003) and by de Boer and Nuttall (2006). In Fig. 6, where f_2/f_1 is equal to 1.2, the response situation is somewhat easier to interpret. Especially from the course of the phase curve, the two waves are clearly discernible. Figure 7, finally, shows the case where f_2/f_1 is larger, equal to 1.4; here the peak in $\nu_{DP}(x)$ is closer to the stapes; and it is evident that the right-going wave is undergoing appreciable amplification. This effect is to be expected, of course. On the other hand, from the phase pattern it is evident that a substantial leftgoing wave is traveling toward the stapes. In an actual

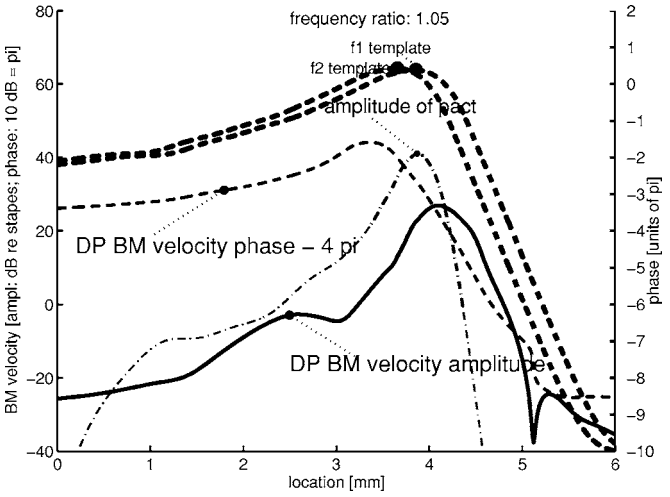


FIG. 5. Simulated DP waves. The template response for the DP frequency (17 kHz) is the same as in the preceding figures, but has been omitted from this figure. The upper two curves (gray dashed lines) illustrate the excitation amplitude by the two primary tones, with frequencies f_1 and f_2, the templates correspond to an excitation level of 60 dB SPL and are shown in this figure as displaced upward by 20 dB (more symbolically than realistically). The peaks of these templates are somewhat wider than the peak of the "DP template" (which corresponds to an excitation level of 20 dB SPL). Frequency ratio (f_2/f_1): 1.05. The curve labeled "amplitude of pact" shows the course of the amplitude of the effective DP pressure $p_{act}(x)$, which is proportional to $A_{DP}(x)$ in Eq. (6). This pressure corresponds to the DP frequency (17 kHz). The simulated DP wave $\nu_{DP}(x)$, generated by $p_{act}(x)$, is shown by the curves labeled "DP BM velocity amplitude" and "DP BM velocity phase." In this and the following two figures, the DP wave is depicted with a constant velocity amplitude at location 4 (mm) (cf. de Boer *et al.*, 2005a). There is no drive at the stapes with the DP frequency, and reflection at the stapes is minimized.

experiment the level of the primary tones has to be increased for increasing frequency ratio f_2/f_1. This is indicated in the figures by the increased level of the templates for the primary tones. The increase shown is only symbolical, of course, because the actual level increases would be much larger. Fur-

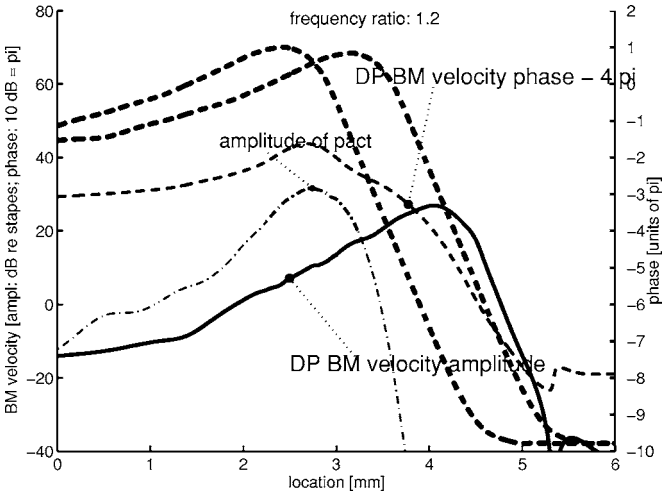


FIG. 6. Simulated DP waves. Layout as in Fig. 5. Frequency ratio (f_2/f_1): 1.20. Note how the generating pressure $p_{act}(x)$ has a peak in the overlap region between the primary templates, but mainly on the side of the f_2 template pattern. This pressure $p_{act}(x)$ causes DP waves to arise. One DP wave is going to the left (reverse wave) and one to the right (forward wave). From the interference pattern of the DP BM velocity it can be concluded that, internally, some degree of interaction exists.

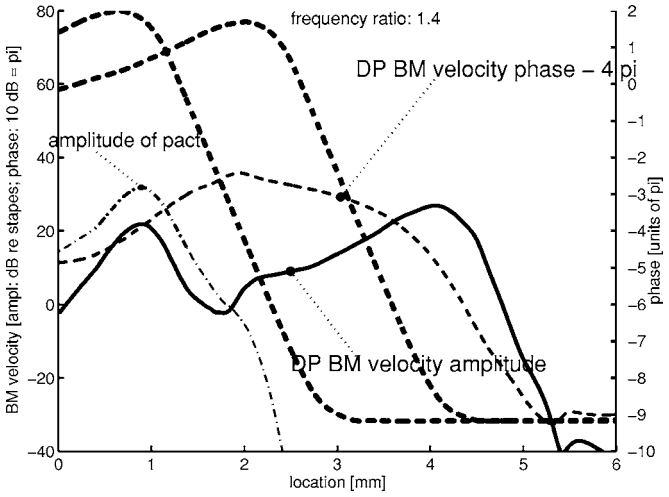


FIG. 7. Simulated DP waves. Layout as in Fig. 4. Frequency ratio (f_2/f_1): 1.4. Under this condition, it is evident that the forward wave is amplified in the same way as the template response. Note that near the stapes the DP wave is traveling backward.

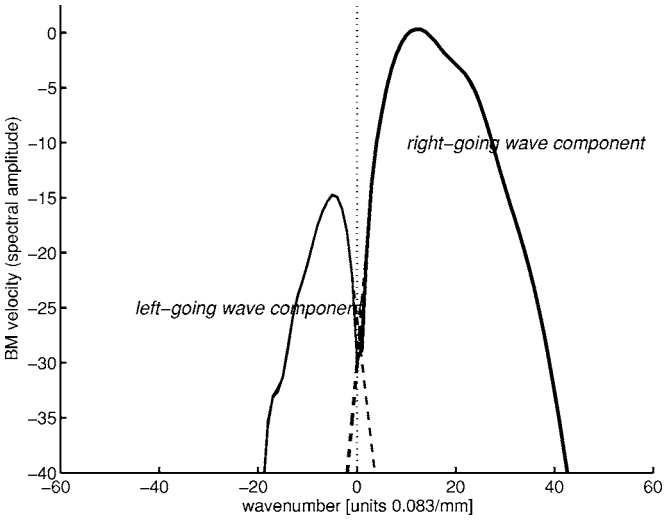


FIG. 8. Waveform spectra used in separation of waves. Wavenumber in units of 0.083 mm^{-1}. Thick curve: amplitude of spectrum $\tilde{v}_{DP}(k)$ of DP BM velocity signal $v_{DP}(x)$, taken from Fig. 6; f_2/f_1 equal 1.2. Thin curves: wavenumber spectra of right- and left-going components; both components are extrapolated into the "opposite" domain (dashed lines).

thermore, we have abstained from using templates adapted to each f_2/f_1 condition—wider and lower for higher primary-tone levels–because in our experience the effect of wider primary-tone responses is small with this type of computation.

Let us return to Fig. 5. With f_2/f_1 near 1, a good deal of wave interference takes place; this is the reason why the DP velocity at the stapes is fairly small. With increasing frequency ratio (Fig. 6), this velocity increases, in the same manner as the data in de Boer *et al.* (2005a). A decreasing degree of wave interference would constitute an explanation of this property. Consider next Fig. 7. When the level of the primaries were the same as before, the wave arriving at the stapes would be smaller than in Fig. 6. That it is actually larger is, first, due to the fact that the level of the primaries has to be higher in order to achieve a constant DP level at the DP place, and, second, due to the nearness of the peak evoked by the f_2 tone to the stapes—a corresponding situation is not usually encountered in experiments on DPOAEs. For these reasons we observe only half of the familiar "band-pass" character of DPOAEs.

VI. SEPARATION OF WAVES

In principle it seems simple to separate the two waves that are visible in Figs. 5 to 7. The function $v_{DP}(x)$, the DP component in the cochlear response pattern, can be Fourier transformed to the wavenumber domain $\tilde{v}_{DP}(k)$. Wave components with $k>0$ would refer to rightgoing and components with $k<0$ to leftgoing waves. This is, however, a too simplistic attitude. Consider, for a moment, the "template" response $v_{tem}(x)$ of Fig. 1. In the figure it appears as a monotonic rightgoing wave, yet its Fourier transform $\tilde{v}_{tem}(k)$ proves to have substantial nonzero terms for $k<0$. It is easy to understand where these come from: the wave is monotonic in its phase course, but it does not have a constant amplitude nor a constant wave-propagation velocity. Therefore, its wavenumber spectrum is not a single peak (delta function) but appears to be convolved with some kind of modulation function. The convolution causes the spectrum to spread out, and even to cross the zero line. Some degree of internal reflection could add to the same effect.

It appears unavoidable, then, that the wave spectrum $\tilde{v}_{DP}(k)$ should be decomposed into two parts that are both *two sided* in the wavenumber domain. One part will have its major contribution in the positive and the other in the negative k domain. This type of decomposition is not unique, however, and we have few rules and reasons to guide us. We have decided to use the following procedure. Figure 8 shows the wavenumber spectrum $\tilde{v}_{DP}(k)$ corresponding to $v_{DP}(x)$ of Fig. 6, on a logarithmic ordinate scale (amplitude only). We have isolated the part with $k>0$ and extended it into the region $k<0$ with a fixed slope (actually, that slope is a complex number); see the thick dashed line. In this way we believe to have obtained a good estimate of the complete spectrum of the rightgoing component wave. We have carried out the corresponding procedure with the isolated part of $\tilde{v}_{DP}(k)$ with negative wavenumbers, extending it into the domain of positive wavenumbers (the thin dashed line). This leads to an estimate of the spectrum of the leftgoing component wave. The two resulting spectra have the property that they are continuous in value and slope at and around the point $k=0$. In the x domain the resulting two wave-component patterns can be expected to be smooth and "well behaving."

Figure 9 shows, for the value of f_2/f_1 (1.2) of Fig. 6, the decomposition of the $v_{DP}(x)$ wave. The figure is a simplified version of Fig. 6 and shows the extracted rightgoing and leftgoing component waves by thin curves—solid curves for the amplitude and dotted curves for the phase. Indeed, the right-going component wave, to be called $v_{DP}^{right}(x)$, approximates the right-hand part of the $v_{DP}(x)$ function quite well. Note that over the entire x domain its phase curve has an almost monotonic slope. The course of the leftgoing component wave, to be called $v_{DP}^{left}(x)$, is somewhat less regular. It should be noted that the separation between the original $v_{DP}(x)$ amplitude curve and the amplitudes of the two com-

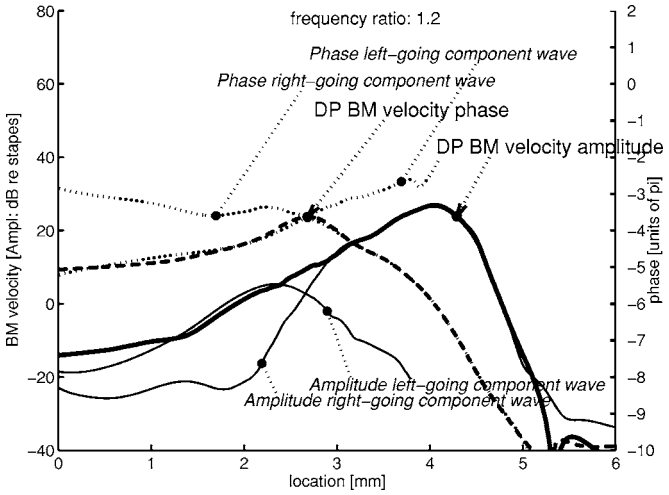


FIG. 9. Decomposition of DP wave signal $\nu_{DP}(x)$ into a right- and a left-going "component wave." Thick continuous curve: amplitude of $\nu_{DP}(x)$. Thick dashed curve: the associated phase. Thin continuous curves: the two constituent wave components, amplitudes. Thin dotted curves: the phase courses associated with these wave components. These phase curves have almost monotonic slopes. Note how amplitudes and phases of the two constituent waves more or less smoothly approximate amplitude and phase of the original $\nu_{DP}(x)$ signal. Phase curves have been shifted by integer multiples of 2π to achieve the correct match.

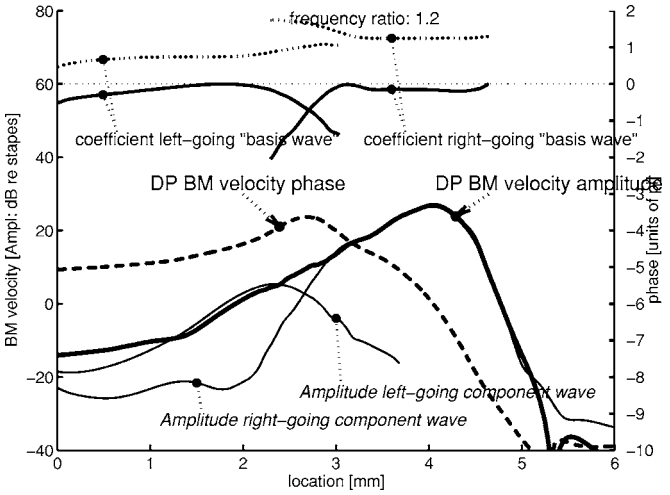


FIG. 10. The two component waves have an intrinsic behavior that resembles very much that of the "basis waves" (see the text). The upper part of figure shows the coefficients $A(x)$ and $B(x)$ over the region where they are significant. Solid curves: amplitudes, dotted curves: phases. Amplitudes of the components have been shifted so that their maximal levels lie at +60 dB, phases have been shifted by integer multiples of 2π to make them come closest to the zero line (dotted horizontal line in the figure).

ponent waves is never very large, averaging 15 dB for the leftgoing component and being somewhat larger for the rightgoing component. This reveals that our separation method may be an approximate one, but it may well be optimal in the sense that a "better" separation cannot be achieved. We have applied the same separation method to the results shown by Figs. 2 and 3, and we found the separation of the two wave components to have the same properties as shown by Fig. 9. In summary, it should be noted that the separation of waves shown by Fig. 9 is based on Fourier transforms, that is, on *average* properties of the actual wave $\nu_{DP}(x)$, averaged over x.

For one- and two-dimensional cochlear models the theoretical study of internal reflection has been facilitated by the definition of "basis waves" (Shera and Zweig, 1991; Talmadge *et al.*, 1998a; Shera *et al.*, 2005). Usable forms of such basis waves can be obtained from specialized resynthesis results of our three-dimensional model. Consider Fig. 4, the resynthesized response is a wave that appears to be solely consisting of a rightgoing wave, a few small ripples signify an additional component. Applying a moderate amount of smoothing to the BM impedance functions and doing resynthesis produces a wave function that is smoother and unidirectional in a more pronounced way. We will call that wave the first basis wave, $\nu_{basis}^{right}(x)$. Consider next Fig. 3; injection occurs here far to the right and the resulting wave is dominated by a leftgoing wave. Putting the injection point still more to the right will yield a wave that can serve as the second basis function, $\nu_{basis}^{left}(x)$. We may consider these two responses as due to "extreme" types of stimulation. Let us now return to the component waves $\nu_{DP}^{right}(x)$ and $\nu_{DP}^{left}(x)$ of Fig. 9, obtained from simulation of DP generation. Write each of these wave components in the form of a slowly varying coefficient times one of the basis waves:

$$\nu_{DP}^{right}(x) = A(x) \cdot \nu_{basis}^{right}(x), \tag{8a}$$

$$\nu_{DP}^{left}(x) = B(x) \cdot \nu_{basis}^{left}(x). \tag{8b}$$

The functions $A(x)$ and $B(x)$ play a part similar to that of the "osculating parameters" (Shera and Zweig, 1991). Figure 10 illustrates the character of this decomposition. The figure shows, for orientation, the main curves of Fig. 9 and, in the top part of the picture, amplitude (solid curves) and phase (dotted curves) of the coefficients $A(x)$ and $B(x)$. Each of these is shown only over the range where a basis function is expected to approximate a component wave function. The amplitudes of the coefficients $A(x)$ and $B(x)$ have been shifted so that their maximal levels lie at the level of +60 dB, the phase curves have been shifted by integer multiples of 2π to make them come closest to the zero line (the dotted horizontal zero-phase line in the figure). From this figure it is clear that over a certain range the amplitudes of the two coefficients are nearly constant, the phase of $A(x)$ is almost π, and the phase of $B(x)$ has a small slope. Indeed, the coefficient functions are varying more slowly than the basis functions, and the two extracted component waves therefore have an intrinsic behavior that resembles very much that of the "basis waves." If the coefficients were constant, the component waves could be thought to originate from one point (as the basis waves do). Apparently, the distribution of excitation as involved in the generation of a DP is to a certain degree equivalent with pointwise excitation. There appears to be relatively little interaction between the two waves and little evidence of internal reflection. On a final note, it is stressed that in the analysis demonstrated in Fig. 9 the creation of the component waves is based on wavenumber spectra, whereas the present description of basis waves is founded on similarity of wave-pattern *shapes*, as indicated by the slow variations of the factors $A(x)$ and $B(x)$.

VII. CONCLUDING COMMENTS

In this study we have shown that a simulation of the process of creation of a DP [in this case always the cubic difference tone, i.e., the DP with the frequency given by Eq. (3)] *invariably produces two waves*: one, the forward wave, traveling in the direction of the apex and the other one, the reverse wave, traveling in the direction of the stapes. It proved possible to isolate these two waves to a certain extent (see Fig. 9). In view of our type of modeling the cochlea, it is the reverse wave, which, having arrived at the stapes, actually gives rise to a DPOAE observable in the external ear canal. The existence of the two waves is inextricably connected with the physical-mathematical nature of the hydrodynamics of the system. In terms of the variable x all expressions are symmetrical. This does not necessarily mean that under all circumstances such a pair of symmetrical waves arises.[1] This depends on the type of excitation.

In the case treated in the present paper the excitation that simulates DP generation is distributed over only a part of the length of the model. To be more precise the injection is effective (i.e., produces DPs) only inside the strip, where the model is "active" for both of the primary frequencies. From every point in this strip two DP waves emerge: a forward and a reverse wave. Therefore, distributed (Figs. 5 to 7) as well as pointwise injections (Figs. 2 to 4) have the same basic property: they excite *both* of the wave components. In searching through our database, and applying the analysis described in this paper, we have found nine experiments where the BM impedance function was sufficiently smooth to allow the analysis of this paper. In all these the results were equivalent to those shown here. In a further six experiments the results were somewhat less clear but still easily interpretable. We have not found a case where the reverse wave component dominates in the stapes region.

In Sec. VI it was attempted to separate the composite DP wave into a forward-propagating and a backward-propagating component wave. That attempt was successful, but the amplitudes of the two component waves generally remain at a limited distance (generally, less than 25 dB). This result may indicate that the component waves can never be separated *completely*. Internal reflection due to irregularity of the BM impedance function may constrain this distance, too. By way of Fig. 10 we have shown that each of the two component waves can be considered as a "basis function" that is only slowly modulated (in amplitude and phase). Starting from the central region (the region of overlap), each wave gradually increases its amplitude, and there is relatively little evidence of interference and internal reflection. It should be noted, finally, that all our conclusions are valid for the type of model used, i.e., a classical model—defined in the Appendix as a model governed by a driving point impedance $Z_{\mathrm{BM}}(x,\omega)$. An extension of our work for nonclassical models is in progress.

In our work we have artificially minimized reflection at the stapes. If we would re-introduce reflection at this point, the amplitude of the forward component wave could be increased. Theoretically, it might thus be possible that the forward component wave dominates the reverse component wave. Such a case we have not been able to find in our data material, however.

ACKNOWLEDGMENTS

Most of the experiments from which we selected the response used as the "template" in this work were executed in close collaboration with Jiefu Zheng. We acknowledge with great pleasure his extreme expertise and diligence in this task. From the many datasets collected, we have selected a "template" response that seems to show the smallest degree of pollution by data errors and a minimum amount of internal reflection. This study received support from NIH NIDCD R01 DC00141 (ALN) and R01 DC003687 (CAS).

APPENDIX

The matrix formulation has earlier been described in de Boer and Nuttall (1999, 2000b). We recapitulate the description here. We will use the following symbols: x is the longitudinal coordinate, ω the radian frequency, ρ the density of the fluid, $v_{\mathrm{BM}}(x,\omega)$ the BM velocity, and $Z_{\mathrm{BM}}(x,\omega)$ is the BM impedance. The height and width of the model are h and b, respectively. The parameter ε is the width of the BM divided by the width b of the model. To prepare for greater generality we have reformulated the solution in the way described by Allen (1977), Allen and Sondhi (1978), and Mammano and Nobili (1993). First, divide the x axis into N discrete points x_i $(\mathrm{i}=1,\ldots,N)$. The pressure in the fluid $p(x,\omega)$—close to the BM in the upper channel—and the BM velocity $v_{\mathrm{BM}}(x,\omega)$ are represented by column vectors $\mathbf{p}$ and $\mathbf{v}$, both of length N. The *hydrodynamics of the fluid* inside the model is described by a matrix $\mathbf{G}$ of size $N\times N$, which represents the Green's function—which function relates pressure to acceleration but is here used in a form to relate pressure to velocity. The pressure $\mathbf{p}$ can be expressed as

$$\mathbf{p} = i\omega\rho(\mathbf{G}\mathbf{v} + \mathbf{s}v_{\mathrm{st}}). \tag{A1}$$

Here $\mathbf{s}$ is a column vector (called the stapes propagator), which represents the way the stapes boundary condition expresses itself in the pressure $p(x,\omega)$, and v_{st} is the stapes velocity. In this approach, $\mathbf{G}$ and $\mathbf{s}$ express the hydrodynamics of the fluid constrained as it is by the geometry of the model and the presence of the stapes and the helicotrema. Furthermore, $\mathbf{G}$ and $\mathbf{s}$ automatically incorporate boundary conditions. Assume further that pressure and velocity are related via the driving-point BM impedance $Z_{\mathrm{BM}}(x,\omega)$. This implies that, mechanically, the BM receives only stimulation from its own local velocity, all other stimulation comes via the fluid, this defines our model as a "classical model." Write the relation between pressure and velocity as[2]

$$\mathbf{p} = -1/2\mathbf{Z}\mathbf{v}, \tag{A2}$$

where $\mathbf{Z}$ is a diagonal $N\times N$ matrix, which has $Z_{\mathrm{BM}}(x_{\mathrm{i}},\omega)$ $(\mathrm{i}=1,\ldots,N)$ in its main diagonal. Substitute $\mathbf{p}$ from Eq. (A2) in Eq. (A1):

$$(i\omega\rho\mathbf{G} + 1/2\mathbf{Z})\mathbf{v} = -i\omega\rho\mathbf{s}v_{\mathrm{st}}. \tag{A3}$$

In the *forward* solution, Eq. (A3) is solved for $\mathbf{v}$ with

given $\mathbf{Z}$ and ν_{st}. In the *inverse* solution the pressure $\mathbf{p}$ is computed directly from Eq. (A1), $\mathbf{v}$ and ν_{st} being given, and Eq. (A2) is applied to find the impedance. By the nature of the problem formulation the matrix $\mathbf{G}$ is a "full" matrix. We have derived $\mathbf{G}$ and $\mathbf{s}$ in the manner described by Mammano and Nobili (1993), but have given our "stylized" model a constant width b, a constant BM width ϵb (ϵ=0.2), and a constant height h. For simplicity, we have made b equal to h, and equal to 1 (mm). The number N of sections is 1024. The length of the model was chosen to be 12 mmof which the first 6 mm are shown in the figures.

The model is normally driven by a given velocity, the stapes velocity, ν_{st}. In order to minimize reflection from the stapes we consider the stapes to be driven via an impedance Z_0, from a source with strength ν_0. In that case the following relation holds:

$$\nu_{st} = \nu_0 - p(1)/Z_0, \tag{A4}$$

where ν_0 is the driving velocity and $p(1)$ is the pressure at the stapes location. The matrix equation (A3) is formulated in terms of velocity, hence Eq. (A5) has to be rewritten as

$$\nu_{st} = \nu_0 + 1/2[Z_{BM}(1)/Z_0]v(1). \tag{A5}$$

The term with $v(1)$ has to be included in the first term of Eq. (A3). To this aim a matrix $\mathbf{G}_{01}$ is created that contains all zeros, except for the first column, where it contains the elements of the vector $\mathbf{s}$. When used in a matrix equation, $\mathbf{G}_{01}$ would operate only on the unknown variable $v(1)$. This leads to the revised model equation,

$$[i\omega\rho(\mathbf{G} + 1/2(Z_{BM}(1)/Z_0)\mathbf{G}_{01}) + 1/2\mathbf{Z}]\mathbf{v} = -i\omega\rho\mathbf{s}\nu_0. \tag{A6}$$

For a long-wave model with given BM impedance $Z_{BM}(0)$, the characteristic impedance Z_0 is given by

$$Z_0 = [1/2i\omega\rho h_{eff} Z_{BM}(0,\omega)]^{1/2}, \tag{A7}$$

where h_{eff} is the "effective height" of the model, the cross-sectional area of the channels divided by the width of the BM. Note that expression (A8) has the same dimension as the BM impedance $Z_{BM}(0)$. As said above, this modification of the model equation will be used to minimize reflection at the stapes. Where there is no stimulation at the stapes we put ν_0 equal to zero.

Equation (A3) is in reality an equation in which each term represents a pressure. When $\mathbf{p}_{act}$ is the extra pressure (injected or generated by OHCs), expressed as a column vector, the equation can be rewritten as

$$(i\omega\rho\mathbf{G} + 1/2\mathbf{Z})\mathbf{v} = -i\omega\rho\mathbf{s}\nu_{st} + \mathbf{p}_{act}. \tag{A8}$$

To test the extended solution method (see Fig. 4), $\mathbf{p}_{act}$ has been given as the actual "active" pressure component generated by the OHCs (computed from the BM velocity and the active component of the BM impedance), for $\mathbf{Z}$ is taken the passive impedance component, and the equation is solved. The resulting response pattern is almost indistinguishable from the "active" resynthesized response of the model—as Fig. 4 demonstrates. Equation (A8) combined with Eq. (A6) is used to compute the distribution of wave components generated by a distributed set of pressure sources with minimized reflection at the stapes (Figs. 2–7).

[1]It is easy to find a counterexample. Consider a simple model with an extremely smooth BM impedance function, stimulated from the stapes. In such a model virtually no reflection occurs, only one of the two possible waves appears to be realistic.

[2]The minus sign and the factor 1/2 are due to sign conventions. Upward movement of the BM is counted positive, $\mathbf{p}$ is the pressure in the upper channel and $-\mathbf{p}$ in the lower channel. Velocity and pressure should be interpreted as averages over the width of the BM.

Allen, J. B. (**1977**). "Two-dimensional cochlear fluid model: New results," J. Acoust. Soc. Am. **61**, 110–119.

Allen, J. B. and Sondhi, M. M. (**1979**). "Cochlear macromechanics—time domain solutions," J. Acoust. Soc. Am. **66**, 123–132.

de Boer, E. (**1997**). "Connecting frequency selectivity and nonlinearity for models of the cochlea," Aud. Neurosci. **3**, 377–388.

de Boer, E. and Nuttall, A. L. (**1997**). "The mechanical waveform of the basilar membrane. I. Frequency modulations ("glides") in impulse responses and cross-correlation functions," J. Acoust. Soc. Am. **101**, 3583–3592.

de Boer, E. and Nuttall, A. L. (**1999**). "The "inverse problem" solved for a three-dimensional model of the cochlea. III. Brushing-up the solution method," J. Acoust. Soc. Am. **105**, 3410–3420.

de Boer, E. and Nuttall, A. L. (**2000a**). "The mechanical waveform of the basilar membrane. II. From data to models-and back," Journal **107**, 1487–1496.

de Boer, E. and Nuttall, A. L. (**2000b**). "The mechanical waveform of the basilar membrane. III. Intensity effects," J. Acoust. Soc. Am. **107**, 1497–1507.

de Boer, E. and Nuttall, A. L. (**2002**). "The mechanical waveform of the basilar membrane. IV. Tone and noise stimuli," J. Acoust. Soc. Am. **111**, 979–989.

de Boer, E. and Nuttall, A. L. (**2006**). "Spontaneous basilar-membrane oscillation (SBMO) and coherent reflection," J. Assoc. Res. Otolaryngol. **7**, 26–37.

de Boer, E., Nuttall, A. L., Hu, N., Zou, Y., and Zheng, J. (**2005a**). "The Allen-Fahey experiment extended," J. Acoust. Soc. Am. **117**, 1260–1267.

Goodman, S. S., Withnell, R. H., and Shera, C. A. (**2003**). "The origin of SFOAE microstructure in guinea pigs," Hear. Res. **183**, 7–17.

Goodman, S. S., Withnell, R. H., de Boer, E., Lilly, D. J., and Nuttall, A. L. (**2004**). "Cochlear delays measured with amplitude-modulated tone-burst evoked OAEs," Hear. Res. **188**, 57–69.

Kemp, D. T. (**1978**). "Stimulated acoustic emission from within the human auditory system," J. Acoust. Soc. Am. **64**, 1386–1391.

Mammano, F., and Nobili, R. (**1993**). "Biophysics of the cochlea: Linear approximation," J. Acoust. Soc. Am. **93**, 3320–3332.

Nuttall, A. L., Grosh, K., Zheng, J., de Boer, E., Zou, Y., and Ren, T. (**2004**). "Spontaneous basilar membrane oscillation and otoacoustic emission at 15 kHz in a guinea pig," J. Assoc. Res. Otolaryngol. **5**, 337–349.

Prijs, V. F., Schneider, S., and Schoonhoven, R. (**2000**). "Group delays of distortion product otoacoustic emissions: Relating delays measured with $f1$- and $f2$-sweep paradigms," J. Acoust. Soc. Am. **107**, 3298–3307.

Ren, T. (**2004a**). "Propagation direction of the otoacoustic emission along the basilar membrane," Association for Research in Otolaryngology, Mid-Winter-Meeting Abstracts, Vol. **27**, p. 343 (#1011).

Ren, T. (**2004b**). "Reverse propagation of sound in the gerbil cochlea," Nat. Neurosci. **7**, 333–334.

Ren, T., He, W., and Nuttall, A. L. (**2006**). "Backward propagation of otoacoustic emissions in the cochlea," in: *Auditory Mechanisms: Processes and Models*, edited by A. L. Nuttall, T. Ren, P. Gillespie, K. Grosh, and E. de Boer (World Scientific, Singapore), pp. 79–85.

Ren, T. and Nuttall, A. L. (**2006**). "Cochlear compression wave: An implication of the Allen-Fahey experiment," J. Acoust. Soc. Am. **119**, 1940–1942.

Ruggero, M. A. (**2004**). "Comparison of group delay of $2f_1-f_2$ distortion product otoacoustic emissions and cochlear travel times," ARLO **5**, 143–147.

Schneider, S., Prijs, V. F., and Schoonhoven, R. (**1999**). "Group delays of distortion product otoacoustic emissions in the guinea pig," J. Acoust. Soc. Am. **105**, 2722–2730.

Shera, C. A. (**2003**). "Wave interference in the generation of reflection- and

distortion-source emissions," in *Biophysics of the Cochlea: From Molecule to Model*, edited by A. W. Gummer, E. Dalhoff, M. Nowotny, and M. P. Scherer (World Scientific, Singapore), pp. 439–454.
Shera, C. A. and Guinan, J. J. (**2003**). "Stimulus-frequency-emission group delay: A test of coherent reflection filtering and a window on cochlear tuning," J. Acoust. Soc. Am. **113**, 2762–2772.
Shera, C. A. and Zweig, G. (**1991**). "Reflection of retrograde waves within the cochlea and at the stapes," J. Acoust. Soc. Am. **89**, 1290–1305.
Shera, C. A., Tubis, A., and Talmadge, C. L. (**2005**). "Coherent reflection in a two-dimensional cochlea: Short-wave versus long-wave scattering in the generation of reflection-source otoacoustic emissions," J. Acoust. Soc. Am. **118**, 287–313.
Siegel, J. H., Cerka, A. J., Recio-Spinoso, A., Temchin, A. N., van Dijk, P., and Ruggero, M. A. (**2005**). "Delays of stimulus-frequency otoacoustic emissions and cochlear vibrations contradict the theory of coherent reflection filtering," J. Acoust. Soc. Am. **118**, 2434–2443.
Talmadge, C. L., Tubis, A., Long, G. R., and Piskorski, P. (**1998a**). "Modeling otoacoustic emissions and hearing threshold fine structures," J. Acoust. Soc. Am. **104**, 1517–1543.
Talmadge, C. L., Tubis, A., Piskorski, P., and Long, G. R. (**1998b**). "Modeling otoacoustic emission," in *Diversity in Auditory Mechanisms*, edited by E. R. Lewis, G. R. Long, R. F. Lyon, P. M. Narins, C. R. Steele, and E. Poinar (World Scientific, Singapore), pp. 462–471.
Tubis, A., Talmadge, C. L., and Tong, C. (**2000a**). "Modeling the temporal behavior of distortion product otoacoustic emissions," J. Acoust. Soc. Am. **107**, 2112–2127.
Tubis, A., Talmadge, C. L., Tong, C., and Dhar, S. (**2000b**). "On the relationships between the fixed-f_1, fixed-f_2, and fixed-ratio phase derivatives of the $2f_1-f_2$ distortion product otoacoustic emission," J. Acoust. Soc. Am. **108**, 1772–1785.
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.

Spectral modulation detection as a function of modulation frequency, carrier bandwidth, and carrier frequency region

David A. Eddins[a)]
Department of Otolaryngology, University of Rochester, Rochester, NY 14642 and International Center for Hearing and Speech Research, Rochester Institute of Technology, Rochester, NY 14623

Eva M. Bero
Department of Audiology, University of Massachusetts Memorial Medical Center, 15 Belmont Street, Worchester, MA 01605

(Received 25 October 2005; revised 21 September 2006; accepted 3 October 2006)

The present study investigates the nature of spectral envelope perception using a spectral modulation detection task in which sinusoidal spectral modulation is superimposed upon a noise carrier. The principal goal of this study is to characterize spectral envelope perception in terms of the influence of modulation frequency (cycles/octave), carrier bandwidth (octaves), and carrier frequency region (defined by lower and upper cutoff frequencies in Hz). Spectral modulation detection thresholds measured as a function of spectral modulation frequency result in a spectral modulation transfer function (SMTF). The general form of the SMTF is bandpass in nature, with a minimum modulation detection threshold in the region between 2 to 4 cycles/octave. SMTFs are not strongly dependent on carrier bandwidth (ranging from 1 to 6 octaves) or carrier frequency region (ranging from 200 to 12 800 Hz), with the exception of carrier bands restricted to very low audio frequencies (e.g., 200–400 Hz). Spectral modulation detection thresholds do not depend on the presence of random level variations or random modulation phase across intervals. The SMTFs reported here and associated excitation pattern computations are considered in terms of a linear systems approach to spectral envelope perception and potential underlying mechanisms for the perception of spectral features. [DOI: 10.1121/1.2382347]

PACS number(s): 43.66.Fe, 43.66.Ba, 43.66.Jh, 43.66.Lj, 43.66.Mk [JHG] Pages: 363–372

I. INTRODUCTION

Normal perception necessitates that sensory systems function as multidimensional pattern recognition devices. Representation of the visual world, for example, requires the perception of complex luminance or color patterns across visual or retinal space. Likewise, representation of the auditory world requires the perception of complex intensity patterns across audio frequency (the spectral envelope) and the resulting vibratory patterns across cochlear space. It is this spectral envelope that differentiates speech sounds such as "kit" and "cat," that allows one to estimate the elevation of sound sources such as a puzzling rattle in an automobile, and that gives a single note played by a clarinet and an oboe a distinctly different timbre. Although several psychophysical techniques have been used to study principles underlying spectral envelope perception, few provide a unifying framework that allows one to make predictions about the perception of novel spectral envelopes. The present study uses a spectral modulation detection task to systematically explore the perception of spectral features superimposed upon noise carriers across a broad range of relevant acoustic parameters including spectral modulation frequency (cycles/octave), noise carrier bandwidth (octaves), and noise carrier frequency region (Hertz). A spectral modulation detection threshold is determined by varying the modulation depth (peak to valley difference in dB) required to discriminate a noise with a flat spectral envelope (unmodulated stimulus) from a noise with a sinusoidal pattern of peaks and valleys across frequency (modulated stimulus). The goal of this study is to provide a general characterization of spectral envelope perception by young, normal-hearing listeners and to determine the dependence, if any, of spectral envelope perception on various stimulus parameters.

The spectral modulation detection task employed here is similar to the technique introduced by Bernstein and Green (1987a) as a variation on the traditional spectral profile analysis paradigm. In that study, the standard stimulus was an equal-amplitude, multi-component tonal complex that formed a flat spectral envelope and the signal to be detected included a sinusoidal pattern of increments and decrements to the standard spectral components. A detection threshold was determined by varying the size of the corresponding increments and decrements. By systematically varying the spacing of the peaks and valleys created by the signal, they could probe the just detectable spectral modulation depth as a function of the frequency of the spectral envelope, generating a spectral modulation transfer function (SMTF).

More recently, several investigators have used this technique with spectral modulation superimposed upon tonal complexes spanning two to nearly five octaves (e.g., Bernstein and Green, 1987a, b; 1988; Summers and Leek, 1994;

[a)]Electronic mail: david_eddins@urmc.rochester.edu

Amagai *et al.*, 1999; Chi *et al.*, 1999). When considered in terms of SMTFs, the results from these studies indicate a minimum spectral modulation detection threshold between two and four cycles/octave with a substantial increase in modulation detection threshold for higher modulation frequencies. The threshold increase for high modulation frequencies may be attributed primarily to limited frequency resolution and the associated reduction in spectral contrast as the auditory system integrates across more closely spaced cycles of modulation. The form of the function below the minimum threshold is not strictly determined by frequency resolution. For example, Summers and Leek (1994) and Amagai *et al.* (1999) observed a progressive increase in modulation depth at threshold as modulation frequency decreased below two cycles/octave. This result would not be predicted based on limited frequency resolution alone and may well depend upon specific experimental variables yet to be described.

Previous studies of spectral modulation detection have used relatively broadband carriers and therefore it is not known whether spectral modulation detection depends on the frequency region of the stimulus (i.e., cochlear place) or the range over which the spectral modulation extends (i.e., bandwidth of the carrier or cochlear distance). Ultimately, the degree to which measures of spectral modulation detection are related to or can be used to predict spectral shape perception in general cannot be assessed until the dependence of spectral modulation detection upon these parameters is known.

Studies of spectral shape perception using the profile analysis technique have focused on the influence of frequency region on spectral shape perception (e.g., Green, 1988; Green and Mason, 1985; Green *et al.*, 1987, 1995). These studies have used the more traditional stimulus paradigm consisting of a standard stimulus composed of a logarithmically spaced, multi-component complex and a signal stimulus created by incrementing one component of the complex. Each of these studies have shown that thresholds are lowest (i.e., better spectral shape discrimination) for standard stimuli centered near 1000 Hz (e.g., Green *et al.*, 1995) and signal increments near 1000 Hz (e.g., Green, 1988; Green and Mason, 1985; Green *et al.*, 1987). As the standard complex or signal increment is moved to lower or higher frequency regions, threshold in the profile analysis paradigm increases. This frequency effect is consistent across studies, despite the fact that other parameters have varied, such as number of components or component density (i.e., multiple components per critical band versus multiple critical bands per component). To the extent that both traditional measures of profile analysis and the current spectral modulation detection task are related to each other and to spectral envelope perception in general, one might predict that spectral modulation detection will vary with frequency region in a manner similar to signal detection in the profile analysis task. Conversely, the stimuli used in the traditional profile analysis and spectral modulation detection tasks are quite different and therefore the pattern of thresholds obtained with the two techniques as a function of any given parameter may not be related in a simple manner.

Several aspects of the spectral modulation detection task are attractive from practical as well as theoretical perspectives. First, spectral modulation detection is a relatively easy task for naïve listeners and informal testing indicates that little practice is required to achieve consistent performance on the task. Second, simple sinusoidal spectral envelopes can easily be used to systematically map the perception of spectral features over a range of relevant parameters. Third, following a linear systems approach, spectral modulation detection thresholds may be used to provide a general characterization of spectral envelope perception from a limited set of measurements. Thus, it may be possible to predict the internal representation of arbitrary spectral envelopes by treating the SMTF as a simple transfer function (see Viemeister, 1979, for a review of this general approach applied to the study of temporal resolution). For example, the spectral envelope of a given sound may be expressed as the magnitude spectrum (magnitude versus log frequency) resulting from a Fourier transformation of the original wave form. A second-order fast Fourier transform (FFT) then represents the same sound in the spectral modulation frequency domain. The second-order magnitude spectrum reveals the relevant spectral modulation features of the sound. For example, analyses of English and Dutch vowels (van Veen and Houtgast, 1985; Liu and Eddins, 2004) and head related transfer functions (Qian and Eddins, 2006) indicate that the distinguishing spectral features are between 0.1 and 2.0 cycles/octave. If the linear systems approach is valid in this case, and it may not be, then the psychophysical SMTF may be applied to the modulation spectrum of any arbitrary sound to predict the internal representation of the spectral envelope of that stimulus. This general approach has been used extensively to characterize the perception of the temporal envelope of acoustic signals and the temporal and spatial details of visual stimuli. The extent to which this approach is valid depends on many factors, including a reasonable selection of the basis function. In the present experiment, log amplitude (dB) and $\log_2$ frequency scales were chosen as a first approximation to the tonotopic organization and intensity coding of the auditory system. Although neither of these parameters is strictly accurate, it is unclear at this point how critical is the selection of these parameters. Ultimately, the utility of the spectral modulation task may lie in the ease of implementation and the ability to directly relate the experimental stimuli to the spectral features of natural stimuli rather than adherence to a strict linear systems approach.

Many stimulus parameters potentially influence spectral modulation detection. The present study uses the spectral modulation detection paradigm to evaluate the dependence of spectral envelope perception on spectral modulation frequency, carrier frequency region, and carrier bandwidth. Knowledge of the extent to which modulation detection depends on these parameters will influence the theoretical understanding of spectral shape perception and will largely determine the general utility of this technique, as a strong dependence on either carrier bandwidth or frequency region would increase the number of measurements required to provide a comprehensive assessment of spectral envelope perception.

TABLE I. Bandwidth and upper-cutoff frequency specifications for the 12 noise carrier conditions.

Bandwidth	Upper cutoff frequency (Hz)
Six octave	12 800
Three octave	1600, 12 800
Two octave	800, 3200, 12 800
One octave	400, 800, 1600, 3200, 6400, 12 800

II. METHOD

A. Listeners

Three listeners ranging in age from 22 to 28 years participated in this experiment. All listeners were inexperienced in psychoacoustic listening tasks. Each had normal middle ear function (based on screening tympanograms) and normal hearing sensitivity (<20 dB HL) from 250 to 8000 Hz (re: ANSI, 1996) and from 8000 to 13 000 Hz (re: laboratory norms). Listeners were paid an hourly wage for their participation. The data collection for each listener was completed in a series of approximately 16 to 18 two-hour sessions.

B. Conditions

Spectral modulation detection was measured by determining the modulation depth (peak-to-valley difference in dB) required to discriminate a noise carrier with flat spectrum from that with a sinusoidally modulated spectrum. The modulator was sinusoidal on a logarithmic frequency ($\log_2$) and logarithmic amplitude (dB) scale. For each carrier condition, thresholds were obtained for seven to ten modulation frequencies between 0.25 and 10.0 cycles/octave, with the requirement that at least 1.5 cycles of modulation span the nominal passband. The main experiment included 12 carrier conditions in which bandwidth (in octaves) and frequency region (between 200 and 12 800 Hz) were manipulated as experimental parameters, as shown in Table I.

The modulation phase was randomly selected from a uniform distribution (0 to 2π rad) in an effort to minimize the possibility that detection was based on local changes in intensity. As a result, spectral peaks and valleys were randomly located along the audio frequency axis. To further discourage the use of local intensity differences as cues for discrimination, the overall level of the stimulus was randomly varied over a 10 dB range about the nominal presentation level. This level variation is referred to as ROVE following Green *et al.* (1983). Randomization of overall level and modulation starting phase are both control parameters designed to render local spectral cues less useful and to encourage a comparison of intensity across a broad audio-frequency range. For example, it is possible that the cue for the presence of spectral modulation could be a change in the stimulus intensity within a narrow audio-frequency region across presentation intervals. In this case, the task would depend more on local intensity resolution than the perception of global spectral envelope features. Randomizing the modulation starting phase renders the relevant local frequency regions unpredictable. Together, these two manipulations encourage the listener to focus on the pattern of intensity across

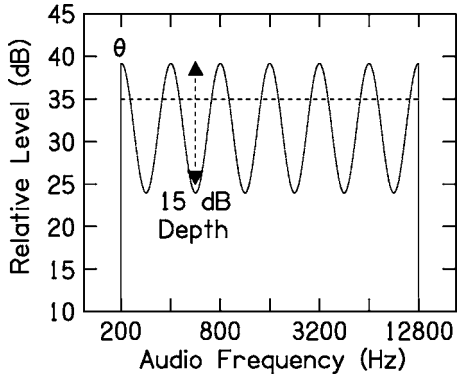


FIG. 1. Stimulus schematic illustrates a noise carrier with a nominal bandwidth from 200–12 800 Hz. The horizontal dashed line illustrates the flat-spectrum standard stimulus and the solid curve represents sinusoidal spectral modulation of the signal stimulus. The starting phase of modulation is 0 rad, the modulation depth is 15 dB (peak-to-valley difference), and the modulation frequency is 1 cycle/octave (six cycles spanning six octaves).

frequency (i.e., the spectral envelope) rather than local stimulus cues. Figure 1 illustrates the stimulus in schematic form for a six octave carrier band (200–12 800 Hz). The dashed horizontal line represents the flat spectrum of the standard stimulus and the solid line illustrates sinusoidal spectral modulation superimposed on the flat spectrum with a modulation frequency of 1 cycle/octave (i.e., six cycles over six octaves), a modulation starting phase of 0 rad (cosine phase), and a modulation depth of 15 dB.

Furthermore, to assess the dependence of the SMTF on local and/or global intensity changes, the present study examined spectral modulation detection in the presence or absence of roving level and with and without randomization of modulation starting phase. To evaluate the effect of random phase on spectral modulation detection, additional conditions were included in which the modulation phase was fixed at 0 rad (cosine phase) relative to the nominal low-frequency cutoff of the carrier. To evaluate the possible effects of roving level, a no-ROVE condition was included in which the modulation phase was random from interval to interval but the overall stimulus level was fixed across intervals and trials. In addition, a no-ROVE plus fixed-phase condition was included to examine any differences in threshold when the availability of local intensity cues was at a maximum.

For all subjects, testing began with the 200–12 800 Hz carrier condition. The remaining 11 carrier conditions of the main experiment were run in a different random order for each subject. For a given carrier condition, the various modulation frequencies were tested in a different random order for each subject and thresholds for all modulation-frequency conditions were collected before proceeding to a new carrier condition. The three additional conditions (no-ROVE condition, fixed-phase condition, and no-ROVE, fixed phase condition) were completed in sequential order following the completion of the main experiment. Within each of these conditions, the order of testing across modulation frequency or condition number was random across listeners.

C. Stimuli

All stimuli were generated using a digital array processor (TDT AP2) and a 16-bit D/A converter with a sampling period of 24.4 μs (40 983 Hz). The sequence of operations in the stimulus generation was as follows. Sinusoidal spectral modulation was generated by filling an 8192-point buffer with a sinusoid computed on a logarithmic frequency axis, with the appropriate modulation frequency (cycle per octave), phase, and modulation depth (peak-to-valley difference in dB). The modulator was converted to linear units and multiplied by a second buffer filled with random numbers from a Gaussian distribution. This buffer was then multiplied by a third 8192-point buffer filled with values corresponding to the magnitude response of a Butterworth filter (−32 dB/octave) with condition-specific lower and upper cutoff frequencies. This sequence of operations was repeated, forming a complex buffer pair. A real inverse FFT was performed on the complex buffer pair resulting in the desired 400-ms wave form with a spectral density (spacing between frequency components) of 2.5 Hz. The wave form was then shaped with a 10-ms $\cos^2$ window and scaled to the desired presentation level.

In conditions without roving level, the stimulus presentation level was determined in a manner that minimized the interval-by-interval variations in overall level. First, the rms level of each wave form (modulated and unmodulated) was normalized to a fixed value. Next, a programmable attenuator was set to a value that produced a spectrum level of 35 dB SPL for the standard (unmodulated) stimulus. The same attenuation setting was used for the modulated signal so that the unmodulated standard and modulated signal for a given condition had the same overall level. The resulting spectrum level for the signal stimulus depended on the actual modulation depth, with a maximum peak difference of about 4 dB relative to the unmodulated standard. Finally, in the ROVE conditions, the overall level of the stimulus was adjusted on an interval-by-interval basis by either increasing or decreasing the attenuator setting by a value selected randomly (in 1-dB steps) from a uniform distribution with a mean of 0 dB and a range of 10 dB.

D. Procedures

Stimuli were presented and responses obtained using a cued, two-interval, two alternative forced-choice method. The standard stimulus (flat spectrum) was always presented in the first stimulus interval as an anchor or reminder stimulus. A second standard stimulus and the signal stimulus were randomly presented in the following two intervals. This procedure was adopted following a previous spectral modulation detection experiment using a standard two-interval task in which listeners reported that they could hear the difference between the standard and signal stimuli but could not easily identify the signal interval as such. Spectral modulation detection thresholds were estimated using an adaptive tracking algorithm with a three-down, one-up rule, corresponding to the 79.4% point on the psychometric function (Levitt, 1971). Thus, following a sequence of three correct responses, modulation depth (peak-to-valley difference in dB) was decreased, and following one incorrect response, modulation depth was increased. An initial step size of 2 dB was used for the first three reversals in the adaptive track, after which the step sized was fixed at 0.4 dB for the remainder of the 60-trial block. Threshold for a given track was based on the average of the modulation depths corresponding to each of the last even number of reversals in the adaptive track, excluding the first three reversals. Final thresholds were based on the average of three such threshold estimates.[1] The stability of thresholds across these three estimates is indicated by the within-subject standard error of the mean which was 0.37 dB when averaged across the three listeners and 88 conditions of the main experiment (see the error bar in the upper left corner of the upper left panel of Fig. 2). Each observation interval was 400 ms in duration and successive intervals were separated by a 400-ms interstimulus interval. Stimulus intervals were marked with light emitting diodes (LEDs) on a button box, listeners responded by button press corresponding to the second or third stimulus interval, and correct feedback was provided via LEDs corresponding to intervals two and three. Listeners were seated in a double-walled sound attenuating chamber and stimuli were delivered via ER-2 (Etymotic) insert earphone to the left ear.

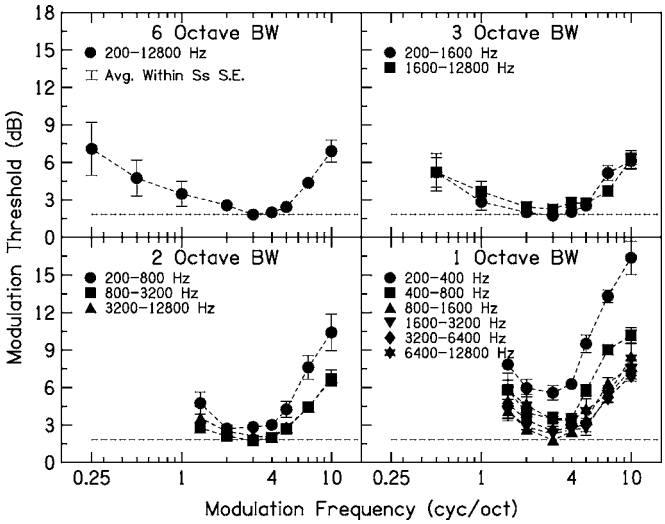


FIG. 2. Spectral modulation transfer functions (SMTFs) averaged across three normal-hearing subjects. Modulation depth at threshold (peak-to-valley difference in dB) is shown as a function of modulation frequency (cycles/octave). Carrier frequency region is indicated by the different symbols. Each panel represents data for a different carrier bandwidth. Error bars represent standard error of the mean. To facilitate comparisons of sensitivity across bandwidth (panels), the dashed line indicates the minimum contrast threshold for the six-octave carrier.

III. RESULTS

A. Spectral modulation transfer functions

Spectral modulation detection thresholds averaged across the three listeners are shown in Fig. 2 for each of the conditions in the main experiment. Panels represent different carrier bandwidths. Within a panel, the symbols indicate different carrier frequency regions. Spectral modulation detection thresholds ranged from 1.74 to 16.37 dB across the various bandwidth, frequency region, and modulation frequency

conditions. Error bars about each symbol indicate the standard error of the mean. The data for the six-octave carrier (upper left panel) illustrate the characteristic bandpass form of the SMTF for all conditions, with a minimum modulation detection threshold at modulation frequencies between 2 and 4 cycles/octave.

The different functions within a given panel in Fig. 2 correspond to different carrier frequency regions. For a three-octave carrier (upper right panel), there is no change in the shape or the sensitivity of the SMTF with carrier frequency region. For the two-octave carrier conditions (lower left panel), thresholds are slightly elevated for the lowest carrier frequency region (200–800 Hz) relative to the two higher carrier frequency regions. Likewise, for the one-octave carrier conditions (lower right panel), thresholds are quite similar for the mid- and high-frequency carriers but are elevated substantially for the 200–400 Hz carrier and slightly for the 400–800 Hz carrier.

Comparisons across panels in Fig. 2 reveal that the shape and sensitivity of the SMTF is quite robust to changes in carrier bandwidth. Overall, the shape of the function changes little with carrier bandwidth from one to six octaves. The dashed line in each panel represents a threshold of 1.83 dB corresponding to the minimum threshold obtained for the six-octave carrier (at three cycles/octave). For each bandwidth, the most sensitive SMTF intersects this line, indicating that sensitivity to modulation is not strictly dependent on carrier bandwidth. A two-way repeated measures analysis of variance (ANOVA) with spectral modulation frequency and bandwidth as parameters indicated that the difference among the mean SMTFs across bandwidth was significant ($F_{5,15}=31.448$, $p<0.001$). However, a post-hoc Tukey Test revealed that the source of variation arose from a single one-octave condition (200–400 Hz) that was significantly different from all other bandwidth conditions. No other bandwidth conditions were significantly different from each other. Overall, these results indicate little change in spectral modulation detection with carrier bandwidth or with carrier frequency region with the exception of carriers restricted to very low audio-frequency regions. This latter variation may reflect changes in frequency selectivity with audio frequency. This possibility is considered in more detail in the Sec. IV.

Close inspection of the standard error bars in Fig. 2 reveals minimal variability across listeners for the midmodulation frequencies and maximal variability for the lowest and highest modulation frequencies. The most variability was observed for the six-octave carrier condition. The upper panel of Fig. 3 shows the individual data for this condition with bars representing the standard error of the mean. Clearly the greatest variability across subjects occurred at the low modulation frequencies, with thresholds spanning a range of 7 dB. In other carrier bandwidth conditions, the across-subject variability was somewhat less than for the 200–12 800 Hz condition but was always greatest for the lower modulation frequencies. In general, thresholds at 0.25 cycles/octave were consistently highest for *S*3 and lowest for *S*1 except in the two-octave carrier conditions, where thresholds were lowest for *S*1 and similar for *S*2 and *S*3. To better determine

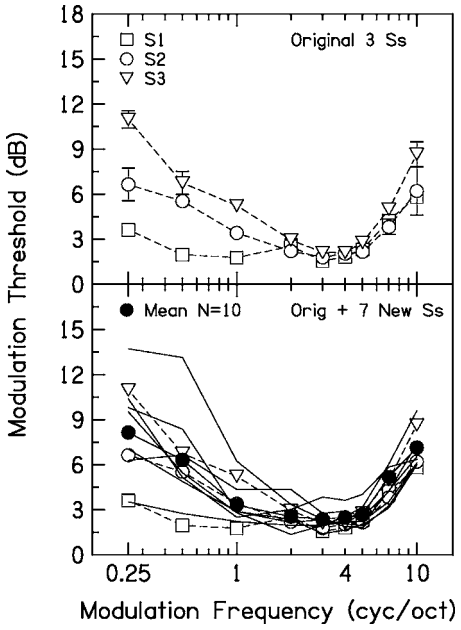


FIG. 3. Individual differences in spectral modulation detection. Upper panel shows individual SMTFs for the three original subjects in the 200–12 800 Hz carrier condition. Lower panel shows the same data (symbols) along with the data from seven new listeners (solid lines). The solid circles represent the mean thresholds across all ten listeners.

whether or not the three subjects were representative of the "typical" listener, seven additional subjects were recruited using the same subject selection criteria met by the original subjects. Modulation detection thresholds were collected for the six-octave carrier conditions and the results are shown in the lower panel of Fig. 3, where thresholds for the seven new listeners are shown by the solid lines and thresholds for the three original listeners are shown by the open symbols. The filled circles represent the mean across all ten listeners. Although the SMTFs for each listener show the characteristic bandpass shape, the range of thresholds across listeners clearly is greatest for the lowest modulation frequencies. Interestingly, for the lower modulation frequencies, the thresholds for *S*1 and *S*3 represent the upper and lower ends of the distribution of thresholds and thresholds for *S*2 closely mimic the mean thresholds for the group.

For the conditions reported above, the phase of the sinusoidal spectral modulation was randomized across presentations and the overall level of the stimulus was randomized over a 10 dB range. Both of these manipulations were included in an effort to render the use of local intensity cues a less reliable listening strategy in the modulation detection task. To evaluate the potential influence of random modulation phase on spectral modulation detection threshold, an additional condition was included in which the starting phase of the sinusoidal spectral modulation was fixed at 0 rad relative to the low frequency cutoff of the nominal passband (800–3200 Hz in this case). The overall level was randomly varied from interval to interval, as in the main conditions. As shown in the upper panel of Fig. 4, no significant differences in modulation detection were observed for fixed (circles) versus random (squares) modulation phase conditions.

Phase effects were also assessed at a single modulation

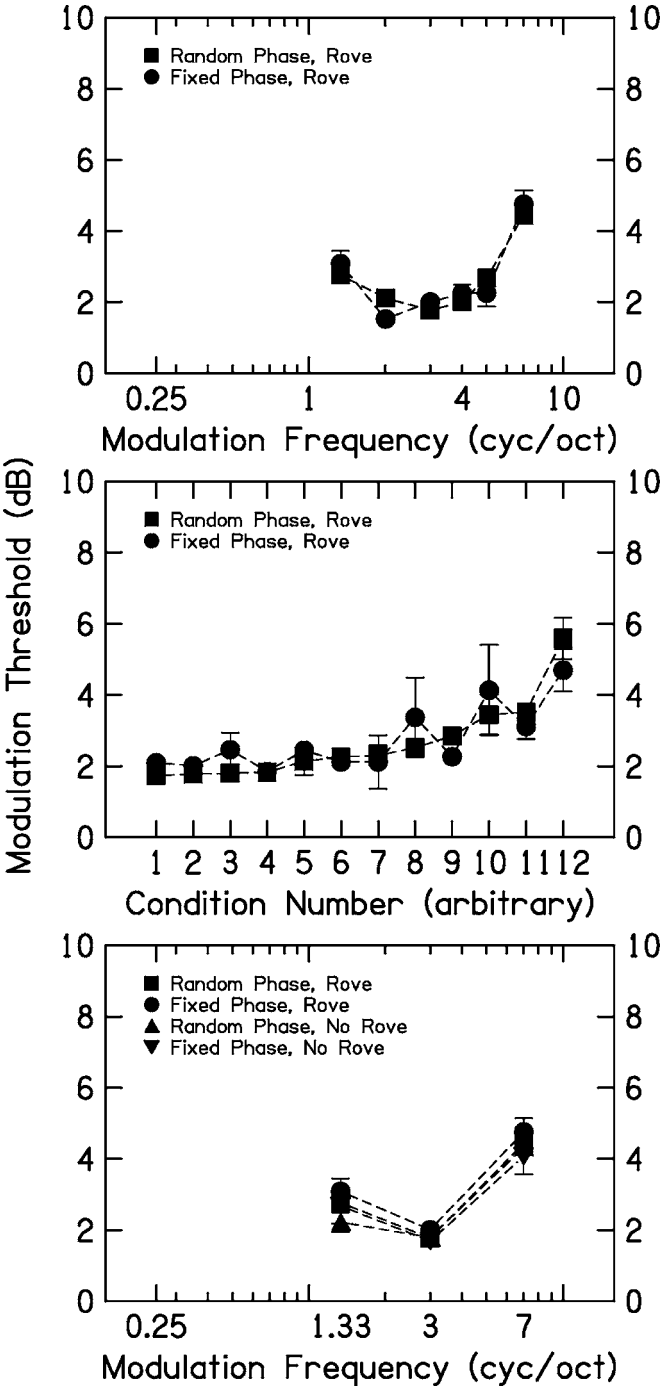


FIG. 4. Influence of spectral modulation phase. Upper panel shows SMTFs with random (squares) or fixed (circles) starting phase. In both cases, the overall level was randomized over a 10 dB range about the nominal stimulus level. Middle panel shows modulation detection thresholds for 3 cycles/octave in each of the 12 different carrier frequency/bandwidth conditions with random (squares) or fixed (circles) modulation phase. Lower panel shows abbreviated SMTFs for the four combinations of random level variation (ROVE or no-ROVE) and phase (random or fixed).

frequency (three cycles/octave) for each of the 12 bandwidth/audio-frequency combinations of the main experiment. The results are shown in the middle panel of Fig. 4. Again it is evident that the starting phase of the sinusoidal spectral modulation has little impact on modulation detection.

For the conditions reported above, the overall stimulus level was varied over a 10-dB range about the nominal spectral level of 35 dB. In conditions where the starting phase of the sinusoidal modulation is random *or* the level of the stimulus is varied within a trial, use of local intensity cues as the basis for detection is inhibited to some extent. To fully allow for the use of local intensity cues, the stimulus must have both fixed phase and static level (no-ROVE). To better gauge the extent to which spectral modulation detection could be based on local (in a narrow frequency range) versus global (across a broad frequency range) intensity cues, modulation detection was measured in conditions that included the four possible combinations of roving level and random modulation phase. The carrier bandwidth was 800–1600 Hz (two octaves) and thresholds were obtained for modulation frequencies of 1.33, 3, and 7 cycles/octave.

As shown in the lower panel of Fig. 4, thresholds varied little across the four different conditions. Nevertheless, it is possible that overall level cues could be robust enough to overcome the random level and phase variations imposed here. To better evaluate this possibility, expected thresholds for a signal increment given a detection strategy based on changes in the overall stimulus level produced by the signal were estimated from the computations reported by Green (1988, pp. 19–21). Accordingly, given the 10-dB rove used here, one would expect a threshold of 3.5 dB based on overall level cues. Thresholds for one of the six modulation frequencies shown in the upper panel of Fig. 4 exceed this value. However, the spectrally modulated signal produces both increments and decrements in level relative to the standard stimulus. If it is assumed that the level comparisons are local in nature, such that the listener focuses on a frequency region corresponding to a spectral peak or spectral valley in the signal stimulus and compares that level to the corresponding level in the standard stimulus, then the change in level does not correspond to the peak-to-valley contrast but the signal-to-standard contrast at that audio frequency. For the spectral modulation task here, this value corresponds to roughly 0.2 times the peak-to-valley contrast at threshold for spectral peaks and approximately 0.8 times the peak-to-valley contrast for spectral valleys. Thus, thresholds based on changes in peak level would be approximately 17.5 dB and changes based on valley level, assuming no filling in of the valleys as a result of limited frequency selectivity, would be 4.375 dB. Given that there likely is some smoothing of the valleys, it appears unlikely that overall level variations can explain the thresholds for any of the ROVE conditions in Fig. 4.

With respect to the random phase, no ROVE conditions, one can use the same analyses as described above to ascertain the thresholds expected based on a change in level. As a result of the random modulation phase, the level at any fixed audio frequency could vary from the level of a spectral peak to the level of a spectral valley on a trial-by-trial basis. Thus, the "ROVE" in local level imposed by random phase was equal to modulation detection threshold in dB, which ranged from 1.5 to 5.2 dB in the conditions shown in Fig. 4. Predicted thresholds based solely on changes in overall level range from approximately 0.7 to 5.5 dB. Again assuming local comparisons of level, peak detection would lead to expected thresholds between 3.5 and 27.5 dB and valley detection without smoothing would lead to expected thresholds

between 0.875 and 6.875 dB. Finally, combined random phase and random level variations equate to roving levels from 11.5 to 15.2 dB, depending on the actual threshold value. Predicted thresholds based on peak detection would correspond to values between 19.5 and 25.5 dB and thresholds based on valley detection, assuming no smoothing, would correspond to values between 4.875 and 6.375 dB. Overall, these results are consistent with the notion that the spectral modulation detection thresholds of the main experiment were based upon a global comparison of intensity across frequency and that listeners can and do focus on global intensity patterns to encode spectral patterns.

It should be noted, however, that the fixed-phase and static level conditions reported above were tested after the listeners had completed the random phase conditions random level of the main experiment (see Fig. 2, noting the change in *y*-axis range). Thus, the listeners had considerable experience listening to stimuli where local cues were unreliable. While it is possible that this experience influenced the degree to which the listeners were likely to adopt a local listening strategy, the minimum detectable modulation depth of about 1.5–2.5 dB reported here is close to the minimum detectable change in level for a broadband noise in an intensity discrimination task (e.g., 0.5 to 3.0 dB; Miller, 1947; Bos and de Boer, 1966). Since intensity discrimination for noise stimuli is inversely related to noise bandwidth, intensity discrimination for a broadband noise indicates that there is little possibility of improvement in threshold beyond the spectral modulation detection thresholds obtained in these conditions.

IV. DISCUSSION

The form of the spectral modulation transfer functions reported here for noise carriers is similar to SMTFs measured by previous investigators using tonal complexes (e.g., Bernstein and Green, 1987a, b; Summers and Leek, 1994; Chi *et al.*, 1999), both in terms of sensitivity to modulation and the pattern of thresholds with modulation frequency. This similarity indicates that spectral modulation detection is not strongly dependent on the nature of the stimulus carrying the spectral envelope, at least for carriers with moderately high density in the audio-frequency domain. Furthermore, the present results indicate that the form of the SMTF depends little on changes in bandwidth (from one to six octaves) and carrier frequency region (between 400 and 12 800 Hz). Spectral modulation detection was significantly poorer for the lowest one-octave condition (200–400 Hz) than the other carrier conditions. In this frequency region, hearing sensitivity is considerably worse than at mid-to-high frequency regions and thus one possibility is that reduced audibility or loudness might have led to higher thresholds in the 200–400 Hz condition. To test this possibility, one subject was retested on this condition with the stimulus level increased by 15 dB; however the resulting spectral modulation detection thresholds were not significantly different across levels. Furthermore, intensity resolution is not strongly dependent upon frequency region (e.g., Jesteadt *et al.*, 1977) so there is no reason to believe that variations in intensity across frequency are encoded less precisely at low- than high-audio frequencies. It may be that the perception of global spectral shape simply is not as good at low audio frequencies as it is at higher audio frequencies. Using the spectral envelope of a vowel as an example, perhaps the auditory system focuses on individual spectral peaks at low-audio frequencies, corresponding to individual harmonics, and more global spectral shape at higher-audio frequencies, which would provide information about the pattern of peaks and valleys, their relative amplitudes, and distances between peaks, etc.

Each of the SMTFs shown in Fig. 2 show a minimum between two and four cycles/octave. Above this frequency, it is likely that limited frequency selectivity results in a smoothing of the spectral modulation and thus provides a lower limit to modulation detection threshold. Below this minimum, however, spectral smoothing due to the limits of frequency selectivity predict a gradual decrease in spectral modulation threshold with decreasing modulation frequency, whereas the data show a gradual increase in threshold with decreasing modulation frequency. It is over this modulation frequency region where the spectral modulation detection paradigm can provide information about the representation of spectral contrast not available from traditional measures of frequency selectivity. To further evaluate the role of frequency selectivity in spectral modulation detection, excitation patterns were computed based on the model first proposed by Moore and Glasberg (1987) and later modified by Moore and Glasberg (2004) in the context of their loudness pattern model. This model assumes that the auditory periphery can be represented by a bank of overlapping bandpass filters, the characteristics of which were determined on the basis of the results of several investigations of auditory filter shape using the notched noise masking paradigm (e.g., Patterson, 1976). With the same stimulus generation software used in the main experiments, average stimulus spectra were computed based on 100 signal and standard samples for each of the 88 conditions in the main experiment. To reduce the number of possible computations, the modulation phase was fixed at 0 deg relative to the nominal low-pass cutoff frequency and the stimulus level was fixed rather than randomly varied. The modulation depth for each signal condition was based on the average threshold across the three original subjects. As described by Moore and Glasberg, the excitation pattern model included a middle ear correction factor. Absolute thresholds were set to a default value of 0 dB HL and excitation patterns were estimated over the range of 3 to 40 equivalent rectangular bandwidths (ERBs) or roughly 90 to 16 800 Hz. The internal representation of spectral modulation was indexed by subtracting the excitation pattern in response to the standard from the excitation pattern in response to the signal and then computing the maximum modulation depth over the range of audio frequencies extending 1/3 octave below and above the nominal lower and upper cutoff frequencies of the carrier condition under study.

The results of these computations are shown in Fig. 5 in a manner parallel to the SMTFs shown in Fig. 2. If the SMTF were strictly determined by frequency selectivity, then the output of the excitation pattern computations would yield a horizontal function relating the change in excitation pro-

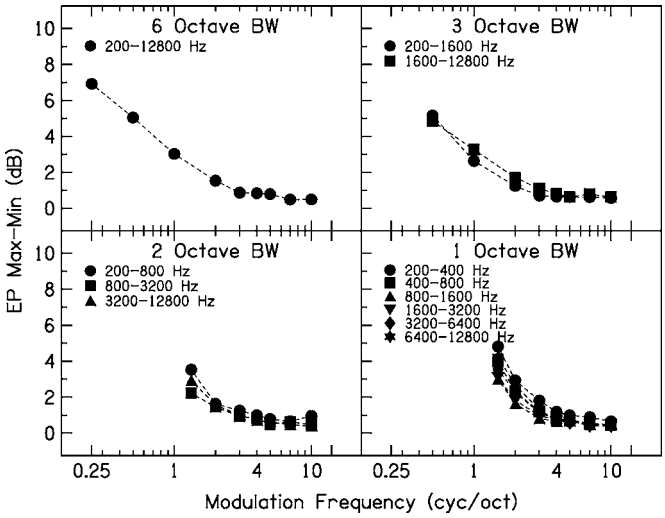


FIG. 5. Excitation pattern differences as a function of modulation frequency for the four carrier bandwidth conditions (panels) and the various carrier frequency conditions (symbols within a panel). See text for computational details.

duced by spectral modulation-to-modulation frequency, reflecting the constant change in modulation depth required to detect the presence of spectral modulation. Indeed such a horizontal function is achieved for high-modulation frequencies, where one would expect frequency selectivity to dominate. The increase in modulation depth maintained in the excitation pattern as modulation frequency decreases, however, clearly indicates that the spectral modulation detection paradigm is measuring something in addition to frequency selectivity. Indeed, one could obtain a rough estimate of the critical band by fitting a negatively sloping and a horizontal line to the excitation pattern difference values and estimating the breakpoint in the function. By eye, this would be roughly three to four cycles/octave, corresponding to a bandwidth of roughly 1/4 octave in width.

In addition to highlighting the manner in which the SMTF captures spectral shape perception at low-modulation frequencies, the functions in Fig. 5 also help to explain differences in the SMTFs for the lowest-carrier frequency regions. In particular, the markedly higher-modulation thresholds for the 200–800 Hz carrier relative to other two-octave carriers, shown in the lower left panel of Fig. 2, do not correspond to markedly higher changes in the excitation pattern, as shown in the lower left panel of Fig. 5. Thus, for two-octave carrier conditions, one may conclude that threshold differences across audio frequency are due to changes in frequency selectivity with audio frequency. For the one-octave carriers, the relatively large range of threshold values for high-modulation frequencies is not evident in the excitation pattern differences (compare lower right panels of Figs. 2 and 5), though there remains some variation in excitation pattern differences across the one-octave carrier conditions. An analysis of variance computed on the excitation pattern differences for the one-octave carrier conditions revealed a significant effect of carrier frequency region ($F_{5,30}=8.787$, $p<0.001$) and a post-hoc Tukey (HSD) test revealed that the 200–400 Hz condition was significantly different from all other conditions, whereas none of the other conditions were significantly different from each other. These results support the conclusion that the threshold differences across various carrier conditions reflect the changes in frequency selectivity across audio frequency with the exception of the 200–400 Hz carrier condition. In this condition, the estimates of frequency selectivity incorporated in the excitation pattern computations do not account for the observed threshold differences.

While the SMTFs reported here indicate that the perception of spectral envelope features depends little on the carrier bandwidth or carrier frequency region, these results differ substantially from the consistent frequency effects reported for the profile analysis task. This relationship between threshold and signal frequency has often been described as a "bowl" effect, in which thresholds are lowest for midfrequencies (near 1000 Hz) and are progressively higher for lower- and higher-signal frequencies (e.g., Green and Mason, 1985; Green *et al.*, 1987). Such a discrepancy may well indicate that the two tasks (profile analysis and spectral modulation perception) actually reflect different perceptual processes. Indeed, the fact that the spectral modulation detection task requires very little familiarity or practice to achieve consistent performance while the profile analysis task often requires considerable training supports the notion that the two tasks differ in some fundamental way that is not well understood.

The lack of an effect of bandwidth and frequency region on spectral envelope perception has important ramifications for understanding the perception of spectrally complex stimuli. For example, it suggests that transposition of spectral features across audio frequency should have little effect on the perception of those features. Likewise, it is consistent with the hypothesis that the spectral features inherent in speech stimuli (such as vowels) and spectral features imposed upon broadband stimuli by the pinnae [and other factors influencing head related transfer functions (HRTFs)] might be processed in qualitatively and quantitatively similar ways.

On practical and theoretical levels, the spectral modulation detection task offers potential predictive power that may prove useful in the future. Specifically, if one adopts a linear systems approach, then the SMTF may be considered as a filter function, that when applied to input spectra, represents the internal representation of a given spectral envelope. Of course, the auditory system is known to be nonlinear in many ways. Nevertheless, a linear systems approach is adopted routinely by most auditory scientists and clinicians when studying or characterizing audition. In general, if a system can be considered linear and time invariant under specific conditions, then the response of that system may be characterized by a transfer function, which, in theory, will allow the output of the system to be predicted for any possible input. In sensory systems, this function is known as the modulation transfer function (MTF) or contrast sensitivity function (CSF) and reflects a mapping of the system output relative to the input over the sensitivity range of the system.

Linearity assumes (1) additivity (superposition): The response of a complex signal can be predicted from the algebraic sum of responses to simple stimuli (e.g., sinusoids);

and (2) homogeneity: the system is uniform or position invariant over a specific dimension (e.g., cochlear place). Although many sensory systems are nonlinear in numerous ways, a growing body of evidence suggests that such a linear systems approach captures many of the most relevant properties of sensory systems. In the visual system, this has been clearly demonstrated in behavioral and physiological studies of sensitivity to luminance (e.g., Davidson, 1968;—behavioral; De Valois *et al.*, 1974—physiological); chromaticity (van der Horst and Bouman, 1969—behavioral; De Valois *et al.*, 1966; Thorell *et al.*, 1984—physiological), and temporal variations in spatial gratings (e.g., Kelly, 1975—behavioral; McLean and Palmer, 1994—physiological). These studies indicate that such CSFs represent processing by an array of band-pass channels tuned to a best frequency. In the auditory system, the linear systems approach is commonly applied in both behavioral and physiological studies of temporal resolution (e.g., Viemeister, 1979—behavioral; Langner, 1992—physiological), in studies of spectral shape discrimination (e.g., Hillier, 1991—behavioral; Shamma *et al.*, 1993—physiological), and in clinical practice via the audiogram. While such an approach cannot explain many aspects of sensory systems, its generality and broad applicability have proven extremely beneficial, and at a minimum have encouraged systematic analyses of complex systems. The similarity of SMTFs across frequency region provides a simple test of position invariance, one criterion that supports the use of a linear systems approach in the study of spectral envelope perception. The SMTFs for the one-, two-, and three-octave carriers are consistent with the property of position invariance, although the SMTF for the one-octave carriers at the lowest-carrier frequency region deviates from this rule. As mentioned in the Introduction, the stimuli used here provide only a rough approximation to the tonotopic nature of the auditory periphery, and would be more appropriate for a constant Q system. Most estimates of frequency selectivity, however, indicate that the relative auditory filter width increases at low-audio frequencies. Thus, a better test of position invariance might use an ERB, bark, or mel frequency scale rather than a $\log_2$ scale. Future work should also evaluate the principle of superposition as it might apply to spectral modulation.

A question that arises from the discussion above is whether or not the SMTF reflects a single filter characteristic or the combined output of multiple filters that differ in modulation frequency. Similar to auditory temporal and visual spatial perception, one hypothesis is that the SMTF reflects the combined output of multiple channels tuned to different modulation frequencies (and perhaps phases). Such channels might originate from families of cells with specific spectral or spectrotemporal receptive fields (see Chi *et al.*, 1999, for a more detailed discussion). To date, there is little behavioral evidence to support or refute the existence of spectral modulation channels in the auditory system. Regardless of the underlying mechanisms for spectral envelope perception, the present data provide a systematic description of spectral modulation detection across a broad range of relevant parameters that may serve as a basis for future investigations of the nature of spectral envelope perception.

V. CONCLUSIONS

Spectral modulation detection thresholds were measured as a function of spectral modulation frequency (cycles/octave) by superimposing sinusoidal spectral modulation on noise carriers. To limit the utility of local intensity cues, overall stimulus level and modulation starting phase were randomized from interval to interval. The resulting spectral modulation transfer functions were used to characterize spectral shape perception in young, normal-hearing listeners as a function of noise carrier bandwidth and audio-frequency region. The SMTFs were bandpass in nature, with maximal sensitivity to spectral modulation in the region of two to four cycles/octave and progressively worse sensitivity at lower and higher modulation frequencies. The form of the SMTFs and overall sensitivity to modulation were not strongly dependent on carrier bandwidth (from one to six octaves) or carrier frequency region (between 400 and 12 800 Hz). Sensitivity to spectral modulation was poorer for noise carriers at very low-audio frequencies (200–400 Hz). Excitation pattern computations confirmed that spectral modulation detection for high-modulation frequencies is limited by frequency selectivity; however, spectral modulation detection can be used to describe important aspects of spectral shape perception at low-modulation frequencies not captured by measures of frequency selectivity.

ACKNOWLEDGMENTS

Work supported by NIH NIDCD R01 DC04403 and NIA P01 AG09524. The authors wish to express their sincere appreciation for the helpful comments of two anonymous reviewers and the associate editor.

[1]Estimation of threshold based on three 60-trial blocks was based in part on a pilot experiment in which six 60-trial blocks were collected. That experiment included two carrier bandwidths (100–6400 and 200–12 800 Hz), ten modulation frequencies between 0.25 and 14 cycles/octave and three subjects inexperienced in psychoacoustic listening tasks. Stimulus generation was identical to the present study with the exception that the bandpass filter slopes were steeper (96 versus 32 dB/oct) and the nominal spectrum level was 40 dB. Threshold averages based on blocks 1 through 3 and 4 through 6 were compared across the three subjects and 20 conditions and revealed an increase in threshold of by a factor of 0.991 and a decrease in standard error by a factor of 1.178 from the first to the second three blocks. On this basis, it was decide that three blocks of trials provided a sufficiently accurate threshold estimate.

Amagai, S., Dooling, R. J., Shamma, S., Kidd, T. L., and Lohr, B. (**1999**). "Detection of modulation in spectral envelopes and linear rippled noises by budgerigars (Melopsittacus undulates)," J. Acoust. Soc. Am. **105**, 2229–2235.

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

Bernstein, L. R., and Green, D. M. (**1987a**). "The profile analysis bandwidth," J. Acoust. Soc. Am. **81**, 1888–1895.

Bernstein, L. R., and Green, D. M. (**1987b**). "Detection of simple and complex changes of spectral shape," J. Acoust. Soc. Am. **82**, 1587–1592.

Bernstein, L. R., and Green, D. M. (**1988**). "Detection of changes in spectral shape: Uniform vs. non-uniform background spectra," Hear. Res. **32**, 157–166.

Bos, C. E., and de Boer, E. (**1966**). "Masking and discrimination," J. Acoust. Soc. Am. **39**, 708–715.

Chi, T., Gao, Y., Guyton, M. C., Ru, P., and Shamma, S. (**1999**). "Spectro-temporal modulation transfer functions and speech intelligibility," J. Acoust. Soc. Am. **106**, 2719–2732.

Davidson, M. L. (**1968**). "Perturbation approach to spatial brightness interaction in human vision," J. Opt. Soc. Am. **58**, 1300–1309.
De Valois, R. L., Abramov, I., and Jacobs, G. H. (**1966**). "Analysis of response patterns of LGN cells," J. Opt. Soc. Am. **56**, 966–977.
De Valois, R. L., Morgan, H., and Snodderly, D. M. (**1974**). "Psychophysical studies of monkey vision. III. Spatial luminance contrast sensitivity tests of macaque and human observers," Vision Res. **14**, 75–81.
Green, D. M. (**1988**). *Profile Analysis: Auditory Intensity Discrimination* (Oxford University, New York).
Green, D. M., Dai, H., and Saberi, K. (**1995**). "*Shape discrimination of spectra with various bandwidths*," in Advances in Hearing Research, edited by G. A. Manley, G. A. Klump, C. Koppl, and H. Fastl (World Scientific, Singapore).
Green, D. M., Kidd, G., Jr., and Picardi, M. C. (**1983**). "Successive versus simultaneous comparison in auditory intensity discrimination," J. Acoust. Soc. Am. **73**, 639–643.
Green, D. M., and Mason, C. R. (**1985**). "Auditory profile analysis: frequency, phase, and Weber's law," J. Acoust. Soc. Am. **77**, 1155–1161.
Green, D. M., Onsan, Z. A., and Forrest, T. G. (**1987**). "Frequency effects in profile analysis and detecting complex spectral changes," J. Acoust. Soc. Am. **81**, 692–699.
Hillier, D. A. (**1991**). "Auditory processing of sinusoidal spectral envelopes," Dissertation, Sever Institute of Technology, Washington University, St. Louis, Missouri.
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.
Kelly, D. H. (**1975**). "Luminous and chromatic flickering patterns have opposite effects," Science **188**, 371–372.
Langner, G. (**1992**). "Periodicity coding in the auditory system," Hear. Res. **60**, 115–142.
Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.
Liu, C., and Eddins, D. A. (**2004**). "Spatial frequency modulation in vowel identification," J. Acoust. Soc. Am. **115**, 2631.
McLean, J., and Palmer, L. A. (**1994**). "Organization of simple cell responses in the three-dimensional frequency domain," Visual Neurosci. **11**, 295–306.
Miller, G. A. (**1947**). "Sensitivity to changes in the intensity of white noise and its relation to masking and loudness," J. Acoust. Soc. Am. **19**, 609–619.
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," Hear. Res. **28**, 209–225.
Moore, B. C. J., and Glasberg, B. R. (**2004**). "A revised model of loudness perception applied to cochlear hearing loss," Hear. Res. **188**, 70–88.
Qian, J., and Eddins, D. A. (**2006**). "Virtual sound localization using head related transfer functions modified in the spectral modulation frequency domain," Assoc. Res. Otolaryngol. Abs., 147.
Patterson, R. D. (**1976**). "Auditory filter shapes derived with noise stimuli," J. Acoust. Soc. Am. **59**, 640–654.
Shamma, S. A., Fleshman, J. W., Wiser, P. R., and Versnel, H. (**1993**). "Organization of response areas in ferret primary auditory cortex," J. Neurophysiol. **69**, 367–383.
Summers, V., and Leek, M. R. (**1994**). "The internal representation of spectral contrast in hearing-impaired listeners," J. Acoust. Soc. Am. **95**, 3518–3528.
Thorell, L. G., De Valois, R. L., and Albrecht, D. G. (**1984**). "Spatial mapping of monkey V1 cells with pure color and luminance stimuli," Vision Res. **24**, 751–769.
van der Horst, G. J. C., and Bouman, M. A. (**1969**). "Spatiotemporal chromaticity discrimination," J. Opt. Soc. Am. **59**, 1482–1488.
van Veen, T. M., and Houtgast, T. (**1985**). "Spectral sharpness and vowel dissimilarity," J. Acoust. Soc. Am. **77**, 628–634.
Viemeister, N. F. (**1979**). "Temporal modulation transfer functions based upon modulation thresholds," J. Acoust. Soc. Am. **66**, 1364–1380.

Effect of duration on the frequency discrimination of individual partials in a complex tone and on the discrimination of fundamental frequency

Hedwig E. Gockel[a)]
MRC Cognition and Brain Sciences Unit, 15 Chaucer Road, Cambridge CB2 2EF, UK

Brian C. J. Moore
Department of Experimental Psychology, University of Cambridge, Downing Street, Cambridge CB2 3EB, UK

Robert P. Carlyon
MRC Cognition and Brain Sciences Unit, 15 Chaucer Road, Cambridge CB2 2EF, UK

Christopher J. Plack
Department of Psychology, Lancaster University, Lancaster LA1 4YF, UK

(Received 10 May 2006; revised 25 September 2006; accepted 4 October 2006)

Thresholds for the discrimination of fundamental frequency (*F*0DLs) and frequency difference limens (FDLs) for individual partials within a complex tone (*F*0=250 Hz, harmonics 1–7) were measured for stimulus durations of 200, 50, and 16 ms. The FDLs increased with decreasing duration. Although the results differed across subjects, the effect of duration generally decreased as the harmonic number increased from 1 to 4, then increased as the harmonic number increased to 6, and finally decreased for the seventh harmonic. For each duration, *F*0DLs were smaller than the smallest FDL for any individual harmonic, indicating that information is combined across harmonics in the discrimination of *F*0. *F*0DLs predicted from the FDLs corresponded well with observed *F*0DLs for the 200- and 16-ms durations but were significantly larger than observed *F*0DLs for the 50-ms duration. A supplementary pitch-matching experiment using two subjects indicated that the contribution of the seventh harmonic to the pitch of the 16-ms complex tone was smaller than would be predicted from the FDL for that harmonic. The results are consistent with the idea that the dominant region shifts upward with decreasing duration, but that the weight assigned to individual harmonics is not always adjusted in an optimal way. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2382476]

PACS number(s): 43.66.Fe, 43.66.Hg, 43.66.Ba [JHG] Pages: 373–382

I. INTRODUCTION

When a complex tone contains many harmonics, its pitch is usually determined primarily by harmonics falling in a relatively restricted frequency region called the "dominant region for pitch." This has been investigated mainly using inharmonic complex tones, in which one or more of the components are shifted from their nominal frequency values and the effect of this on pitch is determined. Ritsma (1967) suggested that, for complex tones with fundamental frequency (*F*0) in the range 100 to 400 Hz, the dominant region lay around the third, fourth, and fifth harmonics; these harmonics are thought to be well resolved in the auditory system (Plomp, 1964; Moore and Ohgushi, 1993). Plomp (1967) and Patterson and Wightman (1976) proposed that the dominant region was not fixed in harmonic number, but tended to shift downward in harmonic number as *F*0 increased; for high *F*0s, the fundamental component itself became dominant. Moore *et al.* (1985) found large individual differences in which harmonics were dominant, but, for *F*0s of 100 to 400 Hz, the dominant harmonics were always among the lowest six. Dai (2000), using *F*0s from 100 to 800 Hz, found that the dominant region was best described by a fixed spectral region centered around 600 Hz.

It has generally been assumed that the dominant region does not correspond to a fixed spectral region, but is affected by the value of *F*0 (decreasing in harmonic number with increasing *F*0), and perhaps to a lesser extent by the relative levels of the harmonics (Moore *et al.*, 1985). However, recently, Gockel *et al.* (2005) demonstrated that the dominant region is affected also by the duration of the complex tone. They used a method similar to that of Moore *et al.* (1985). The dominance of individual harmonics was determined for 16- and 200-ms complex tones containing the first seven harmonics of an *F*0 of 250 Hz. A tone was presented with one of the harmonics mistuned upward or downward by 3%, followed 500 ms later by a perfectly harmonic tone of the same duration. Listeners adjusted the *F*0 of the harmonic tone so that its pitch matched that of the first tone. The extent to which the pitch of the first tone was influenced by the shift in frequency of the single harmonic was taken as a measure of the relative dominance of that harmonic. For the 200-ms duration, the first and second harmonics were the most domi-

[a)]Author to whom correspondence should be addressed. Electronic mail: hedwig.gockel@mrc-cbu.cam.ac.uk

nant ones for all five subjects. For the 16-ms duration, the dominance of higher harmonics increased, especially for the third, fourth, and fifth harmonics, while the dominance of the fundamental component decreased. Thus, the dominant region shifted toward higher harmonics for the shorter duration.

Gockel *et al.* (2005) considered the ability of two models to account for the effect of duration on the dominant region. One model was an autocorrelation model of pitch perception (Meddis and Hewitt, 1991; Meddis and O'Mard, 1997). It was found that this model could give a rough qualitative prediction of the effect of duration, but only with some rather arbitrary choices of the parameters of the model. In particular, the point in time at which the autocorrelation function was calculated (the "offset") had to be adjusted so that a whole number of stimulus cycles was included in the calculation of the autocorrelation function. This effectively provides extra information to the model which would not necessarily be available to a human listener.

The other model considered by Gockel *et al.* (2005) was Goldstein's (1973) pattern recognition model of pitch. In this model, the contribution of a given component to the pitch of a complex tone is assumed to depend upon the precision with which the frequency of that component is represented at the input to the central pitch processor; a higher precision is associated with greater dominance. The precision of the internal representation of the frequency of a sinusoid can be estimated from the frequency difference limen (FDL) for that sinusoid. FDLs for sinusoids increase with decreasing duration, and the increase is greater for low frequencies than for high frequencies (Liang and Chistovich, 1961; Sekey, 1963; Moore, 1973). This could account for the upward shift in the dominant region with decreasing duration. Note, however, that the appropriate measure of the precision of the internal representation of a frequency component in a complex tone is not the FDL for that component when presented in isolation, but rather is the FDL for that component when presented within the complex tone (with the other components fixed in frequency), as proposed by Moore *et al.* (1984). Data on the FDLs of components within short-duration complex tones were not available to Gockel *et al.* (2005), so they generated predictions based on FDLs measured as a function of duration for isolated sinusoids. In the current study, we present data on the FDLs of components within complex tones as a function of duration. The data were intended to provide a stronger test of whether the shift in the dominant region with duration can be explained in terms of pattern-recognition models of pitch.

Another major purpose of the current study was to test the hypothesis of Moore *et al.* (1984) that thresholds for discrimination of the $F0$ of a complex tone can be predicted, using a modification of Goldstein's (1973) model of pitch, from the FDLs of the components within the complex tone. The predictions are based on the assumption that information about the frequencies of the components in the complex tone is weighted optimally by the central pitch processor. Components whose frequencies are represented most precisely are assumed to have the greatest weight in the determination of pitch. The data of Moore *et al.* were consistent with this hypothesis for the 420-ms tones used by them. We assessed whether the hypothesis held for tones of shorter duration.

II. METHOD

A. Stimuli

The stimuli were almost identical to those used by Gockel *et al.* (2005). The test stimulus was a complex tone with a nominal $F0$ of 250 Hz, containing the first seven partials added in sine phase and with equal amplitude. Note that the frequencies of all partials fell within the range where subjects are able to make consistent matches to the pitches of mistuned components (Hartmann *et al.*, 1990). When measuring the threshold for discrimination of $F0$ ($F0$DL), all partials had frequencies which were exact integer multiples of the $F0$. When measuring the FDL of the partial corresponding to harmonic number j, all partials except the jth had frequencies at integer multiples of the $F0$. The jth partial had a frequency which was below $jF0$ in one interval of a trial, and an equal amount above $jF0$ in the other interval. The duration of the tones was 200, 50, or 16 ms, including 8-ms raised-cosine ramps.

The complex tones were presented at a level of 65 dB sound pressure level (SPL) per component in a continuous pink background noise that had a spectrum level of 9 dB at 1 kHz (re 20 μPa). The pink noise was used so that the individual components would all have been approximately 38 dB above their masked threshold. The tones were generated digitally, and were played out using a 16-bit digital-to-analog converter (CED 1401 plus), with a sampling rate of 20 kHz. Stimuli were passed through an antialiasing filter (Kemo 21C30) with a cutoff frequency of 8.6 kHz (slope of 96 dB/octave), and presented monaurally using Sennheiser HD250 headphones. Subjects were seated individually in an IAC double-walled sound-attenuating booth.

B. Procedure

A two-interval, two-alternative forced choice (2I-2AFC) task was used. When measuring the $F0$DL, the $F0$ was $250+\Delta/2$ Hz in one interval and $250-\Delta/2$ Hz in the other interval. The subjects were required to indicate the interval containing the tone with the higher $F0$. In this condition, they were instructed to listen to the low, overall pitch and not to any individual tones that might be heard out. When measuring the FDL for the jth partial, the frequency of that partial was $jF0+\Delta/2$ Hz in one interval and $jF0-\Delta/2$ Hz in the other interval. Note that the partial under investigation was below its nominal (harmonic) frequency in one interval and above its nominal frequency (by the same amount) in the other interval. Thus, both intervals contained inharmonic stimuli, and the degree of mistuning was the same in the two intervals. Therefore, beats or roughness cues would not differentiate between the two intervals (Moore *et al.*, 1984).

Subjects were required to indicate the interval containing the partial with the higher frequency. In this condition, subjects were instructed to listen to the pitch of the specific partial, i.e., to try and hear it out as well as they could; the relatively large change in frequency used at the start of a run

helped to draw attention to the partial whose frequency was to be discriminated. A 1-up 2-down adaptive procedure was used, tracking the 70.7% point on the psychometric function. The value of Δ was increased by a factor of 1.4 after one wrong response and decreased by the same factor after two correct responses. The value of Δ was not allowed to exceed 250 Hz. When the adaptive procedure called for a value of Δ above 250 Hz, the value of Δ was set to 250 Hz. If the value of Δ remained at 250 Hz for three successive trials, the FDL was deemed to be unmeasurable, and the run was aborted. Except in such cases, 12 reversals were obtained. The threshold was defined as the geometric mean of the values of Δ over the last eight reversals. The data reported are the geometric means of at least 11 threshold measurements (see below).

The interval between the two stimuli within a trial was fixed at 500 ms. Each interval was visually marked (by a change in color of a corresponding square on a screen) and visual feedback was provided following each response. The total duration of a single session was about 2 h, including rest times.

Within a single session, either the *F*0DL or the FDL for *one fixed* partial was measured. This was done to allow subjects to "tune in" to the specific condition, i.e., listen to the *F*0 or listen to an individual partial and, in the case of the FDL measurement, to tune in to that specific frequency and hear out the partial in question. The feedback was intended to help the subject to achieve this. In most cases, optimal performance in discriminating the frequency of a single harmonic will be achieved if the subject attends to that harmonic and optimal performance in *F*0 discrimination is achieved if subjects attend to the virtual pitch or to multiple harmonics.

In each session, subjects ran through six blocks, each of which was followed by a short break. Each block consisted of six threshold measurements. In each of the first three blocks there were two threshold estimates for the 200-ms duration, followed by two threshold estimates for the 50-ms duration, and then two estimates for the 16-ms duration. As hearing out a component is easier for long- than for short-duration tones (Moore *et al.*, 1986), the different durations were tested in this order to help subjects to hear out the mistuned partial. In each of the last three blocks of a session there were two threshold estimates for the 50-ms duration, followed by two estimates for the 200-ms duration, followed by two estimates for the 16-ms duration. This was done to balance the order of testing across the two longer durations used. The 16-ms duration was not run at the beginning of a block because it was very difficult to hear out an individual partial for this short duration, and, following a break, subjects would not necessarily be tuned in to a specific partial anymore. One session was run for each partial in turn, before additional sessions were run for that partial. Each subject participated in at least two, usually three sessions, for each partial. If performance was better for the second than for the first session for a given partial, the data from the first session were discarded. If the first measured threshold in a given session was higher than the later ones for the same condition, this threshold was discarded. Thus, the overall threshold estimate reported here is based on at least 11 measurements for each condition and subject. For the purpose of the experimental design, the measurement of the *F*0DL was treated like the measurement of the FDL for a specific partial.

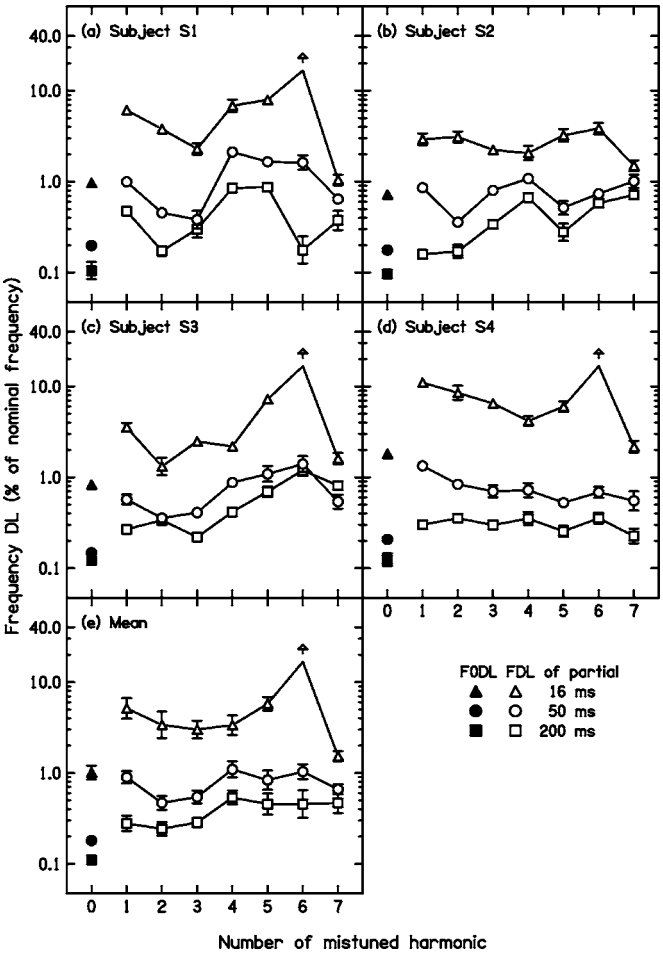


FIG. 1. The open symbols show FDLs for each individual harmonic within a complex tone, expressed as a percentage of the nominal frequency of the harmonic and plotted as a function of harmonic number. The filled symbols show *F*0DLs for the complex tone as a whole. The parameter is signal duration: 200 ms (squares), 50 ms (circles), and 16 ms (triangles). The top four panels show results for the individual subjects. In these panels, error bars indicate ± one standard error across repeated runs. The bottom-left panel shows mean results across subjects. In this panel, error bars indicate ± one standard error across subjects. Up-pointing arrows indicate cases where the FDL could not be measured because of the limitation imposed on the frequency change (see text for details).

C. Subjects

Four subjects, with various degrees of musical experience, participated. They ranged in age from 18 to 27 years, and their quiet thresholds at octave frequencies between 250 and 4000 Hz were within 15 dB of the ISO (2004) standard. To familiarize subjects with the procedure and equipment, they were given about 2 h of practice.

III. RESULTS AND DISCUSSION

The results for each subject and for the mean are plotted in Fig. 1. The FDL for each harmonic is plotted as a percentage of its nominal frequency, with duration as parameter (open symbols). Filled symbols show *F*0DLs, i.e., thresholds

for discrimination of the *F*0 of the whole complex (when all harmonics were shifted by the same percentage, so that the tone remained harmonic). The pattern of variation of the FDLs with harmonic number differed across subjects, especially for the two longer durations. Individual differences of a similar magnitude were observed by Moore *et al.* (1984). The origin of these individual differences is not understood, although it has been suggested that they are related to individual differences in the relative dominance of individual harmonics in determining the pitch of a complex tone (Moore *et al.*, 1985). Large individual differences have been found in all studies of the relative dominance of individual harmonics (Moore *et al.*, 1985; Moore, 1987; Dai, 2000; Gockel *et al.*, 2005). Despite the individual differences, some general trends can be observed.

For the 200-ms duration, the FDLs were generally lowest for the first three harmonics. However, the FDLs for *S*4 were almost independent of harmonic number, and *S*1 had a small FDL for the sixth harmonic. For the 50-ms duration, the FDLs were generally lowest for the second and third harmonics, but again there were individual differences, with *S*2 showing a small FDL for the fifth harmonic and *S*4 showing a slight trend for FDLs to decrease with increasing harmonic number. For the 16-ms duration, FDLs were unmeasurable for the sixth harmonic for *S*1, *S*3, and *S*4; in these cases, the line (without any symbol) is plotted at the point corresponding to the limitation imposed by the procedure (Δ=250 Hz, $\Delta f/f$=16.7%) with an arrow above it. However, all subjects showed relatively small FDLs for the seventh harmonic, which probably reflects the fact that this was the highest harmonic in the complex tones; the "edge" components in complex tones appear to be especially easy to hear out (Plomp, 1964; Moore and Ohgushi, 1993; Moore *et al.*, 2006), and the frequencies of edge components are often discriminated very well (Moore *et al.*, 1984).[1] In the mean data, the harmonics with the lowest FDLs were the second, third, fourth, and seventh. It is noteworthy that, for each duration, the *F*0DL was lower than the FDL of any single harmonic. This is consistent with the results of Moore *et al.* (1984) for 420-ms tones, and it implies that the *F*0DLs depend upon information being combined across harmonics.

To assess whether these data are consistent with the shift in the dominant region with duration, as found by Gockel *et al.* (2005), it is instructive to examine how the FDLs change with duration for each harmonic. These changes are shown in Fig. 2, which plots the FDLs and *F*0DLs for the two shorter durations relative to the FDLs and *F*0DLs for the 200-ms duration; these are referred to as normalized FDLs. For comparison, normalized FDLs for 16- and 50-ms sinusoids presented in isolation are shown in the panel for the mean data (triangles and circles joined by dashed lines; data interpolated from Moore, 1973). The pattern of results varies somewhat across subjects. However, in the mean data, the normalized FDLs for the 16-ms duration decrease progressively as the harmonic number increases from 1 to 4. This means that the worsening of the FDL with decreasing duration is less for the fourth harmonic than for the first harmonic, a trend that is consistent with the upward shift in the dominant region with decreasing duration found by Gockel *et al.* (2005). For the

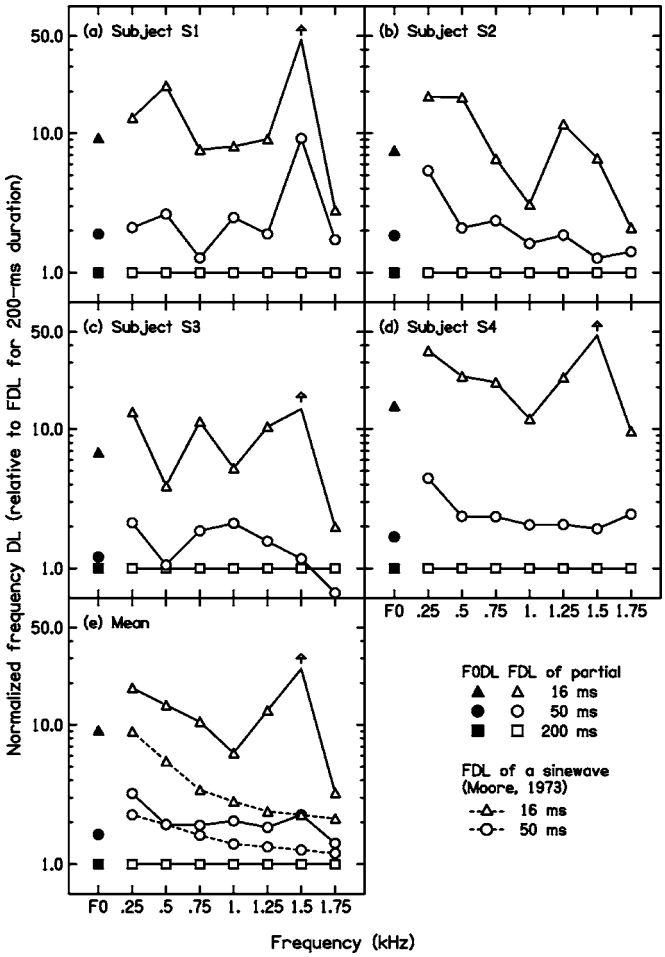


FIG. 2. The same results as in Fig. 1, but with FDLs and *F*0DLs expressed as normalized values, i.e., relative to the FDL or *F*0DL for the 200-ms duration. The dashed lines with triangles and circles indicate normalized FDLs for the 16 and 50-ms durations, respectively, interpolated from the data of Moore (1973) for isolated sinusoids.

50-ms duration, the normalized FDL is larger for the first harmonic than for the remaining harmonics, which is also consistent with an upward shift in the dominant region with decreasing duration.

It is noteworthy that the effect of decreasing duration on the FDLs for harmonics within the complex tone is markedly greater than the effect of duration found for isolated sinusoids by Moore (1973). Furthermore, the normalized FDLs derived from the data of Moore for the 16- and 50-ms durations decrease monotonically with increasing frequency over the range shown, whereas the FDLs for harmonics within the complex tone vary non-monotonically with frequency, showing a distinct peak at 1.5 kHz (the sixth harmonic) for the 16-ms duration and a smaller peak for the 50-ms duration. These differences may reflect difficulty in hearing out the harmonic whose frequency is to be discriminated. Frequency components that are gated on and off synchronously tend to fuse perceptually, especially when the overall duration is short (Moore *et al.*, 1986; Beerends, 1989; Bregman, 1990). This may account for the fact that the normalized FDLs for the 16-ms duration are generally markedly higher for the harmonics within the complex than for the isolated sinusoids. In addition, it becomes progressively harder to hear out in-

dividual harmonics as the harmonic number increases (Plomp, 1964). This effect, combined with the perceptual fusion produced by gating with a short duration, may account for the fact that FDLs for harmonics within the complex tone could not be measured for three of the four subjects for the sixth harmonic and the 16-ms duration. Probably, those three subjects were unable to hear out the sixth harmonic, so they could not discriminate its frequency. However, this does not imply that information about the frequency of the sixth harmonic would be completely unavailable to the central pitch processor. Note that "energetic" masking, due to the spectral spread associated with shortening of the stimulus duration, is unlikely to be the reason for the increase in FDLs for the sixth harmonic at the 16-ms duration. This is because masking due to spectral spreading would be more effective at lower frequencies, where the auditory-filter bandwidth is smaller, than at higher frequencies, where the auditory-filter bandwidth is larger. Thus, the FDLs should increase more for the lower harmonics than for the higher harmonics. In fact, the FDLs for the second, third, and fourth harmonics are markedly lower than that for the sixth harmonic. Also, 8-ms ramps are effective in preventing effects of spectral spreading for center frequencies above 1 kHz (Bacon and Viemeister, 1985).

Based on Goldstein's (1973) theory, Moore *et al.* (1984) proposed that the precision with which a listener could estimate the low pitch of a complex tone (as measured by the *F*0DL) could be predicted from the FDLs of the components within the complex tone, provided that the number of harmonics *N* was sufficiently large that the pitch of the complex tone was unambiguous. This prediction is expressed by

$$\left(\frac{1}{F0\mathrm{DL}_{\mathrm{pred}}}\right)^2 = \sum_{k=1}^{N}\left(\frac{1}{\mathrm{FDL}_k}\right)^2, \tag{1}$$

where $F0\mathrm{DL}_{\mathrm{pred}}$ is the predicted *F*0DL (expressed as a percentage of *F*0) and FDL_k is the FDL for the *k*th harmonic (expressed as a percentage of the frequency of that harmonic) measured when the harmonic is within the complex tone, as here. Note that, in contrast to what was assumed by Goldstein (1973), Moore *et al.* (1984) argued that the channels conveying information about the frequencies of the harmonics to the central processor did not introduce additional noise; hence *F*0DLs could be estimated directly from the values of FDL_k using Eq. (1).

Figure 3 compares values of $F0\mathrm{DL}_{\mathrm{pred}}$ calculated using Eq. (1) with the *F*0DLs actually obtained, for each subject and each duration. The figure also shows the FDLs for the two harmonics with the lowest FDLs, and indicates what the numbers of those harmonics were. For example, for *S*1 for the 200-ms duration, harmonics 2 and 6 had the lowest FDLs. With decreasing duration, both the observed and the predicted *F*0DLs increase in size. To assess whether there was any systematic deviation between the observed and predicted *F*0DLs across durations, a repeated measures two-way analysis of variance (ANOVA) was carried out with factors duration (three values) and type of score (predicted or observed), using the logarithms of the predicted and the mean observed *F*0DL for each subject and condition as input. The

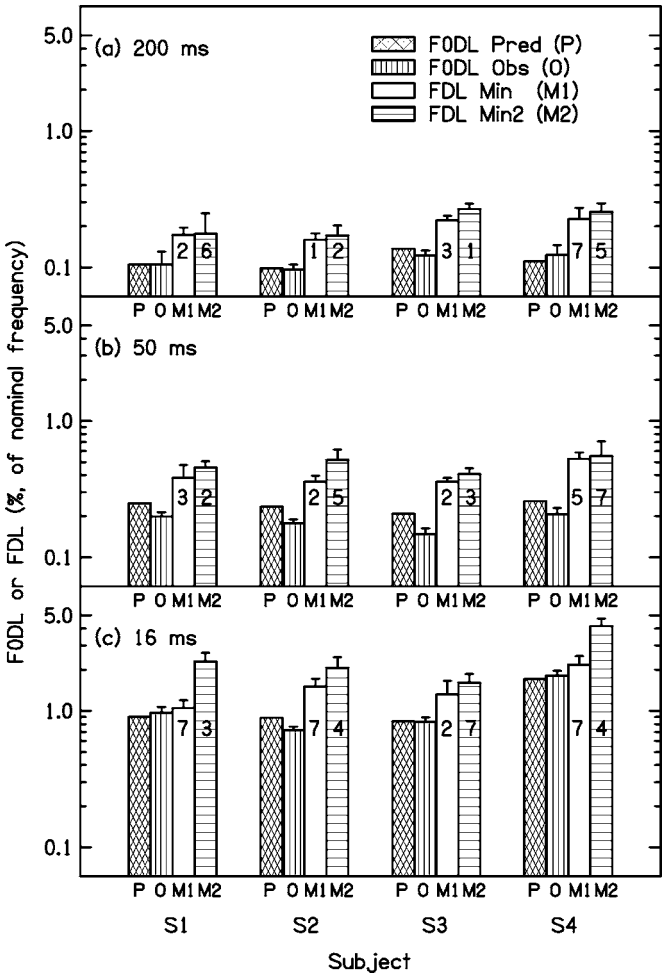


FIG. 3. Comparison of *F*0DLs predicted from Eq. (1) (*P*), observed *F*0DLs (*O*), and FDLs for the two harmonics that led to the lowest FDLs (*M*1 and *M*2). The numbers of these harmonics are indicated within the bars. The FDLs and *F*0DLs are expressed as a percentage of the nominal frequency and are plotted separately for each subject (*S*1–*S*4). Each panel shows results for one duration. Error bars indicate one standard error across repeated runs.

ANOVA revealed a significant main effect of duration $[F(2,6)=120.5, p<0.001]$,[2] but no significant main effect of type of score $[F(1,3)=6.8, p=0.08]$, indicating that overall there was no significant difference between the predicted and observed *F*0DLs. However, there was a significant interaction between duration and type of score $[F(2,6)=15.5, p=0.01]$, indicating that the effect of type of score varied across durations. Thus, we calculated *post hoc* contrasts based on Fisher's least significant difference procedure (Keppel, 1991), separately for each duration. The results of these will be presented below, where we discuss the results for each duration in more detail.

For the 200-ms duration, the predicted and obtained *F*0DLs were similar for all subjects and did not differ significantly ($p=0.87$). Furthermore, the predicted and obtained *F*0DLs were always markedly smaller than the smallest FDLs. This pattern of results is consistent with that found by Moore *et al.* (1984) for a duration of 420 ms. The results are consistent with the idea that, for long-duration tones, information is combined across harmonics in the manner described by Eq. (1).

For the 50-ms duration, the values of $F0DL_{pred}$ were consistently slightly larger than the obtained $F0$DLs, and the difference was significant ($p<0.01$). Again, the predicted and obtained $F0$DLs were always markedly smaller than the smallest FDLs. The discrepancy between the obtained and predicted $F0$DLs cannot be the result of additional noise in the channels conveying information about the frequencies of the harmonics to the central pitch processor, since such a noise, if it existed, would lead to obtained $F0$DLs that were larger than predicted, which is the opposite of what was found. The most likely explanation for the discrepancy is that the values of FDL_k are underestimates of the precision with which the frequencies of the harmonics are represented at the input to the central processor (i.e., the values of FDL_k are too large), because performance in discriminating the frequency of a single harmonic within a complex tone is adversely influenced by difficulty in hearing out individual harmonics when the duration of the sound is short.

Curiously, the predicted and obtained values of the $F0$DLs were similar for all subjects for the 16-ms duration, and did not differ significantly ($p=0.77$). One might have expected that the individual harmonics would be even harder to hear out for this duration, leading to values of FDL_k that considerably underestimate the precision with which the frequencies of the harmonics are represented at the input to the central processor, which in turn would lead to values of $F0DL_{pred}$ that were well above obtained values. This suggests that the increase in FDLs resulting from the difficulty in "hearing out" individual harmonics at a 50-ms duration must have been counteracted by some other factor when the duration was reduced further to 16 ms. Possible reasons for this are discussed in Sec. IV.

We turn now to a more detailed comparison of the present data with the data of Gockel *et al.* (2005) on the dominance of individual partials as a function of duration. To make the comparison, we first derive the pitch shift that would be expected from shifting the frequency of a single harmonic in an otherwise harmonic complex tone, using Goldstein's (1973) model. According to this model [Goldstein's Eq. (13)], for a complex tone with harmonics 1 to N, where N is sufficiently large that the pitch is unambiguous, the mean value of the $F0$ is estimated from the frequencies of the partials f_k according to

$$\overline{F0} = \sum_{k=1}^{N} \frac{kf_k}{\sigma_k^2} \Bigg/ \sum_{k=1}^{N} \frac{k^2}{\sigma_k^2}, \tag{2}$$

where σ_k is the standard deviation of the estimate of the value of f_k at the input to the central processor. We assume here that the value of σ_k is not affected by a small change in f_k. If the jth harmonic is shifted by a small amount, from f_j to $(f_j+\Delta f_j)$, then the change in the mean $F0$ estimated by the central processor is given by

$$\overline{\Delta F0} = \left[\left(\sum_{k=1}^{j-1} \frac{kf_k}{\sigma_k^2} \right) + \frac{j(f_j+\Delta f_j)}{\sigma_j^2} + \left(\sum_{k=j+1}^{N} \frac{kf_k}{\sigma_k^2} \right) - \left(\sum_{k=1}^{N} \frac{kf_k}{\sigma_k^2} \right) \right] \Bigg/ \sum_{k=1}^{N} \frac{k^2}{\sigma_k^2}, \tag{3}$$

which simplifies to

$$\overline{\Delta F0} = \frac{j\Delta f_j}{\sigma_j^2} \Bigg/ \sum_{k=1}^{N} \frac{k^2}{\sigma_k^2}. \tag{4}$$

When the frequency shift of the harmonic is a fixed proportion P of the frequency of the harmonic, as in the experiments of Gockel *et al.* (2005), then, for all j

$$\overline{\Delta F0} = \frac{Pj^2F0}{\sigma_j^2} \Bigg/ \sum_{k=1}^{N} \frac{k^2}{\sigma_k^2}. \tag{5}$$

Dividing both sides by $F0$ gives the change in estimated $F0$ expressed as a proportion of the $F0$

$$\frac{\overline{\Delta F0}}{F0} = \frac{Pj^2}{\sigma_j^2} \Bigg/ \sum_{k=1}^{N} \frac{k^2}{\sigma_k^2} = \frac{Pj^2F0^2}{\sigma_j^2} \Bigg/ \sum_{k=1}^{N} \frac{k^2F0^2}{\sigma_k^2}. \tag{6}$$

We assume here, following Moore *et al.* (1984), that the value of σ_j/f_j is directly proportional to the value of FDL_j (where FDL_j is expressed as a proportion of f_j). Using this assumption, and taking into account the fact that $f_k=kF0$, we get

$$\frac{\overline{\Delta F0}}{F0} = P\frac{1}{\mathrm{FDL}_j^2} \Bigg/ \sum_{k=1}^{N} \frac{1}{\mathrm{FDL}_k^2}. \tag{7}$$

Based on this, we can define a measure of the predicted relative dominance D_j of the jth partial in a complex tone with N harmonics as the predicted pitch shift (expressed as proportion of $F0$) divided by P, i.e., as

$$D_j = \frac{1}{FDL_j^2} \Bigg/ \sum_{k=1}^{N} \frac{1}{\mathrm{FDL}_k^2}. \tag{8}$$

When relative dominance is expressed in this way, the sum of the dominance values across harmonics is equal to 1. The upper panel of Fig. 4 shows the mean values of D_j and standard errors across subjects, calculated using Eq. (8) for the mean FDL data of each subject of the present experiment, separately for the two durations, 16 and 200 ms, used by Gockel *et al.* (2005).

To express the pitch shifts found by Gockel *et al.* (2005) in a similar way, the shift in the pitch of a complex tone produced by a 3% shift in the frequency of each harmonic was divided by the sum of the pitch shifts across harmonics. The results are shown in the lower panel of Fig. 4. The comparison of the predicted values of D_j and the pitch shifts found by Gockel *et al.* should be made with caution, as individual differences were marked both in the present study and in the study of Gockel *et al.*, and the subjects differed across the two studies.

Considering, for the moment, only the results for the first four harmonics, there is an overall similarity between

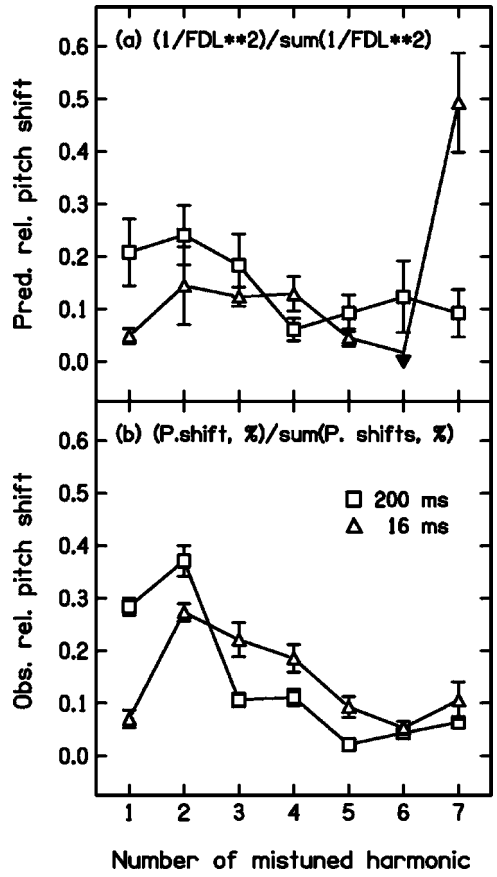


FIG. 4. Comparison of the relative pitch shifts (dominance values) predicted from Eq. (8) (top panel) with the relative pitch shifts found by Gockel *et al.* (2005) (bottom panel). The stimulus duration was 200 ms (squares) or 16 ms (triangles).

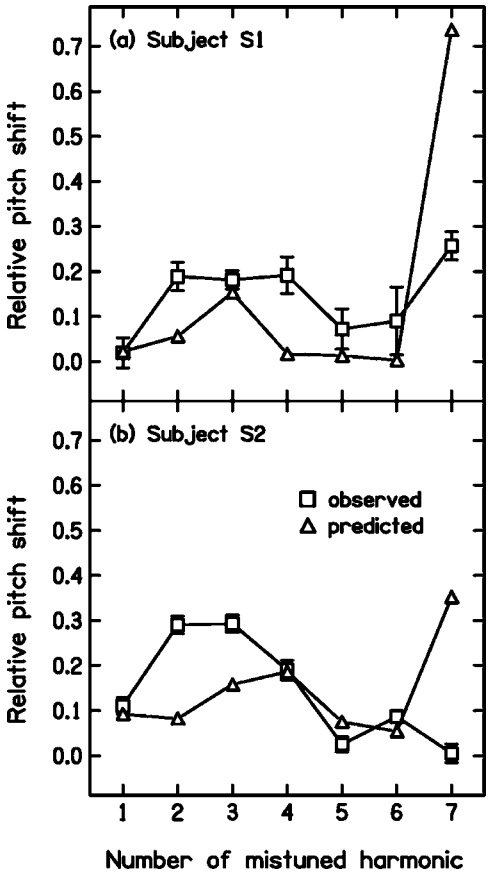


FIG. 5. Results for two individual subjects for 16-ms tones, comparing the relative pitch shifts (dominance values) predicted from Eq. (8) (triangles) with the relative pitch shifts found in a pitch-matching task (squares).

predicted and obtained dominance values. Specifically, for the 200-ms duration, harmonics 1 and 2 are the most dominant in both predicted and obtained values. For the 16-ms duration, harmonics 2, 3, and 4 are the most dominant in both predicted and obtained values. However, the predicted and obtained dominance values differ dramatically for the seventh harmonic at the 16-ms duration. This large discrepancy arises from the fact that, for the 16-ms duration, the values of the FDLs obtained in the present experiment were lowest for the seventh harmonic for three out of the four subjects, while only one of the five subjects of Gockel *et al.* (2005) showed a reasonably strong dominance of the seventh harmonic.

To assess whether the large discrepancy for the seventh harmonic might have been caused by individual differences, two of the subjects of the present experiment, *S*1 and *S*2, were tested in a pitch-matching paradigm identical to that used by Gockel *et al.* (2005), to assess directly the relative dominance of each harmonic in the 16-ms complex tone. Figure 5 compares the observed relative pitch shifts (squares, defined in the same way as for Fig. 4) and the pitch shifts (relative dominance values) predicted using Eq. (8) (triangles). For both subjects, the observed shift for the seventh harmonic was much smaller than the predicted shift. Indeed, subject *S*2 showed no pitch shift for the seventh harmonic. On the other hand, the observed shifts for the second, third, and fourth harmonics were generally larger than predicted. These results confirm that, for the 16-ms duration, the pitch shifts predicted using Eq. (8) do not correspond accurately to the observed shifts, especially for the highest (seventh) harmonic. This means either that the pitch mechanism does not weigh information about the frequencies of individual harmonics in the optimal manner predicted by Eq. (8), or that the FDLs for harmonics within complex tones do not provide good estimates of the precision with which the frequencies of the harmonics are represented at the input to the central processor, when the tones are of short duration (or both).

IV. DISCUSSION

A. Weights applied to harmonics according to a modified version of Goldstein's model

As noted earlier, an assumption underlying Eqs. (1) and (8), which are based on Goldstein's (1973) model, is that information about the frequencies of the individual harmonics is weighted optimally by the central pitch processor. In other words, the pitch processor is assumed to "learn" or to "know" the precision with which the frequencies of the individual harmonics are represented. The weights are not based solely on the frequency of a harmonic, or on harmonic number, as is illustrated by the results of Moore *et al.* (1984). In some of their conditions, they used complex tones containing relatively high harmonics (5–12, 6–12, or 7–12) of a

nominal $F0$ of 200 Hz. For these tones, the FDLs for the 12th harmonic (when measured when the harmonic was presented as part of the complex tone) were relatively small, and that harmonic was therefore predicted to make a substantial contribution to the precision of the estimate of $F0$. When a sinusoid with a fixed frequency of 2600 Hz (corresponding to the 13th harmonic of the nominal $F0$), was added to the complex tones to mask the upper side of the excitation pattern of the 12th harmonic, the FDLs for the 12th harmonic increased markedly, as expected. Correspondingly, $F0$DLs for the complex tone with high-numbered harmonics were smaller when the fixed 2600-Hz component was absent than when it was present. Furthermore, the $F0$DLs were smaller (when expressed as a percentage) than the FDL for any individual harmonic within the complex tones. In contrast, in other conditions, where the complex tone contained harmonics 1–12 and the FDL for the 12th harmonic was not among the smallest, adding a sinusoid with fixed frequency of 2600 Hz also increased the FDL for the 12th harmonic, but did not increase the $F0$DL for the complex tone. Thus, adding a fixed frequency component at the upper spectral edge of the complex tone did not in itself impair $F0$ discrimination. It seems that the 12th harmonic contributed substantially to $F0$ discrimination only in the absence of the low harmonics that usually have low FDLs.

The results of Moore *et al.* (1984) clearly show that the contribution of a given harmonic to the discrimination of $F0$ is not fixed for a given harmonic number. Furthermore, the dominance of an individual harmonic, as measured using a pitch-matching paradigm, increases with increasing relative level of that harmonic (Moore *et al.*, 1985). Taken together, these results appear to be consistent with the idea that the pitch processor can adjust the weight attached to a harmonic according to the precision with which the frequency of that harmonic can be discriminated. However, our supplementary experiment using 16-ms tones showed that, although the highest harmonic produced the lowest FDL, it made a relatively small contribution to the pitch of the complex (Fig. 5). Hence it appears that the weight applied to a given harmonic in a pitch-matching task cannot always be predicted by the FDL for that harmonic.

It seems likely that, both in pitch matching and in $F0$-discrimination tasks, the pitch mechanism applies a suboptimal weight to the highest harmonic for short-duration tones. For example, the pitch processor may have some knowledge acquired over the longer term, especially for tones of relatively long duration, say 100 to 400 ms, since musical tones often have durations in this range (Fraisse, 1982). With repeated exposure to musical tones, or to the human voice (Terhardt, 1974), the pitch processor may acquire knowledge about which harmonics are represented most precisely, and it may then operate in a near-optimal manner. However, human listeners are rarely exposed to tones with durations as short as 16 ms. For such tones, the pitch processor may not apply optimal weights, but rather may apply weights that are partly influenced by knowledge obtained from longer-duration tones. For long tones, the seventh harmonic does not usually play a strong role; it typically lies somewhat above the dominant region for an $F0$ of 250 Hz. According to this explanation, although the weight assigned to individual harmonics does change when the duration is altered (Gockel *et al.*, 2005), this change is suboptimal, and the highest harmonic does not receive the weight that would be expected from its good discriminability.

The idea that the weights applied to the highest harmonic are suboptimal for short-duration tones could explain an intriguing aspect of the results shown in Fig. 3. Recall that, although the predicted F0DLs [Eq. (1)] exceeded those obtained for the 50-ms tones, the match between theory and data was good when the duration was reduced further to 16 ms. In Sec. III we suggested that the discrepancy for the 50-ms duration could be due to subjects' difficulty in hearing out individual harmonics at short durations, and that at a duration of 16 ms some other factor intervened to counteract this effect. Such a factor could arise if a sub-optimal weight were applied to the highest harmonic. For the 16-ms duration, the values of FDL_k were lowest for the seventh harmonic, causing it to have a strong influence on the values of $F0DL_{pred}$. However, if this harmonic received a suboptimal weight, this would have increased the $F0$DLs actually obtained. Alternatively, it could be that the efficiency with which information is combined across harmonics decreases at short durations.

B. "Weights" predicted by an autocorrelation model

The discussion so far has focused on the idea, based on Goldstein's (1973) model, that the central pitch processor applies weights to the estimates of the frequencies of the individual harmonics, and that the appropriate choice of these weights requires knowledge about the precision of the frequency estimates. However, it may be the case that such knowledge is not actually required; rather, the weighting may occur in an automatic way. Consider, for example, the autocorrelation model of pitch mentioned in the Introduction (Meddis and Hewitt, 1991; Meddis and O'Mard, 1997). According to this model, autocorrelation functions (ACFs) are calculated based on the neural activity in individual frequency channels, and the ACFs are then summed across channels to give a summary ACF (SACF). The extent to which an individual harmonic contributes to the SACF presumably depends upon the extent to which neurons are phase locked to that harmonic, and this in turn may be closely related to the FDL for that harmonic (when measured within the complex tone). For example, there may be considerable phase locking to the highest ("edge") harmonic in a complex tone, because the upper side of the excitation pattern evoked by that harmonic is not disrupted by any higher harmonics. Hence, neurons tuned at and above the frequency of the highest harmonic would all show phase locking to that harmonic. Similarly, increasing the level of a single harmonic relative to that of adjacent harmonics will lead to an increase in the number of neurons whose responses are phase locked to the incremented harmonic. This could explain why, at long durations, $F0$DLs can be predicted with reasonable accuracy from the FDLs for the harmonics within the complex tone, without any requirement for the possession of knowledge by

the pitch processor. However, this account does not explain why the prediction becomes less accurate when the duration is reduced to 50 ms.

V. SUMMARY AND CONCLUSIONS

*F*0DLs and FDLs for individual partials within a seven-component complex tone were measured for three stimulus durations, 200, 50, and 16 ms. The following are the main findings:

1. There were marked individual differences in the way that the FDLs varied with harmonic number, as found in previous similar work.
2. The FDLs increased with decreasing duration. The effect of duration decreased as the harmonic number increased from 1 to 4, but then increased as the harmonic number increased to 6. FDLs for the sixth harmonic were unmeasurable for three of the four subjects for the 16-ms duration, perhaps reflecting difficulty in hearing out this harmonic. FDLs decreased for the seventh harmonic, and for the 16-ms duration were lower than for any other harmonic, for three of the four subjects.
3. For each duration, *F*0DLs were smaller than the smallest FDL for any individual harmonic. This indicates that information is combined across harmonics in the discrimination of *F*0.
4. For the 200-ms duration, *F*0DLs predicted from the FDLs, using a modification of Goldstein's (1973) model proposed by Moore *et al.* (1984), corresponded well with observed *F*0DLs. However, for the 50-ms duration, predicted *F*0DLs were significantly larger than observed *F*0DLs. This may have happened because the measured FDLs underestimate the precision with which the frequencies of the individual partials are represented at the input to the pitch processor. For the 16-ms duration, the predicted *F*0DLs were close to the observed *F*0DLs. However, this correspondence may result from two competing factors: (1) The FDL for the seventh harmonic was small, but it probably did not make a strong contribution to F0 discrimination; (2) The measured FDLs underestimate the precision with which the frequencies of the individual partials are represented at the input to the pitch processor.
5. A supplementary pitch-matching experiment using two subjects indicated that the contribution of the seventh harmonic to the pitch of the 16-ms complex tone was smaller than would be predicted from the FDL for that harmonic.

Overall, these results, in combination with those of Gockel *et al.* (2005), suggest that the dominant region shifts upward (towards higher harmonic numbers) with decreasing duration, but that the weight assigned to individual harmonics may not always be adjusted in an optimal way. In particular, the highest harmonic does not receive the weight that would be expected from its good discriminability.

ACKNOWLEDGMENTS

This work was supported by EPSRC Grant No. EP/D501571/1. Author B.M. was supported by the MRC. The authors thank Laurent Demany, John Grose, and an anonymous reviewer for their comments.

[1]The small FDLs for the seventh harmonic probably occurred because the upper side of the excitation pattern evoked by that harmonic was relatively free from interference or masking produced by adjacent harmonics, so the frequency could be extracted precisely from phase-locking information evoked by the high-frequency side of the pattern (Moore and Ohgushi, 1993; Moore *et al.*, 2006). To check whether the FDLs for the seventh harmonic were linked to the fact that it was the highest harmonic in the complex tone, FDLs for the seventh harmonic were measured for three subjects (*S*1, *S*2, and *S*3) in the absence and presence of a half-octave wide band of noise. The noise had a lower cutoff frequency of 2 kHz (corresponding to the frequency of the eight harmonic) and the noise level per ERB_N within its passband corresponded to the level of each component in the complex (65 dB SPL). For the two shorter durations, the noise resulted in a clear increase in the FDLs. Indeed, for the 16-ms duration, FDLs for the seventh harmonic in the presence of the noise were not measurable. These results confirm that, especially for the shortest duration, the small measured FDL for the seventh harmonic in the main experiment was strongly determined by the fact that it was an edge component. *F*0DLs were also measured for the same three subjects in the presence and absence of the noise. The *F*0DLs were not significantly affected by the presence or absence of the noise for any of the three durations. Thus, despite the seventh harmonic having, on average, the smallest FDL of all harmonics for the shortest duration in the main experiment, it did not appear to make a significant contribution to the *F*0DL, since *F*0 discrimination was not affected by a noise that markedly increased the FDL.

[2]Throughout the paper, if appropriate, the Huynh-Feldt correction was applied to the degrees of freedom (Howell, 1997). In such cases, the corrected significance value is reported.

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

Beerends, J. G. (**1989**). "The influence of duration on the perception of pitch in single and simultaneous complex tones," J. Acoust. Soc. Am. **86**, 1835–1844.

Bregman, A. S. (**1990**). *Auditory Scene Analysis: The Perceptual Organization of Sound* (Bradford Books, MIT, Cambridge, MA).

Dai, H. (**2000**). "On the relative influence of individual harmonics on pitch judgment," J. Acoust. Soc. Am. **107**, 953–959.

Fraisse, P. (**1982**). "Rhythm and tempo," in *The Psychology of Music*, edited by D. Deutsch (Academic, New York).

Gockel, H., Carlyon, R. P., and Plack, C. J. (**2005**). "Dominance region for pitch: Effects of duration and dichotic presentation," J. Acoust. Soc. Am. **117**, 1326–1336.

Goldstein, J. L. (**1973**). "An optimum processor theory for the central formation of the pitch of complex tones," J. Acoust. Soc. Am. **54**, 1496–1516.

Hartmann, W. M., McAdams, S., and Smith, B. K. (**1990**). "Hearing a mistuned harmonic in an otherwise periodic complex tone," J. Acoust. Soc. Am. **88**, 1712–1724.

Howell, D. C. (**1997**). *Statistical Methods for Psychology* (Duxbury, Belmont, CA).

ISO 389-8 (**2004**). "Acoustics—Reference zero for the calibration of audiometric equipment—Part 8: Reference equivalent threshold sound pressure levels for pure tones and circumaural earphones" (International Organization for Standardization, Geneva).

Keppel, G. (**1991**). *Design and Analysis: A Researcher's Handbook* (Prentice-Hall, Upper Saddle River, New Jersey).

Liang, C.-A., and Chistovich, L. A. (**1961**). "Frequency difference limens as a function of tonal duration," Sov. Phys. Acoust. **6**, 75–80.

Meddis, R., and Hewitt, M. (**1991**). "Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I: Pitch identification," J. Acoust. Soc. Am. **89**, 2866–2882.

Meddis, R., and O'Mard, L. (**1997**). "A unitary model of pitch perception," J. Acoust. Soc. Am. **102**, 1811–1820.

Moore, B. C. J. (**1973**). "Frequency difference limens for short-duration tones," J. Acoust. Soc. Am. **54**, 610–619.

Moore, B. C. J. (**1987**). "The perception of inharmonic complex tones," in *Auditory Processing of Complex Sounds*, edited by W. A. Yost and C. S. Watson (Erlbaum, Hillsdale, N.J.).

Moore, B. C. J., Glasberg, B. R., Low, K.-E., Cope, T., and Cope, W.

(**2006**). "Effects of level and frequency on the audibility of partials in inharmonic complex tones," J. Acoust. Soc. Am. **120**, 934–944.
Moore, B. C. J., Glasberg, B. R., and Peters, R. W. (**1985**). "Relative dominance of individual partials in determining the pitch of complex tones," J. Acoust. Soc. Am. **77**, 1853–1860.
Moore, B. C. J., Glasberg, B. R., and Peters, R. W. (**1986**). "Thresholds for hearing mistuned partials as separate tones in harmonic complexes," J. Acoust. Soc. Am. **80**, 479–483.
Moore, B. C. J., Glasberg, B. R., and Shailer, M. J. (**1984**). "Frequency and intensity difference limens for harmonics within complex tones," J. Acoust. Soc. Am. **75**, 550–561.
Moore, B. C. J., and Ohgushi, K. (**1993**). "Audibility of partials in inharmonic complex tones," J. Acoust. Soc. Am. **93**, 452–461.
Patterson, R. D., and Wightman, F. L. (**1976**). "Residue pitch as a function of component spacing," J. Acoust. Soc. Am. **59**, 1450–1459.
Plomp, R. (**1964**). "The ear as a frequency analyzer," J. Acoust. Soc. Am. **36**, 1628–1636.
Plomp, R. (**1967**). "Pitch of complex tones," J. Acoust. Soc. Am. **41**, 1526–1533.
Ritsma, R. J. (**1967**). "Frequencies dominant in the perception of the pitch of complex sounds," J. Acoust. Soc. Am. **42**, 191–198.
Sekey, A. (**1963**). "Short-term auditory frequency discrimination," J. Acoust. Soc. Am. **35**, 682–690.
Terhardt, E. (**1974**). "Pitch, consonance, and harmony," J. Acoust. Soc. Am. **55**, 1061–1069.

The effect of fundamental frequency on the brightness dimension of timbre

Jeremy Marozeau[a)]
Communication Research Laboratory, Dept. of Speech-Language Pathology & Audiology (106A FR), Institute of Hearing, Speech and Language, Northeastern University, 360 Huntington Avenue, Boston, Massachusetts 02115

Alain de Cheveigné
Equipe Audition, FRE 2929 LPP, CNRS-Université Paris5, Ecole Normale Supérieure, 29 rue d'Ulm 75230, Paris, France

(Received 10 May 2006; revised 2 October 2006; accepted 9 October 2006)

The dependency of the brightness dimension of timbre on fundamental frequency (*F0*) was examined experimentally. Subjects compared the timbres of 24 synthetic stimuli, produced by the combination of six values of spectral centroid to obtain different values of expected brightness, and four *F0*'s, ranging over 18 semitones. Subjects were instructed to ignore pitch differences. Dissimilarity scores were analyzed by both ANOVA and multidimensional scaling (MDS). Results show that timbres can be compared between stimuli with different *F0*'s over the range tested, and that differences in *F0* affect timbre dissimilarity in two ways. First, dissimilarity scores reveal a term proportional to *F0* difference that shows up in the MDS solution as a dimension correlated with *F0* and orthogonal to other timbre dimensions. Second, *F0* affects systematically the timbre dimension (brightness) correlated with spectral centroid. Interestingly, both terms covaried with differences in *F0* rather than chroma or consonance. The first term probably corresponds to pitch. The second can be eliminated if the formula for spectral centroid is modified by introducing a corrective factor dependent on *F0*. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2384910]

PACS number(s): 43.66.Jh, 43.66.Hg, 43.66.Lj, 43.75.Cd, 43.75.Bc [DD] Pages: 383–387

I. INTRODUCTION

In a previous study (Marozeau *et al.*, 2003) we looked at the change with fundamental frequency (*F0*) of the timbre of sounds produced by a set of 12 musical instruments. The timbre of most instruments was found to be stable as a function of the note played. Nevertheless, small *F0* dependencies were observed that appeared to be specific to particular instruments. The choice of natural instrument sounds rather than synthetic stimuli in that study ensured musical relevance, but the large interinstrument variations of timbre within the stimulus set made the investigation of small *F0*-dependent effects difficult. Because distance has a quadratic dependency on the difference along each dimension, large interinstrument differences along one dimension can "swamp" smaller differences along other dimensions.

This study used a set of synthetic instrumental sounds designed to differ along the single physical dimension of spectral centroid (in addition to *F0*). Variations along other dimensions known to affect timbre (temporal envelope shape and spectral spread) were minimized so as to maximize sensitivity along the dimension of interest. The perceptual correlate of spectral centroid is termed "brightness." The centroid is defined from the spectral envelope that determines amplitudes of all partials, and thus does not depend directly on *F0*, but there are several reasons to expect a dependency of the perceptual correlate, brightness, on *F0*. One is that *F0* determines pitch, which is known to have a complex multidimensional nature, involving a cyclic chroma dimension related to position within the octave (*F0* modulo a ratio power of 2), a "tone height" dimension related to *F0*, and possibly a "spectral pitch" dimension determined by the overall spectral distribution (Shepard, 1999). The latter can certainly be expected to interact with brightness. Another reason is that *F0* dependencies of *vowel timbre* have been documented in several studies. Specifically, it appears that in order to maintain constant vowel identity over a one-octave increase of *F0*, formant frequencies must be shifted upward by about 10% (Slawson, 1968; Nearey, 1989). One might think that such dependencies are specific to speech sounds, but Slawson (1968) found similar effects when subjects were instructed to treat the stimuli as musical sounds rather than vowels.

The standard methodology for timbre studies is to apply multidimensional scaling (MDS) analysis to dissimilarity matrices obtained by asking subjects to rate the dissimilarity between pairs of stimuli. Our previous study (Marozeau *et al.*, 2003) extended it to allow timbre comparisons between sounds that differed in *F0*. Subjects were instructed to ignore the resulting difference in pitch. They were quite successful in doing so, but there was nevertheless some evidence of an effect of *F0* difference—or the pitch difference that it induces—on timbre. The present study aimed at understanding the nature of this interaction.

Supposing that *F0*-induced *pitch* differences affect timbre, one can speculate on the form of the dependency. Do

[a)]Electronic mail: marozeau@neu.edu

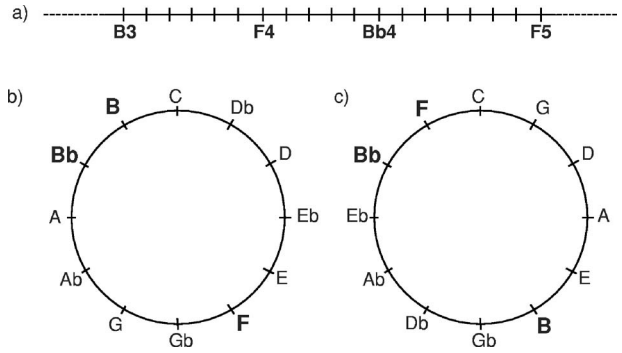


FIG. 1. Hypothetical geometrical structures to predict dissimilarity between sounds that differ in *F0*: (a) logarithmic frequency axis, (b) chroma circle, and (c) circle of fifths.

timbre differences vary monotonically with the *frequency* difference between notes (on a linear or log scale) [Fig. 1(a)]? Are they instead a function of the difference in chroma? If so, they should vary as distance along a chroma circle [Fig. 1(b)]. Because the stimuli are presented pairwise they might also be function of the *consonance* between notes. If so, they should vary as distance along a circle of fifths [Fig. 1(c)] (Shepard, 1982). Stimulus frequencies in this study were chosen so as to test each of these hypotheses.

II. METHOD

A. Stimuli

Stimuli were produced according to a simple model of additive synthesis. Each had a spectrotemporal envelope shaped as the outer product of a frequency-independent temporal envelope multiplied by a time-independent spectral envelope function. The temporal envelope, common to all stimuli, comprised a linear 50-ms onset (attack) followed by a 22.5-ms linear decay of 20% of the maximum amplitude, a 382.5-ms constant amplitude part (sustain), and a linear offset (release) of 45 ms (so-called ADSR or attack decay sustain release profile). The spectral envelope was shaped as a Gaussian when expressed as "partial loudness" (intensity per critical band raised to the power 0.3) as a function of frequency on an equivalent rectangular bandwidth (ERB) (Moore, 2003) scale. The width of the Gaussian was 5 ERB and the same for all stimuli. The centroid of the Gaussian envelope took on six values equally spaced on an ERB-rate scale from 17 to 22 ERB-rate (1196, 1358, 1539, 1739, 1963, and 2212 Hz, respectively). Figure 2 illustrates the waveform and spectra of two stimuli with the same *F0* (247 Hz) but different centroids (17 and 22 ERB-rate).

For each of these envelopes, stimuli were produced at four *F0*'s: 247, 349, 466, and 698 Hz (notes *B3*, *F4*, *Bb4*, and *F5*), for a total of 24 stimuli. This *F0* set was designed to produce intervals both small and large in terms of frequency, chroma, and consonance.

Stimuli were sampled at 44.1 kHz with a resolution of 16 bits. They were presented diotically over earphones at approximately 75 dBA. The term "instrument" will be used in the following to designate the set of stimuli with the same centroid.

B. Listeners

Fourteen subjects (including seven women and seven musicians) participated in the experiment. Musicians were defined as having played an instrument for at least 3 years.

C. Procedure

The subjects were asked to rate the dissimilarity of the 276 possible pairs of the 24 stimuli. They were instructed to

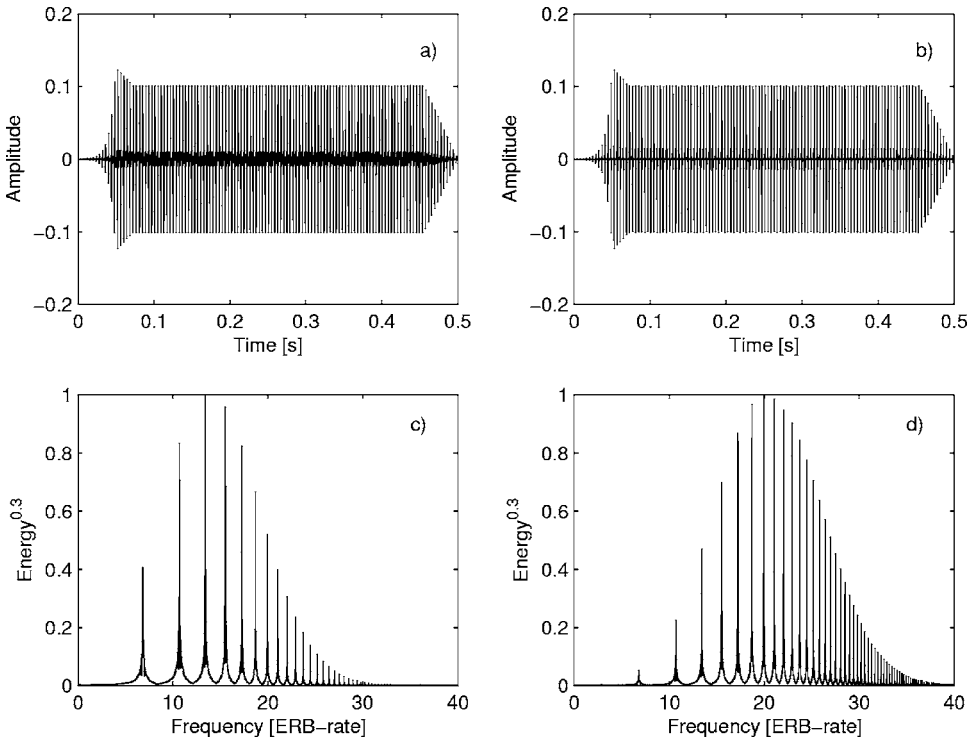


FIG. 2. Waveforms (top) and spectra (bottom) of two stimuli used in our experiments. Both had the same F0 (247 Hz) but different spectral centroids: 17 ERB rate [(a) and (c)] and 22 ERB rate [(b) and (d)].

base their judgments only on the timbre of each stimulus and to ignore the pitch difference. Previous experiments showed that this task was possible, at least for *F0* differences below one octave. The order within pairs and the order of pairs were random (a different randomization was used for each session and subject). The experiment was run inside an audiometric booth, and stimuli were presented diotically over Sennheiser 520 II headphones. Further procedural details can be found in Marozeau *et al.* (2003).

III. RESULTS

A. Outliers, effect of musical experience

Correlation coefficients between dissimilarity scores were calculated for all pairs of subjects. These scores were submitted to a hierarchical cluster analysis to identify eventual outlier subjects. No outlier was found. An ANOVA was performed with between-subjects factor musical experience (2) and within-subjects factor instrument pair (276), to reveal an eventual effect of musical experience. No effect of musical experience was found, either as a main effect or as an interaction. Results were therefore averaged across all subjects.

B. MDS analysis

The data were first transformed with a hyperbolic arctangent function to attempt to correct for the effect of the bounded response scale (Schonemann, 1983; Marozeau, 2004). MDS produces a spatial configuration such that distances between points fit observed dissimilarities between instruments. Distances being unbounded whereas dissimilarities are bounded, the solution is necessarily distorted. The transformation reduces this distortion. The hyperbolic tangent has a slope that is close to 1 for arguments between 0 and 0.5, and that increases exponentially to infinity as its argument approaches 1. It thus leaves unchanged the dissimilarities smaller than 0.5 and expands dissimilarities close to 1.

Transformed scores were analyzed using the MDSCAL procedure, implemented according to the SMACOFF algorithm (Borg and Groenen, 1997). A two-dimensional solution was selected because higher-dimensional solutions did not decrease significantly the stress of the model (Borg and Groenen, 1997). As the MDSCAL solution is rotationally undetermined, the solution was rotated with a procrustean

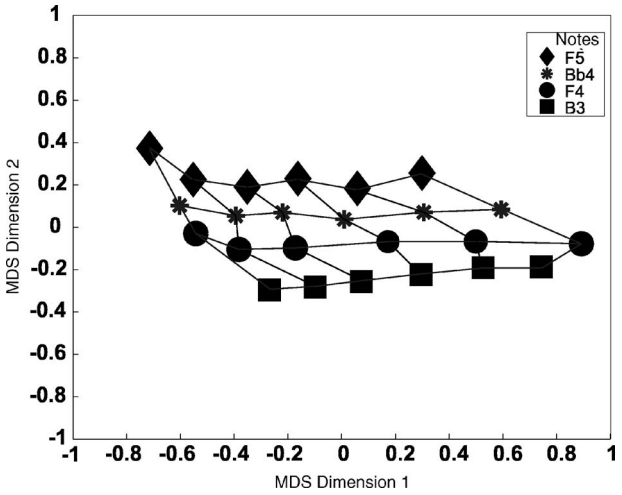


FIG. 3. Two-dimension multidimensional scaling (MDS) solution. The solution was rotated to maximize correlation between dimension 1 and spectral centroid.

procedure in order to maximize the correlation between spectral centroid and position along the first MDS dimension for stimuli with a 247 Hz (*B3*) *F0*. Figure 3 shows this solution.

Stimuli at *B3* are represented by squares, those at *F4* by circles, those at *Bb4* by stars, and those at *F5* by diamonds. Each stimulus is connected by a segment to the two stimuli with the closest spectral centroid and the same *F0* and to the two stimuli with the closest *F0* and the same spectral centroid. Roughly speaking, stimuli are distributed along the first dimension in order of their spectral centroid, and along the second dimension in order of *F0*. Stimuli with a given *F0* tend to follow a horizontal line, indicating that the dimension related to *F0* is not affected by spectral centroid. In contrast, stimuli with a given centroid follow a line slanted to the left, indicating that the dimension related to the centroid is affected by *F0*. In other words, an *increase* in *F0* has an effect similar to a *decrease* in the centroid.

C. Horizontal shift

To test if this shift is significant, an ANOVA was performed on a restricted set of the data. To better understand this analysis let us consider only two instruments (X and Y) differing along a timbre dimension. The positions of these two instruments along this dimension are represented at two different *F0*'s by X_1, Y_1 and X_2, Y_2, respectively, in Fig. 4.

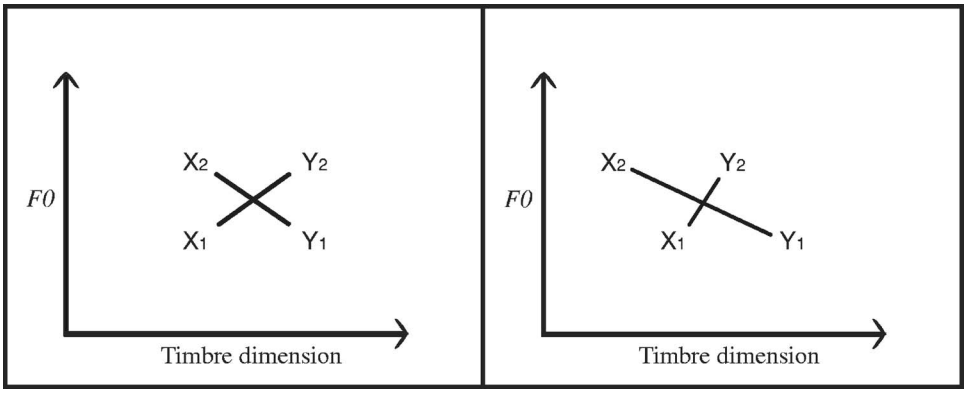


FIG. 4. Hypothetical geometrical structures describing the dependency on *F0* of the timbres of two instruments. Left panel represents the case where the timbres do not change (along the horizontal dimension) with *F0*. Right panel represents the case where the timbres do change with *F0*.

TABLE I. Amount of variance (R^2) of timbre dissimilarity accounted for by each factor manipulated in the experiment. Only the values of effects significant at the $p=0.005$ level are included, as determined by an ANOVA. Factors are instrument pair (IP, df=15) and *F0* order (*F0*, df=2).

	B3			F4			Bb4		
	IP	*F0*	IP×*F0*	IP	*F0*	IP×*F0*	IP	*F0*	IP×*F0*
F4	42.87	8.53	5.11						
Bb4	28.49	9.5	n.s.	43.41	11.39	3.92			
F5	8.31	12.47	n.s.	21.17	27.45	3.85	29.18	17.45	n.s.

Two cases may be considered. In the first, represented in the left panel, the position of the instruments remains invariant along the timbre dimension with a change of *F0*. The distance $\overline{X_1Y_2}$ is then the same as the distance $\overline{X_2Y_1}$. In the second case, represented in the right panel, the position of the instruments moves uniformly along the timbre dimension as function of *F0*; the distances $\overline{X_1Y_2}$ and $\overline{X_2Y_1}$ are in this case different.

To test if an instrument shifts along the first timbre dimension with respect to another instrument, an ANOVA was performed with the dissimilarities of all the pairs containing both instruments at different *F0*'s. The two main factors were the instrument pair (X_1Y_3 vs. X_2Y_4,) and the *F0* order inside the pair (X_1Y_3 vs. X_3Y_1). A lack of change with *F0* along the timbre dimension would imply a lack of significant effect for the latter factor. Conversely, a significant effect would imply a timbre change with *F0*.

Table I shows the result of this analysis for each instrument pair.

The results show that the factor of the *F0* order is always significant. Therefore the shift can be considered as significant for every instrument.

IV. DISCUSSION

A. The dependency of timbre on spectral centroid

Stimuli fall along the first dimension of the MDS solution in order of their spectral centroid, consistent with the pattern found for natural instruments (Marozeau *et al.* 2003). This is not surprising given that they were designed to vary according to that dimension. Their approximately equal spacing along the psychological dimension agrees with their equal spacing along the physical dimension. It is, however, unlikely that our methods could reveal a discrepancy in this respect, should it exist, because of experimental noise. The leftward shift with increasing *F0* suggests that this perceptual dimension is also slightly dependent on *F0*.

B. The effect of *F0*

Stimuli fall along the second dimension of the MDS solution in order of their *F0*. Evidence for a similar *F0*-dependent dimension was found for natural instruments by Marozeau *et al.* (2003) for a smaller range of *F0*'s. Scatter along this dimension is about 40% of the scatter along the first dimension, indicating that the effect of *F0* is moderate compared to the effect of the centroid. The question raised in the Introduction, concerning the dependency on *F0 difference*, chroma similarity, or consonance, is answered unambiguously: the component of dissimilarity induced by an *F0* difference is not a function of the difference in chroma, or the consonance between notes, but rather the size of the difference in *F0*.

A similar remark holds for the small variations observed along dimension 1: these appear to covary with a linear scale of *F0*, rather than a circular scale of chroma or a circular scale of consonance (Fig. 1).

C. An improved "spectral centroid" descriptor

The stimuli were created according to a definition of the spectral centroid as described by Marozeau *et al.* (2003). If this descriptor were accurate to predict the first dimension of timbre (brightness) independently of *F0*, we should have observed no shift along that dimension with *F0*. The significant effect that we did observe can be interpreted either as implying that perceptual dimension *does* depend slightly on *F0*, or that we should search for a better descriptor that ensures that it does not. There is no definite way to choose between these rival interpretations, but for practical applications it would be nice to have a signal-based descriptor of this dimension that is *not* sensitive to *F0*. The purpose of this paragraph is to present such an improved descriptor.

The first stages are the same as in Marozeau *et al.* (2003). Briefly, the waveform was first filtered to model the sensitivity of the outer and middle ear (Killion, 1978). Then it was filtered by a gammatone filterbank (Patterson *et al.*, 1992) with channels spaced at half-ERB intervals on an ERB-rate scale (z) between 25 Hz and 19 kHz (Hartmann, 1998). Instantaneous power was calculated within each channel and smoothed by delaying it by $1/4f_c$ (where f_c is the characteristic frequency of the channel), adding it to the undelayed power, and convolving the sum with an 8-ms window. Smoothed power was then raised to the power 0.3 to obtain a rough measure of "partial loudness" for each channel. The partial loudness-weighted average of ERB-rate was taken over channels, the result being an "instantaneous spectral centroid" function of time according to

$$\bar{Z}(t) = \sum_z z\psi_z(t) \Big/ \sum_z \psi_z(t), \tag{1}$$

where $\psi(t)$ is the "partial loudness" of the channel z at instant t. Finally, the instantaneous centroid $\bar{Z}(t)$ was weighted by "instantaneous loudness" (sum over channels of partial loudness) and averaged over time to obtain a single descriptor value, $\bar{Z}$, to characterize the entire signal.

To better predict position along the first timbre dimension, a correction of the spectral centroid is now proposed. First, the descriptor is converted from ERB-rate to Hz according to the formula:

$$\bar{f} = (\exp(\bar{Z}/9.26) - 1)/0.004\,37, \tag{2}$$

where $\bar{f}$ is the value of the spectral centroid in Hz (Hartmann, 1998). Then the value of the *F0* of the stimulus is subtracted from $\bar{f}$:

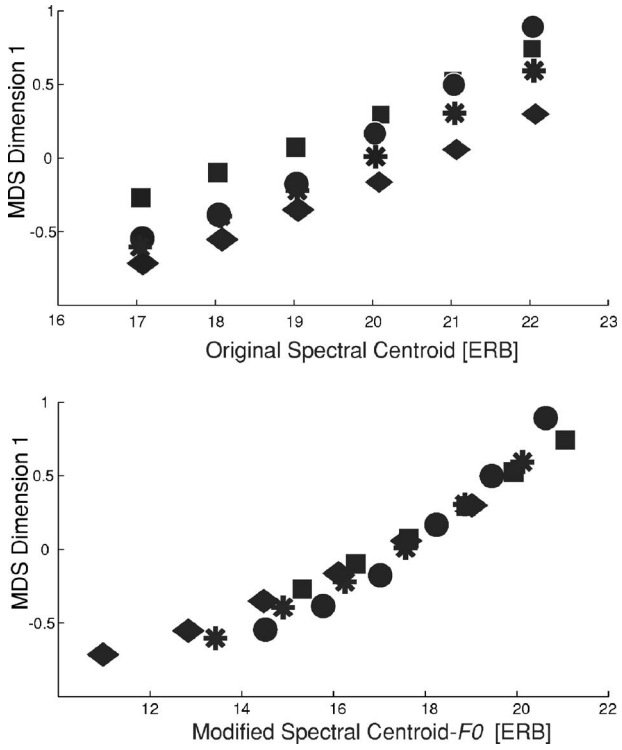


FIG. 5. Scatter plots showing the dependency of the first MDS dimension on signal-based predictors of brilliance. Upper panel: uncorrected spectral centroid. Lower panel: corrected spectral centroid.

$$\overline{f_{\text{corrected}}} = \bar{f} - F0, \tag{3}$$

where $\overline{f_{\text{corrected}}}$ corresponds to the corrected value of $\bar{f}$. Finally the value is converted back in ERB-rate:

$$\overline{Z_{\text{corrected}}} = 9.26 \ln(0.004\,37\overline{f_{\text{corrected}}} + 1), \tag{4}$$

where $\overline{Z_{\text{corrected}}}$ corresponds to the corrected value of $\bar{Z}$ in ERB-rate. The upper panel of Fig. 5 shows a scatter plot of $\bar{Z}$ with the projection along the first timbre dimension. The lower panel shows the scatter plot of $\overline{Z_{\text{corrected}}}$ with the projection along the first dimension. The stimuli are better aligned along the regression line. This can be quantified by the coefficient of correlation of 0.97 (0.91 before correction), explaining more than 93% of the variance (82% before correction) ($dl=22$; $p<0.001$).

A similar correction procedure has already been invoked in studies of vowel perception (Traunmuller, 1981; Hoemeke and Diehl, 1994). Specifically, it has been proposed to "correct" the values of vowel formants, in particular the first formant, by subtraction of the value of *F0* from that of the formant frequency.

V. CONCLUSIONS

We found that cross-*F0* comparisons of timbre were possible up to at least 18 semitones' *F0* difference, corresponding to one and a half octaves. However, dissimilarities appeared to contain a term that increased proportionally with the difference in *F0*. This showed up in MDS solutions as a dimension correlated with *F0* and orthogonal to that correlated with the spectral centroid. Along this dimension, the stimuli were ordered according to their *F0* and not according to their chroma or consonance. The MDS solution also showed a dependency on *F0* of the dimension that covaries with spectral centroid (brightness). That dependency can be reduced by modifying the definition of the spectral centroid. Although this correction seems to be well adapted to the data of this experiment, it needs to be tested and validated over a wider range of stimuli and a wider range of *F0*. If this correction is confirmed, it could be useful in applications that use timbre descriptors to discriminate, categorize, or generate instrumental sounds.

ACKNOWLEDGMENTS

This work was performed in partial fulfillment of the Ph.D. obligations of the first author. The project was funded in part by the Swiss National Science Foundation and by the European Union Project CUIDADO. It was conducted within the Music Perception and Cognition team at IRCAM. The authors wish to thank Diana Deutsch and two anonymous reviewers for their helpful comments.

Borg, I., and Groenen, P. J. F., (**1997**). *Modern Multidimensional Scaling: Theory and Applications* (Springer, New York).

Hartmann, W. M., (**1998**). *Signals, Sound, and Sensation* (Springer, New York).

Hoemeke, K. A., and Diehl, R. L. (**1994**). "Perception of vowel height: the role of *F1-F0* distance," J. Acoust. Soc. Am. **96**, 661–674.

Killion, M. C., (**1978**). "Revised estimate of minimum audible pressure: Where is the missing 6 dB?," J. Acoust. Soc. Am. **63**, 1501–1508.

Marozeau, J., (**2004**). "L'Effet de la Fréquence Fondamentale sur le Timbre," (The Effect of the Fundamental Frequency on Timbre) Ph.D dissertation. University Pierre et Marie Curie, Paris VI (Lulu. com, Paris).

Marozeau, J., de Cheveigné, A., McAdams, S., and Winsberg, S., (**2003**). "The dependency of timbre on fundamental frequency," J. Acoust. Soc. Am. **114**, 2946–2957.

Moore, B. C. J., (**2003**). *An Introduction to the Psychology of Hearing* (Academic, San Diego, CA).

Nearey, T. M., (**1989**). "Static, dynamic, and relational properties in vowel perception," J. Acoust. Soc. Am. **85**, 2088–2113.

Patterson, R. D., Robinson, K., Holdsworth, J., McKeown, D., Zhang, C., and Allerhand, M., (**1992**). "Complex sounds and auditory images," in *Auditory Physiology and Perception*, edited by Y. Cazals, L. Demany, and L. Horner (Pergamon, Oxford).

Schonemann, P. H., (**1983**). "Some theory and results for metrics of bounded response scales," J. Math. Psychol. **27**, 311–324.

Shepard, R. N., (**1982**). "Geometrical approximations to the structure of musical pitch," Psychol. Rev. **89**, 305–333.

Shepard, R. N., (**1999**). "Structural representation of musical pitch," in *The Psychology of Music*, edited by D. Deutsch (Academic, New York).

Slawson, A. W., (**1968**). "Vowel quality and musical timbre as functions of spectrum envelope and fundamental frequency," J. Acoust. Soc. Am. **43**, 87–101.

Traunmuller, H., (**1981**). "Perceptual dimension of openness in vowels," J. Acoust. Soc. Am. **69**, 1465–1475.

Evaluation of multichannel reproduced sound: Scaling auditory attributes underlying listener preference

Sylvain Choisel
Sound Quality Research Unit, Department of Acoustics, Aalborg University, 9220 Aalborg, Denmark
Bang & Olufsen A/S, Peter Bangs vej 15, 7600 Struer, Denmark

Florian Wickelmaier
Sound Quality Research Unit, Department of Acoustics, Aalborg University, 9220 Aalborg, Denmark

(Received 18 November 2005; revised 10 October 2006; accepted 11 October 2006)

A study was conducted with the goal of quantifying auditory attributes that underlie listener preference for multichannel reproduced sound. Short musical excerpts were presented in mono, stereo, and several multichannel formats to a panel of 40 selected listeners. Scaling of auditory attributes, as well as overall preference, was based on consistency tests of binary paired-comparison judgments and on modeling the choice frequencies using probabilistic choice models. As a result, the preferences of nonexpert listeners could be measured reliably at a ratio scale level. Principal components derived from the quantified attributes predict overall preference well. The findings allow for some generalizations within musical program genres regarding the perception of and preference for certain spatial reproduction modes, but for limited generalizations across selections from different musical genres. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2385043]

PACS number(s): 43.66.Lj, 43.66.Ba, 43.38.Md, 43.38.Vk [AK] Pages: 388–400

I. INTRODUCTION

One of the goals of research in sound quality is to understand the mechanisms underlying listener preference. Complex stimuli are typically involved in sound quality assessments, giving rise to various sensations, or *auditory attributes*, which potentially contribute to perceived overall quality. The identification and quantification of these sensations are necessary before their relation to preference can be established.

Apart from pioneering studies on multichannel recording and playback (Nakayama *et al.*, 1971), most work on quality of reproduced sound has focused on timbral aspects of monophonic reproduction (e.g., Gabrielsson and Sjögren, 1979). As multichannel audio formats are growing in popularity, the question arises how the various reproduction modes influence the listener's perception. Of particular interest is how spatial auditory sensations are affected by the introduction of center and surround loudspeakers in a multichannel setup (ITU-R BS.775-1, 1994), or by various processing algorithms. More recent studies have addressed the problem of identifying and quantifying auditory attributes that are relevant to sound quality in the context of multichannel reproduced sound (Rumsey, 1998; Berg and Rumsey, 2006; Zacharov and Koivuniemi, 2001; Guastavino and Katz, 2004). The first three employed combinations of recording and playback techniques to evoke various auditory sensations, and the latter used Ambisonics (Gerzon, 1985), a versatile recording and playback technique in which the sound signals are optimally decoded for each loudspeaker configuration.

By contrast, the present study aimed at investigating more specifically the perceptual differences between reproduction modes typically encountered in home audio systems: Selected musical excerpts—originally produced for five-channel reproduction—were reproduced in various formats (mono, stereo, and several multichannel formats). In a recent study, Zieliński *et al.* (2003) have focused on the overall perceptual evaluation—the so-called *basic audio quality*, defined in ITU-R BS.1116 (1997)—of reproduction modes similar to the ones used in the present work. Rumsey *et al.* (2005) investigated the influence of timbral, frontal, and surround fidelity changes on basic audio quality. The present investigation, however, intended to seek explanations for such global differences in terms of more specific auditory attributes. It was part of a larger-scale study, the goals of which were to (1) identify the auditory attributes that are relevant in the context of multichannel music reproduction, (2) verify that listeners can judge upon them in a consistent manner, (3) quantify them on meaningful scales, and (4) determine their relation to overall preference. The identification of attributes relevant for this study has been reported elsewhere (Choisel and Wickelmaier, 2006a), and in the present paper emphasis is placed on the remaining three goals.

In all the earlier investigations cited above, auditory attributes and/or overall quality were directly estimated using rating scales with either numerical or verbal labels, or graphical (visual analog) scales. Such direct scaling procedures are the *de-facto* standard in sound quality assessments. As an example, consider the ITU-T recommendation P.800 (1996) for transmission quality, or the ITU-R recommendation for small (ITU-R BS.1116, 1997) and intermediate (ITU-R BS.1534, 2003) impairments in audio systems. The validity of such scales, however, relies on many implicit and untested assumptions.

First, it is usually assumed that the order of the scale values corresponds to an order of the sounds along the investigated attribute. This is problematic, at least for multidimensional stimuli, because subjects might not be able to

 0001-4966/2007/121(1)/388/13/$23.00

combine the different dimensions into a single one (e.g., overall quality). Classical studies on human choice behavior (May, 1954; Tversky, 1969) have demonstrated that two or three dimensions already lead to predictable inconsistencies. Focusing on different aspects depending on the stimuli being compared can result in intransitive judgments, such as preferring stimulus A over B, B over C, but C over A. It is evident that a preference order of the stimuli cannot be established in this case. While paired comparisons easily reveal these intransitivities, in direct scaling procedures problems associated with multidimensionality will go unnoticed, which casts doubt on the validity of such directly obtained scales. Very often researchers are interested, not only in an order of the stimuli, but also in information about their differences or ratios, which requires measurements on higher scale types (interval or ratio scales; Stevens, 1946). The higher the scale level, the more restrictive forms of transitivity (as will be defined later in this paper) must be fulfilled.

Another assumption is the subjects' ability to map their sensation magnitude onto a scale. Often the freedom to choose among the many response categories in direct scaling procedures will result in an idiosyncratic strategy of scale usage. Some subjects might display a bias for certain response categories, for example, the center or the end points of the scale. Methods to deal with scale-usage heterogeneity exist (e.g., Rossi *et al.*, 2001), but they employ involved statistical procedures and are therefore rarely used in practice. Binary paired comparisons, on the other hand, require nothing but simple comparative judgments, and thereby eliminate response biases due to scale usage.

Therefore, a major methodological objective of the present work was to use well-founded scaling techniques based on paired comparisons (so-called probabilistic choice models; Luce, 1959; Tversky, 1972). Such scaling methods have been successfully applied to sound quality evaluation, most notably to auditory unpleasantness (Ellermeier *et al.*, 2004; Zimmer *et al.*, 2004). In the present study, probabilistic choice models are employed both for determining the overall preference and for measuring the strength of more basic auditory attributes, thereby verifying that listeners could judge upon them in a consistent manner. Subsequently, the resulting scale values are applied to formulate an exploratory statistical model in which preference is related to the auditory attributes.

II. METHOD

A. Apparatus and stimuli

1. Experimental setup

The listening tests took place in a 60 m^2 sound-insulated listening room complying with the ITU-R BS.1116 (1997) requirements. Seven loudspeakers (Genelec 1031A) were placed as shown in Fig. 1, at a distance of 2.5 m to the listening position. The height of the tweeters was 108 cm above the floor, which corresponds the average height of the entrance of the listeners' ear canals when seated. Five of the seven loudspeakers were arranged in accordance with the ITU-R recommendation BS.775-1 (1994); two additional speakers (LL and RR) were placed at ±45° for the reproduction of stereo over a wider base angle (defined as the bearing angle between the loudspeaker pair, as seen from the listening position). The setup was hidden from the subject by an acoustically transparent curtain.

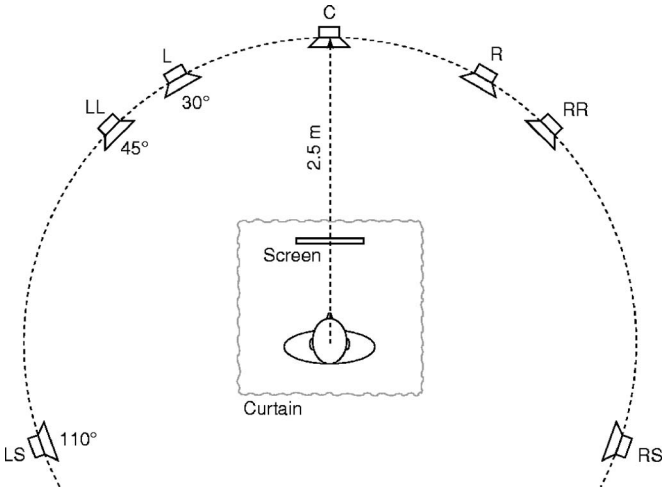


FIG. 1. Playback setup consisting of seven loudspeakers: left (L), right (R), center (C), left-of-left (LL), right-of-right (RR), left surround (LS), and right surround (RS). This setup was symmetrically placed with respect to the width of the room and was hidden from the subject by an acoustically transparent curtain. A computer flat screen was used as a response interface.

The sounds were played back by a computer placed in the control room, equipped with a multichannel sound card (RME Hammerfall HDSP) connected to an eight-channel D/A converter (RME ADI-8 DS) having a flat frequency response from 5 Hz to 21.5 kHz.

The response interface consisted of an optical mouse and a 15 in. flat screen placed in front of the listener, below the loudspeaker level (45 cm above the floor) in order to limit interactions with the sound field. A headrest fixed to the armchair ensured that the subject's head was always centered during the listening test. The head position could be monitored from the control room, via a camera attached to the ceiling above the listener.

The seven loudspeakers were matched in sensitivity and minimum-phase frequency response based on impulse response measurements carried out in an anechoic chamber using a 14th order maximum length sequence (MLS) at 48 kHz. The equalization was implemented as FIR filters applied to the corresponding channels in the sound files. In order to verify that the interchannel level alignment was preserved in the listening room, the A-weighted sound-pressure level of bandpassed pink noise (200 Hz–2 kHz) was measured at the listening position for each channel. As a result, the interchannel level differences were within 0.3 dB, and the differences between left/right pairs did not exceed 0.1 dB.

2. Program material

Four musical excerpts (two pop, two classical) were selected from commercially available multichannel material (Table I). Their different musical contents (genre, instrumental versus vocal) as well as the various spatial information present in the multichannel mix (natural room reverb in the

TABLE I. List of musical program material.

Disc	Title	Medium	Track	Time
Beethoven: Piano Sonatas Nos. 21, 23 & 26 – Kodama	Sonata 21, op. 53 (Rondo)	SACD	03	1'51–1'56
Rachmaninov: Vespers – St. Petersburg Chamber Choir conducted by Korniev	Blazen Muzh	SACD	03	2'04–2'09
Steely Dan: Everything Must Go	Everything Must Go	DVD-A	09	0'52–0'57
Sting: Sacred Love	Stolen Car	SACD	06	1'55–2'00

classical recordings, or distributed instruments in pop music) made this selection suitable for eliciting different spatial sensations. The two classical recordings were made with five omnidirectional microphones placed in a circular array, and the two pop recordings were mixed with standard surround panning technique. These excerpts were transferred from their original medium—Super Audio Compact Disc (SACD) or Digital Versatile Disc–Audio (DVD-A)—onto a computer (48 KHz, 24 bit) using a Denon 2200 player connected to an eight-channel A/D converter (RME ADI-8 DS), and carefully cut to include a musical phrase, their duration ranging from 4.7 to 5.4 s.

3. Reproduction modes

From the original five-channel program material (or), seven additional formats were derived, as summarized in Table II. This selection of reproduction modes was made in order to create a wide range of perceptual changes typically encountered in home audio applications. First, the original was mixed down to stereo (st) according to the ITU-R recommendation BS.775-1 (1994):

$$L_{\mathrm{st}} = L_{\mathrm{or}} + \tfrac{1}{\sqrt{2}} C_{\mathrm{or}} + \tfrac{1}{\sqrt{2}} LS_{\mathrm{or}},$$

$$R_{\mathrm{st}} = R_{\mathrm{or}} + \tfrac{1}{\sqrt{2}} C_{\mathrm{or}} + \tfrac{1}{\sqrt{2}} RS_{\mathrm{or}}. \quad (1)$$

From the stereo version, mono (mo) and phantom mono (ph) were computed as described by Eqs. (2) and (3), respectively.

$$C_{\mathrm{mo}} = \tfrac{1}{\sqrt{2}}(L_{\mathrm{st}} + R_{\mathrm{st}}), \quad (2)$$

TABLE II. Reproduction modes: full name, abbreviation, and loudspeakers used for playback (see Fig. 1).

Name	Abbr.	Speakers
Mono	mo	C
Phantom mono	ph	L,R
Stereo	st	L,R
Wide stereo	ws	LL,RR
Matrix upmixing	ma	L,R,LS,RS
Dolby Pro Logic II	–[a]	L,R,C,LS,RS
DTS Neo:6	–[a]	L,R,C,LS,RS
Original 5.0	or	L,R,C,LS,RS

[a]Referred to as u1 and u2 (in no specific order) in the rest of this paper.

$$L_{\mathrm{ph}} = R_{\mathrm{ph}} = \frac{1}{2}(L_{\mathrm{st}} + R_{\mathrm{st}}). \quad (3)$$

The wide stereo format (ws) was identical to stereo, but played on loudspeakers LL and RR, positioned at ±45°. All processing was done in MATLAB using floating point precision, and all intermediate files were stored with 24-bit resolution.

Finally, three upmixing algorithms were used to reconstruct multichannel sound from the stereo downmix; two commercially available algorithms, Dolby Pro Logic II and DTS Neo:6—later referred to as upmixing 1 and 2 (u1 and u2), in no specific order—and a simple matrix upmixing algorithm. Dolby Pro Logic II was implemented on a surround processor (Meridian 861) that was fed with a digital signal (S/PDIF) from the RME sound card, and the five analog output signals were recorded through the RME converter, using 24-bit resolution. In a similar fashion, an audio/video receiver (Yamaha RX-V 640) was used to generate the DTS Neo:6 upmix. The matrix upmixing (ma) was inspired by matrix decoding systems that are typically applied to encoded stereo tracks (cf. Rumsey, 2001). In this study, however, it was applied to a "regular" stereo downmix [Eq. (1)]. The upmixing was implemented in MATLAB in the following way: The left and right surround channels were fed with the difference between the left and right signals ($L-R$ and $R-L$, respectively) attenuated by 6 dB. The front (L and R) channels were left unchanged.

The eight reproduction modes were matched in loudness by eight subjects (not taking part in the main experiments) using a forced-choice adaptive procedure (2AFC, 1-up/1-down, cf. Levitt, 1971; Jesteadt, 1980). On each trial, the task was to decide which of the two presented sounds was louder, one being the standard, the other one being the comparison, in random order. For all four types of program material (Beethoven, Rachmaninov, Steely Dan, and Sting), the standard was chosen to be the stereo reproduction mode. Its playback level was adjusted beforehand to a comfortable level by the experimenters, and measured in the listening position to have A-weighted, energy-equivalent sound pressure levels of 65.8, 59.4, 66.5, and 67.7 dB, respectively (averaged over the duration of the stimuli). The loudness matching procedure was reported in more details in Choisel and Wickelmaier (2006a). The resulting loudness matches were averaged across subjects, and appropriate gains were applied to the stimuli. After equalization and loudness matching, all sounds were saved as multichannel wave files, dithered, and

quantized to 16 bits (±1 LSB triangular probability density function) and with a sampling frequency of 48 kHz. Further information regarding acoustical and psychoacoustical characteristics of the reproduction modes may be found in Choisel and Wickelmaier (2006b).

B. Subjects

Forty listeners (28 male, 12 female) took part in this study. They were mostly university students, had no or little prior experience with this type of experiment, and were naïve with respect to the research questions. In contrast to expert listeners trained to identify subtle differences in a reliable manner, the selected sample was closer to a consumer population. Such a sample was chosen in order to avoid possible biases, such as an excessive display of knowledge during the identification of auditory attributes. It was, however, desired that they possess the ability to perform the tasks required from them. For that purpose, these participants were selected among 78 candidates, according to their auditory and verbal aptitudes. The selection procedure (detailed in Wickelmaier and Choisel, 2005) consisted of pure-tone audiometry, a stereo-width discrimination task, and a verbal-fluency test. These tests were performed in order to ensure that the listeners selected could (1) appreciate spatial differences in sound and (2) readily produce a description of their sensations. All candidates were native Danish speakers, without any known hearing problems. Eight listeners showing a hearing threshold of more than 20 dB HL (*re*. ISO 389-1, 1998) in any ear at any frequency between 250 Hz and 8 kHz were rejected based on this criterion. From the remaining 70, the 40 listeners performing the best in the other two tests (stereo-width discrimination and verbal fluency) were selected to participate in the main experiments. Their age ranged from 21 to 39 years (median=24 years). One of the participants dropped out during the first part of the experiment, the remaining 39 took part in the complete study that extended over approximately six months.

C. Procedure

The study was organized in several larger experimental parts, throughout which the same sample of subjects and the same set of stimuli were used. First, an overall preference was determined for the reproduction modes. Next, auditory attributes salient in the context of these sounds were identified (elicited) using the same sample of subjects; this part is reported in Choisel and Wickelmaier (2006a); the outcome was a set of eight attributes: *width, elevation, spaciousness, envelopment, distance, brightness, clarity, and naturalness*. Subsequently, the strength of these attributes was quantified. Finally, the preference was reevaluated.

1. Quantification of auditory attributes

Quantification of the attributes was carried out by asking the subjects (in Danish) "Which of the two sounds is more…" followed by one of the following adjectives: *wide (bred), elevated (høj oppe), spacious (rummelig), enveloping (omsluttende), far ahead (langt foran), bright (lys), clear (tydelig)* and *natural (naturlig)*. Definitions of these attributes, generated by the authors so as to represent as much as possible the subjects' own descriptors, can be found in the Appendix .

For each of the eight attributes and for four musical excerpts, all possible pairs of reproduction modes were presented to the subjects. Two buttons on a computer screen, labeled A and B, were visually emphasized in turn (by changing their size) during playback to indicate which sound was played. The response was made by clicking the button corresponding to the chosen sound. Each pair was judged only once. The within-pair order was balanced across subjects (David, 1988, Chap. 5) and the between-pair order was random. Each attribute was evaluated for all four program materials in a single block lasting for about 25 min. Each subject evaluated two attributes in a session lasting for one hour, including a break in the middle. Thus, four sessions were required for all eight attributes. The order of the attributes and program materials was balanced across subjects using five different 8×8 Graeco-Latin squares. Each subject gave 28 judgments per program material and auditory attribute.

2. Quantification of overall preference

It was hypothesized that as the study proceeded (especially by taking part in the attribute elicitation, Choisel and Wickelmaier, 2006a) participants would gain experience with the sounds, which potentially influenced their perception. In order to investigate the influence of experience, preference was measured at two points in time: once at the beginning (first measurement) and once at the end of the study (second measurement, about six months after the first data collection). For each pair of reproduction modes the subjects were instructed to indicate which one they preferred. In the first data collection, each pair was presented in both within-pair orders (AB and BA), and a third time in one within-pair order, counterbalanced across subjects. The second data collection on preference only included two judgments per pair (both within-pair orders). Thus, each subject gave 84 (respectively, 56) preference judgments per program material in the first (respectively, second) data collection.

D. Analysis of choice frequencies

Both, for overall preference and for the selected auditory attributes, the pairwise choices among the eight reproduction modes were aggregated across all listeners, resulting in matrices of choice frequencies. In such a matrix it can be seen how often, for example, mono (mo) reproduction was chosen to be more *spacious* than stereo (st), and vice versa. From these frequencies the probability, P_{xy}, of choosing sound x over sound y according to a given criterion was estimated.

The derivation of scales from the choice frequencies crucially depends on the consistency of the judgments given by the subjects. Consistency was analyzed by testing weak (WST), moderate (MST), and strong (SST) stochastic transitivities, which imply that if $P_{xy} \geq 0.5$ and $P_{yz} \geq 0.5$, then

$$P_{xz} \geq \begin{cases} 0.5 & \text{(WST)}, \\ \min\{P_{xy}, P_{yz}\} & \text{(MST)}, \\ \max\{P_{xy}, P_{yz}\} & \text{(SST)}, \end{cases} \tag{4}$$

for all sounds x, y, and z. Whenever the premise holds, but the implication in Eq. (4) does not hold (for any permutation of the triple x,y,z), a transitivity violation is observed. Violations of the different transitivities are of different severity. A systematic violation of WST indicates that subjects were not able to integrate several stimulus dimensions into one (common) percept, and it is therefore impossible to even derive a meaningful ordering of the sounds. Less severe are violations of SST, which suggest a certain context dependency of the choices made. Such a context dependency usually comes into play when there are subgroups of similar sounds based on multiple perceptually salient aspects or features (Carroll and De Soete, 1991).

Transitivity violations might result from either individually inconsistent choice behavior or from subjects disagreeing in their choices. The probabilistic choice models reported in this paper were applied to aggregate data, and it is therefore not possible to separate these two sources of inconsistencies. Models that account for individual differences have been developed (e. g., Böckenholt, 2001), but are not within the scope of the present study.

Counting the number of transitivity violations in a matrix of choice frequencies only yields a descriptive measure of (in)consistency. In an experiment with a limited number of observations, it is conceivable that violations occur at random; a statistical test is therefore required to classify such violations as either systematic, and thus critical, or random.

Two kinds of probabilistic choice models were considered for representing the choice frequencies, with the goal of (1) providing a statistical evaluation of the transitivity violations encountered, and (2) in the presence of only random violations, quantifying the attribute in question. The first model used was the Bradley-Terry-Luce (BTL) model (Bradley and Terry, 1952; Luce, 1959), which predicts P_{xy} as a function of parameters associated with each sound,

$$P_{xy} = \frac{u(x)}{u(x) + u(y)}, \tag{5}$$

where $u(\cdot)$ is a ratio scale of the criterion. Since Eq. (5) implies SST, systematic violations of SST preclude a BTL representation.

The second, less restrictive, model was the so-called *elimination-by-aspects* (EBA) model (Tversky, 1972; Tversky and Sattath, 1979), which is a generalization of the BTL model. According to EBA, one sound is chosen over a second one because of a certain *aspect* that belongs to the first but not to the second sound. EBA predicts P_{xy} by

$$P_{xy} = \frac{\sum_{\alpha \in x' \setminus y'} u(\alpha)}{\sum_{\alpha \in x' \setminus y'} u(\alpha) + \sum_{\beta \in y' \setminus x'} u(\beta)}, \tag{6}$$

where $\alpha, \beta, \ldots$, are the aspects (or features) of the sounds, x' indicates the set of aspects belonging to sound x, and $x' \setminus y'$ denotes the set of aspects belonging to sound x but not to sound y. As in the BTL model, $u(\cdot)$ is a ratio scale of the criterion. EBA only implies MST, and can therefore to some extent cope with multiple-aspect criteria.

The goodness of fit of the choice models was evaluated by comparing the likelihood L_0 of a given (restricted) model to the likelihood L of a saturated (unrestricted) binomial model which perfectly fits the choice frequencies, under the assumption of independent choices. The test statistic, $-2 \log (L_0/L)$, is approximately χ^2 distributed with as many degrees of freedom as the difference in parameters of the two models. A significant likelihood ratio test indicates lack of fit of the restricted choice model, and thereby that the violations of the corresponding stochastic transitivity have been systematic rather than random. If the fit was adequate, scale values for the reproduction modes were derived. Parameter estimation and model testing were performed using software described in Wickelmaier and Schmid (2004).

Probabilistic choice models provide a powerful method for scaling suprathreshold sensations, not only because they allow for *testing* the validity of a scale of a certain attribute (rather than *assuming* it when using direct scaling procedures), but also because these models enable the investigator to test hypotheses about perceived magnitudes in the framework of standard statistical theory. In order to test whether there was a significant change in the scale values of the reproduction modes in different conditions, for example, whether the preference changed between the two times of data collection (before and after elicitation and scaling of the attributes), standard likelihood ratio tests (McCullagh and Nelder, 1989) were performed. The logic of these tests is to investigate if restricting the parameters to be equal in both conditions entails a significant lack of fit, which implies that the conditions have a significant effect on the scale values. This would mean in the example that the preferences have changed from the first to the second measurement. A likelihood ratio test is possible whenever two models are *nested*, that is, one model results from the other one by applying restrictions on its parameters. A significant likelihood ratio test denotes that the restricted model is to be rejected.

III. RESULTS

A. Scaling listener preference

Table III displays the evaluation of the stochastic transitivities [Eq. (4)] of the preference judgments collected before and after the subjects went through the elicitation and scaling of specific auditory attributes. For the evaluation, data were aggregated over all subjects and repetitions, within each type of program material. Thus, the choice probabilities in the first measurement were estimated based on $N=40\times3=120$ observations per stimulus pair for Steely Dan, and on $N=39\times3=117$ for the other program materials, since one subject left the experiment after the first session. In the second measurement, where two replicates were collected, the choice probabilities were based on $N=39\times2=78$ observations. Weak and moderate stochastic transitivities were found to be violated either in none or in very few of the 56 possible tests, indicating that the participants were able to integrate

TABLE III. Transitivity violations and goodness-of-fit test of the BTL model for preference judgments at two points in the study: before and after elicitation and scaling of attributes. Displayed are the number of violations of weak, moderate, and strong stochastic transitivity [Eq. (4)], and the test statistic and p-value of a likelihood ratio test with the null hypothesis that the BTL model holds.

	First measurement					Second measurement				
Excerpt	WST	MST	SST	$\chi^2(21)$	p	WST	MST	SST	$\chi^2(21)$	p
Beethoven	0	2	14	9.13	0.988	0	1	12	9.06	0.989
Rachmaninov	2	4	19	16.96	0.714	0	0	18	8.44	0.993
Steely Dan	0	0	12	18.13	0.640	0	0	11	17.74	0.666
Sting	0	0	13	10.72	0.968	0	0	9	13.66	0.884

their various sensations into a unidimensional preference judgment. Consequently, at least an ordinal preference scale may be derived from the choice frequencies.

In order to evaluate the more frequent violations of SST, and to test whether a preference *ratio* scale could be obtained, a BTL model [Eq. (5)] was fitted to the paired-comparison data. Table III shows the results of the goodness-of-fit tests which support the validity of the model in each of the four program material conditions. Accordingly, the SST violations were classified as random, and preference scales were extracted. Consequently, it was possible to measure listener preference at a ratio scale level using the very simple, but very restrictive BTL model.

The reliability of the judgments was assessed by testing whether there were any changes of preference between the three (respectively, two) repetitions, *within* each measurement. Likelihood ratio tests were devised to compare a BTL model, which allows for preference changes to one with a fixed set of parameters across repetitions. Neither in the first nor in the second measurement, however, did the fixed-parameter model fit significantly worse than the model having variable parameters; this was true for all types of program material. Therefore, the preference values of the reproduction modes can be regarded constant throughout the repetitions within each measurement. This indicates a high degree of reliability of the preference judgments.

Figure 2 displays the parameter estimates of the BTL model, i.e., the *preference scales*, for the four program materials obtained in the two measurements, together with the 95%-confidence intervals. The preference ratio scales are plotted on logarithmic y axes in order to facilitate the comparison among reproduction modes. For example, two-channel phantom mono (ph) was preferred about twice as much as the single channel mono (mo) for the Beethoven excerpt. About the same ratio was observed between wide-angle stereo (ws) and one of the upmixing algorithms (u2). Since the BTL parameters are unique up to multiplication by a positive constant, they were normalized to sum to unity. Consequently, the distance from the line of indifference ($u = 1/8$, which would be the location of the scale values if all pairwise choice frequencies were 0.5) indicates how pronounced the preferences are between the reproduction modes. In all conditions, equality of the scale values can be rejected, which suggests that listeners were far from indifferent, but had rather strong preferences for certain reproduction modes.

Across the four program materials and the two points of measurement, it was observed that mono reproduction (mo) and (ph) was inferior to the other formats. Stereo, on the other hand, was generally among the most preferred, whereas

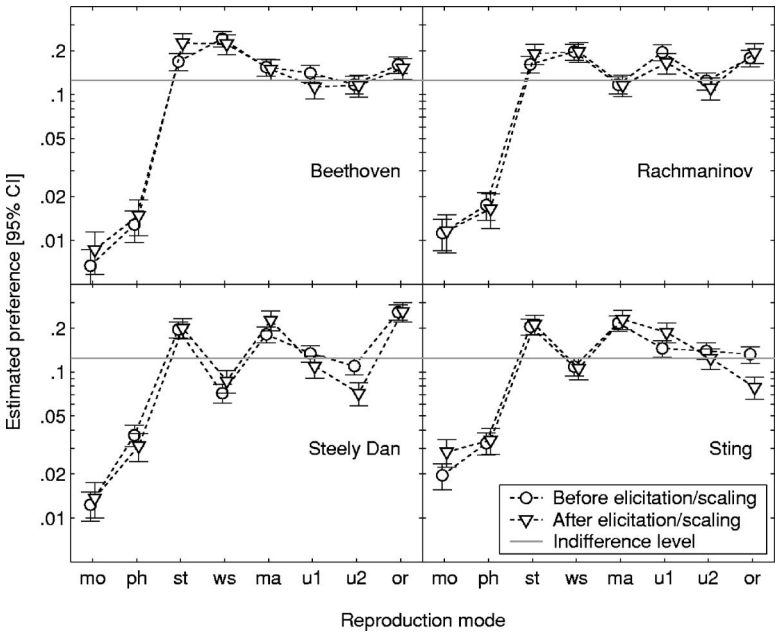


FIG. 2. Ratio scale of preference derived from eight reproduction modes for four musical excerpts. Scale values represent parameter estimates of the BTL model fitted to paired-comparison judgments. Preference was measured at two points in the study (see text), represented by two symbol styles. The reproduction modes were mono (mo), phantom mono (ph), stereo (st), wide-angle stereo (ws), four- (ma) and five-channel upmixing (u1 and u2), and the original five-channel material (or). Error bars show 95%-confidence intervals.

the original five-channel material outperformed stereo only once (Steely Dan). A further interesting relation was observed between ws and matrix upmixing (ma) when comparing classical and pop music. While it was beneficial for the classical music to increase the stereo base angle from 60° to 90°, this had adverse effects for the pop excerpts. Conversely, while ma was less preferred than ws for the classical music, it was favored over ws for the pop music. The results so far indicate common preference patterns, at least within a musical genre, but also excerpt-specific effects.

In a further set of likelihood ratio tests, it was investigated to what extent the preference scales were generalizable across program materials. In spite of the obvious similarities within the classical and the pop genres (see the rows in Fig. 2), the excerpt-specific differences were statistically significant. In the first measurement, a common model for Beethoven and Rachmaninov fared significantly worse $[\chi^2(7)=30.81; p<0.001]$ than a model having two sets of parameters: one for each program material. The same was true for Steely Dan and Sting $[\chi^2(7)=78.15; p<0.001]$. Analogous results were obtained in the second measurement for classical $[\chi^2(7)=21.14; p=0.004]$ and pop music $[\chi^2(7)=146.01; p<0.001]$, respectively. From the magnitudes of the test statistics, it seems that the differences between the classical excerpts were not as striking as between the pop excerpts, and the difference between Steely Dan and Sting even increased in the second measurement. Therefore, the generalizability of the results concerning the preference for certain reproduction modes should not be overestimated, since the dependence on the program material is evident.

Since preference data were collected twice for the same listeners, once *before* elicitation and quantification of the more specific attributes and once *after* that, the effect of experience (with the sounds) on preference may be examined. Figure 2 suggests that there is a close correspondence between the preference scales obtained at the two points in time, indicating that preference was relatively stable even over a period of about six months. Again, likelihood ratio tests were employed for the statistical analyses. This time, it was tested for each type of program material, whether the preference scale had changed between the first and the second measurement. No, significant changes were observed for Beethoven $[\chi^2(7)=12.33; p=0.090]$ and Rachmaninov $[\chi^2(7)=5.90; p=0.551]$, whereas for Steely Dan $[\chi^2(7)=25.37; p=0.001]$ and Sting $[\chi^2;(7)=35.80; p<0.001]$ the changes were significant. These differences might be attributed to listeners becoming more sensitive to subtle differences between the reproduction modes. For example, there were no significant preference differences between the two upmixing algorithms (u1 and u2) and the original five-channel Sting material or in the first measurement (see the bottom right panel in Fig. 2). In the second measurement, however, the ratio between u1 and or extended to about 3:1. A similar argument holds for the ma and u2 reproduction modes of the Steely Dan excerpt (see the bottom left panel in Fig. 2).

TABLE IV. Transitivity violations and goodness-of-fit test of the BTL model for selected attributes. See Table III. Note: $^*p<0.05$.

Attribute	WST	MST	SST	$\chi^2(21)$	p
Beethoven					
Width	0	1	19	24.55	0.267
Elevation	1	11	25	24.63	0.263
Spaciousness	0	2	18	17.80	0.661
Envelopment	0	3	23	22.16	0.391
Distance	3	9	32	22.83	0.353
Brightness	2	3	19	12.25	0.933
Clarity	4	5	27	25.55	0.224
Naturalness	3	5	24	15.41	0.802
Rachmaninov					
Width	1	1	14	21.20	0.447
Elevation	2	7	23	16.08	0.765
Spaciousness	2	7	19	7.35	0.997
Envelopment	2	4	27	16.82	0.722
Distance	2	11	37	21.74	0.414
Brightness	4	4	27	14.49	0.848
Clarity	2	6	21	8.86	0.990
Naturalness	0	2	14	16.46	0.744
Steely Dan					
Width	0	3	14	36.01	0.022*
Elevation	0	2	24	30.64	0.080
Spaciousness	2	2	19	26.66	0.182
Envelopment	0	2	23	39.40	0.009*
Distance	3	13	30	15.89	0.776
Brightness	0	0	15	20.39	0.496
Clarity	0	2	18	14.05	0.867
Naturalness	0	2	18	14.35	0.854
Sting					
Width	0	2	24	29.47	0.103
Elevation	0	0	16	27.30	0.161
Spaciousness	0	4	16	22.60	0.366
Envelopment	1	3	16	15.04	0.821
Distance	0	1	19	21.40	0.435
Brightness	0	0	23	31.54	0.065
Clarity	2	3	16	19.24	0.570
Naturalness	1	1	18	11.72	0.947

B. Scaling auditory attributes

The same logic of consistency checks, model evaluation, and scaling was applied to the more elementary auditory attributes. Table IV displays the violations of the stochastic transittivities for each auditory attribute and program material. Since the pairwise probability estimates were based on 39 observations (every listener judged each pair only once) it was expected to see more (random) violations than for the preference judgments. From the low number of WST violations it follows that at least an ordinal scale of sensation magnitude can be derived in each condition. In order to test for systematic SST violations, a BTL model was applied and evaluated in each case. As shown in Table IV, in general, the model fit is adequate, which suggests that consistency in the judgments was sufficiently high for extracting *ratio* scales. Additional likelihood ratio tests were devised to confirm that each scale was significantly different from the case where all scale values are equal. These tests indicated that for no attribute-excerpt combination did listeners show indifference with respect to the reproduction modes.

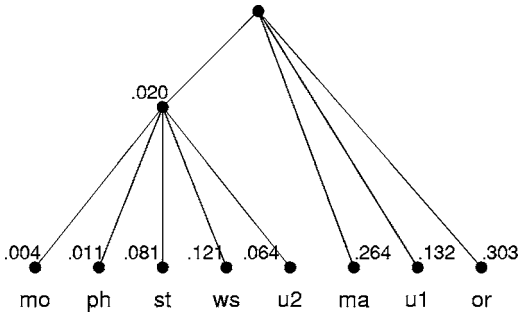


FIG. 3. Elimination-by-aspects (EBA) model structure and parameter estimates for *envelopment* (Steely Dan). Nodes represent aspects shared only by the connected reproduction modes [see Eq. (6)]. Scale values are obtained by adding up the parameters belonging to each reproduction mode.

In only two cases (Steely Dan: *envelopment* and *width*) was there a significant lack of fit of the BTL model. This should not compromise the overall conclusion that the listeners' choice behavior could be described by a simple model, since one might expect about two tests out of 32 to become significant by chance alone on an α level of 5%. The original Steely Dan material, however, was different from the other three excerpts in that it not only contains reverberation but clearly distinct sound sources (e.g., a guitar playing a staccato single-note line) in the surround channels, which might have given rise to a more complex decision strategy. Potentially, the emergence of such a new feature might be more adequately described by an elimination-by-aspects (EBA) model [Eq. (6)]. Among the EBA models with only one additional parameter, the best fitting one for *envelopment* is depicted in Fig. 3. The nodes in the graph denote the features, or *aspects*, of the reproduction modes. Apart from the top node (the aspect shared by all sounds) and the bottom nodes (the unique, individual aspects), the model includes one extra feature shared by all reproduction modes that do *not* reproduce any discrete source at the side of or behind the listener.[1] This simple EBA model was found to fit the data $[\chi^2(20)=26.55; p=0.148]$ the improvement over the BTL model being significant $[\chi^2(1)=12.85; p<0.001]$. The parameter estimates are also displayed in Fig. 3. In order to derive *envelopment* scale values from the model, the parameters belonging to each reproduction mode were added up. For example, $u(\text{mo})=0.02+0.004=0.024$. Similarly, an EBA model was found for the *width* attribute, which accounted for the data $[\chi^2(20)=27.27; p=0.128]$ and outperformed the BTL model $[\chi^2(1)=8.74; p=0.003]$. Here, four reproduction modes (st, ws, ma, and or) shared a common aspect, the interpretation of which is not so straightforward. It is worth noting that, even though these EBA models provided a better fit than the BTL model, the differences in the actual scale values were rather subtle.

Figure 4 shows the derived ratio scales for each auditory attribute and the four types of program material. Within each attribute, a considerable similarity of the scales was observed across program materials, which was even more pronounced within a musical genre (classical and pop music). For example, ws was perceived to be strongly elevated in comparison with the other reproduction modes in the pop material (Steely Dan and Sting); the effect was less distinct, but still visible, for the classical material. The stimuli showed the smallest perceptual differences with respect to *distance*; the mono sounds (mo and ph) were perceived to be nearest to the listener only for the pop music, for the classical music they were further away than most of the other reproduction modes. Except for *distance* and *brightness*, mo and ph were located at the lower end of the sensation scales, which induces correlation also *across* the attributes. Especially the correspondence between *spaciousness* and *envelopment* is

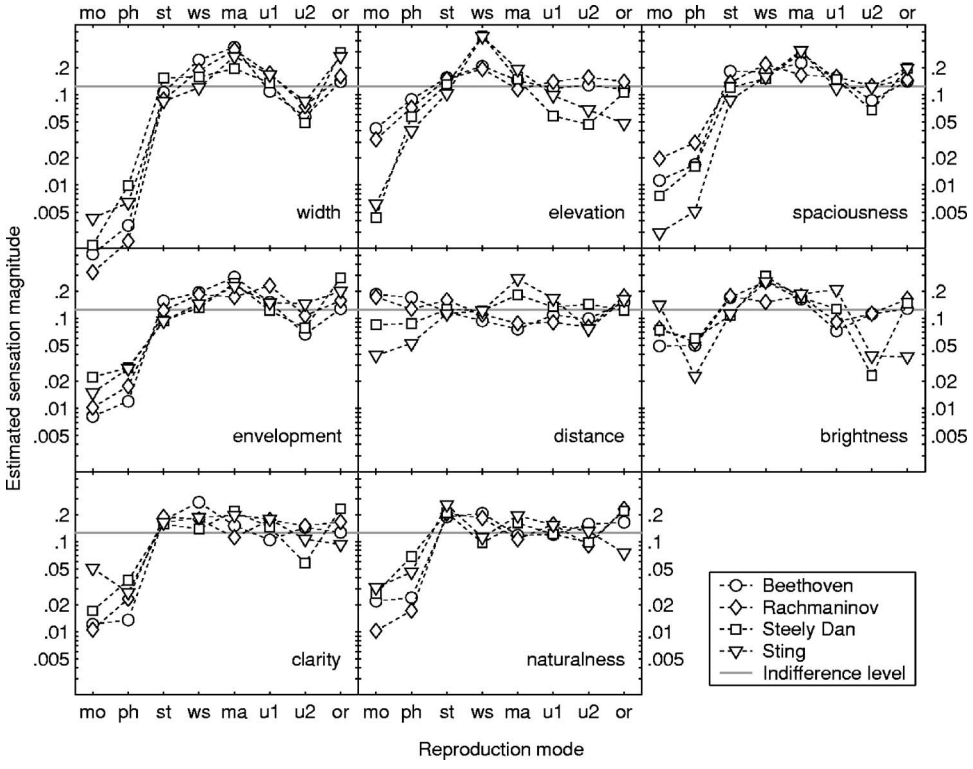


FIG. 4. Ratio scales of eight auditory attributes estimated using BTL and EBA models for four types of program material.

striking. From the high correlations it is evident that the attributes did not vary independently in the stimuli under study, or that listeners were not able to distinguish between all of them.

C. Relation between specific sensations and overall preference

A simple way of relating the auditory attributes to overall preference is multiple regression, where the predicting variables are the attribute scales and the predicted variable is the preference scale. While this approach might at first glance provide a satisfactory goodness of fit, it suffers from two shortcomings, which make it inapplicable in many situations such as the present investigation. First, the low number of reproduction modes (only eight data points to be predicted) compared to the number of possible predictors (eight attributes), makes such modeling trivial and of questionable generality. The second problem is the high correlation between some of the attributes; collinearity of the independent variables in multiple regression often yields an unstable and therefore unreliable model.

In previous studies, several methods have been used to relate the overall quality to more specific perceptual dimensions, while simultaneously addressing the problem of collinear attributes (Nakayama *et al.*, 1971; Zacharov and Koivuniemi, 2001; Mattila, 2002; Rumsey *et al.*, 2005). Such methods are based on multidimensional scaling (MDS), principal component analysis (PCA), or related techniques. PCA reduces the attributes to a few independent factors (or *principal components*) that are orthogonal by construction, as are MDS dimensions. This makes them suitable as predictors in a regression model, circumventing the problem of collinearity mentioned above.

In the present study, multiple regression based on principle components was used to predict preference. In order to increase the generalizability of the model, the data were combined within a musical genre, i.e., classical music (Beethoven and Rachmaninov) and pop music (Steely Dan and Sting), thereby doubling the number of data points to be predicted. This was justified given the similarities observed in the attribute scales across program materials (see Fig. 4). *Naturalness* was excluded from the analysis because it was considered more global than the other (more specific) attributes and not sufficiently separate from preference, the correlation between *naturalness* and preference ranging from 0.94 (Steely Dan) to 0.98 (Rachmaninov).

PCA with varimax rotation was performed on the remaining seven attributes. In the case of the classical music, 87% of the variance in the scale values was explained by the first two factors which, after rotation, accounted for 48 and 39% of the variance, respectively. For the pop music, the first two components accounted for 58 and 30% (88% cumulated) after rotation. The loadings of the attribute scales on the first two factors, calculated as correlation coefficients, are reported in Table V. Although the relationship between the attributes and the two factors is more clear cut for the pop music (because the intercorrelation between the attributes is not as strong as for the classical music), similarities can be observed between the two genres: *brightness* and *elevation* load on the same factor, while the other factor is closely related to *width, spaciousness, envelopment, and distance* (note that *distance* loads negatively for the classical music; see also Fig. 4). Thus, an analogy can be made between Factor 1 in the PCA for classical music and Factor 2 for the pop music, and vice versa, with the following exceptions: *clarity*, which loads on Factor 1 in both cases, and *spaciousness* which loads equally on both factors for the classical material. Figures 5 and 6 show a graphical representation of the attribute loadings and stimulus scores in the two-dimensional factor spaces. The coordinates of the arrow endpoints are calculated as two times the factor loadings.

TABLE V. Attribute loadings on the factors ($F1$ and $F2$) obtained from principal component analysis, and variance explained by these factors after varimax rotation. Loadings higher than 0.6 are indicated in boldface.

	Classical		Pop	
Attribute	F1	F2	F1	F2
Width	0.50	**0.75**	**0.94**	0.17
Spaciousness	**0.68**	**0.68**	**0.93**	0.26
Envelopment	0.56	**0.77**	**0.94**	0.17
Distance	–0.16	**–0.88**	**0.84**	0.13
Clarity	**0.90**	0.35	**0.78**	0.47
Brightness	**0.91**	0.24	0.24	**0.92**
Elevation	**0.83**	0.41	0.15	**0.93**
Var. explained (%)	48	39	58	30

Multiple regression was performed on the two factors (F_1 and F_2) obtained from PCA in order to predict the preference scale values (P) obtained in the second measurement (after attribute scaling). The resulting regression equations are

$$\hat{P} = 0.138 + 0.075F_1 + 0.017F_2 - 0.014F_1^2 \quad \text{(classical)}, \tag{7}$$

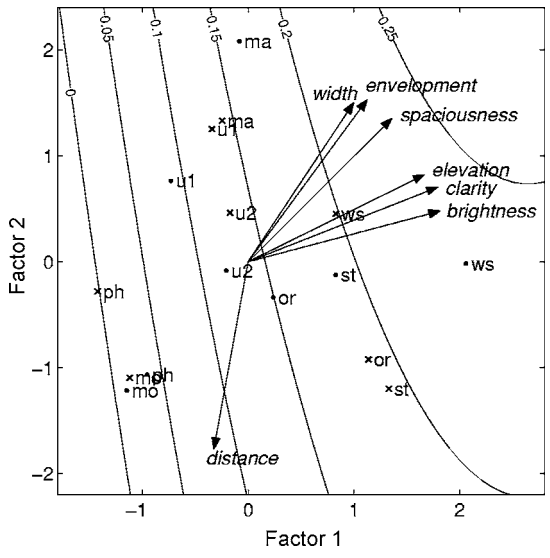


FIG. 5. Graphical representation of the factor space obtained from principal component analysis of the attribute scales, and predicted preference [Eq. (7)] for the classical music material. Factor loadings of the attributes are shown as arrows, and the scores of the reproduction modes along the two factors are represented as dots (Beethoven) or crosses (Rachmaninov). The preference estimated from the two factors is represented by contour lines.

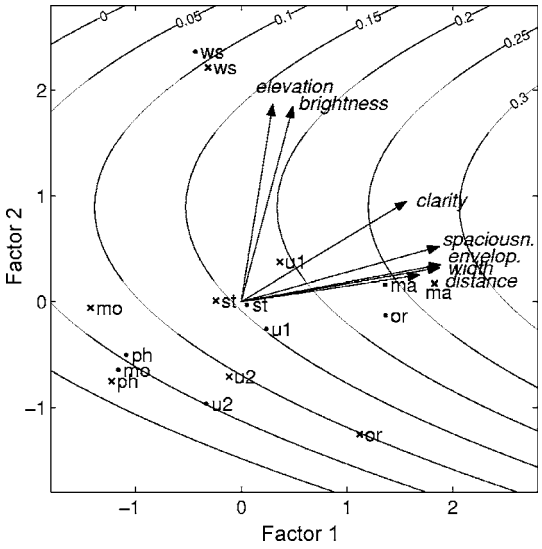


FIG. 6. Graphical representation of the factor space obtained from principal component analysis of the attribute scales, and predicted preference [Eq. (8)] for the pop music material. Factor loadings of the attributes are shown as arrows, and the scores of the reproduction modes along the two factors are represented as dots (Steely Dan) or crosses (Sting). The preference estimated from the two factors is represented by contour lines.

$$\hat{P} = 0.155 + 0.057F_1 + 0.058F_2 - 0.032F_2^2 \quad \text{(pop)}, \qquad (8)$$

all three terms in each equation being significant. In both genres, the quadratic term refers to the factor correlating with *brightness* and *elevation*, suggesting an ideal point on this dimension. The response surfaces resulting from Eqs. (7) and (8) are shown in Figs. 5 and 6 for classical and pop music, respectively. Each contour line connects points of equal preference, as predicted by the regression model. In Fig. 5, for example, the predicted preference increases when moving from the left to the upper right part of the panel. Generally, the two models were found to predict the preference quite well (Figs. 7 and 8), with a total explained variance of 94% (classical) and 84% (pop). The largest prediction errors were obtained for u1 in the classical music, and st in the pop music, both being underestimated.

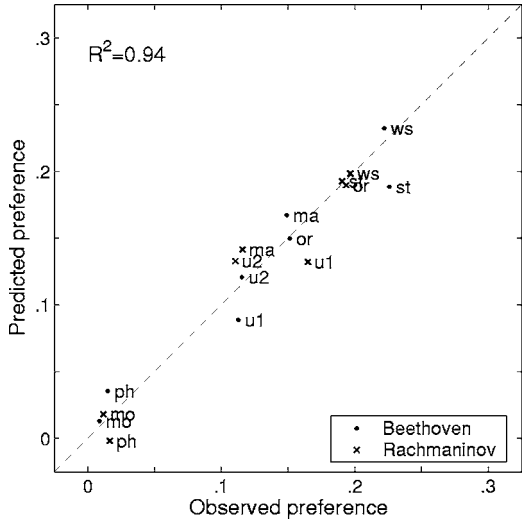


FIG. 7. Predicted [Eq. (7)] versus observed preference for the classical music material (Beethoven and Rachmaninov).

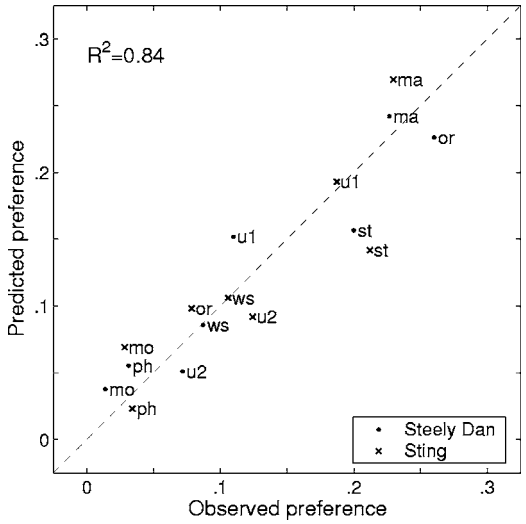


FIG. 8. Predicted [Eq. (8)] versus observed preference for the pop music material (Steely Dan and Sting).

IV. DISCUSSION

A. Scaling auditory attributes using probabilistic choice models

The quantification of attributes that play a role in the context of multichannel reproduced sound is a nontrivial problem because of the complex nature of the stimuli that typically gives rise to several timbral and spatial sensations simultaneously. From the outset, it is by no means clear that the endeavor of deriving a representation of even a single attribute (like, e.g., *spaciousness*) from listener judgments will be successful at all; inconsistent, intransitive behavior might render any numerical scale meaningless. Hence, the present study goes beyond previous work in that scales of both overall preference *and* the underlying—more basic—attributes were obtained using well-founded methodologies. Paired-comparison judgments were collected in order to allow inconsistencies to reveal themselves (which would have been impossible using direct scaling procedures). Subsequently, probabilistic choice models were employed to evaluate statistically the intransitivities encountered, and, whenever possible, to derive scales of sensation magnitude. It was demonstrated that listeners can consistently judge both upon their global preference and on more specific auditory attributes. Although the preference judgments might reasonably be assumed to be based—at least unconsciously—on many different aspects, listeners were evidently able to integrate them into a unidimensional judgment. This agrees with evidence from other fields of sound quality research, where global auditory attributes, for example the *overall unpleasantness*, have been thoroughly investigated with respect to whether listeners can make transitive judgments about heterogeneous sets of environmental sounds (Ellermeier *et al.*, 2004; Zimmer *et al.*, 2004). In the former study, a BTL model was found to represent the choice frequencies, while in the latter one, a simple EBA model was required to account for the complex stimuli. Taken together, these results suggest that it will strongly depend on the context to what extent the multiple aspects of complex stimuli pose a prob-

lem for deriving a meaningful sensation scale. To simply *assume* unidimensionality, however, is hard to justify.

It is worth noting that the aspects represented by EBA parameters might or might not have a direct correspondence to physical characteristics of the stimuli. Rather, the aspects may be viewed as perceptual effects relevant in the decision process. Furthermore, listeners might or might not be fully aware of the aspects when choosing among the sounds, that is to say that—depending on the context of stimuli—aspects may relate to more elementary sensory or higher-level cognitive mechanisms. A sensation scale derived from probabilistic choice models, therefore, reflects to more or less a degree both sensory and judgmental (cognitive) processes (see also Baird, 1997).

It is an encouraging result of the present study that the listeners' overall preference was measurable at a high scale level, and that it was stable over a period of six months. The experience that the listeners had gained during their participation in the experiments had the beneficial effect that subtle differences between the reproduction modes became more salient to them in the course of time. From the preference data collected at two points in time in this study it can be concluded that nonexpert listeners have a clear and stable concept of the versions of reproduced music to which they would prefer to listen.

The highly restrictive BTL model that implies strong stochastic transitivity was found to be less adequate for some of the rather "simple" auditory attributes—especially for *envelopment* and *width* for the Steely Dan excerpt—than for the "complex" overall preference. Therefore, it cannot always be assumed that a seemingly simple question like "How wide is the sound event?" would yield a unidimensional evaluation for any kind of stimulus. In the present case, however, it was possible to find less restrictive EBA models that accounted for the few situations in which the BTL model was violated. The model structures as well as the hypothesis that discrete sound sources in the surround channels might have been responsible for the BTL model to fail should be confirmed in further studies.

It is conceivable that inconsistencies resulting from multidimensional stimuli could be eliminated by training the listeners and breaking up the problematic attribute in several unidimensional "subattributes."[2] Probabilistic choice models therefore constitute a valuable diagnostic tool to reveal such problems, even if they are difficult to point out directly by the listeners (or even the experimenter).

B. Generalizability across program materials

For all attributes as well as for overall preference, the type of program material had a significant effect, suggesting that perceptual effects evoked by the selected reproduction modes depend on the musical signals they are applied to. Nevertheless, certain similarities can be observed across programs. For instance, it appears clearly from Fig. 4 that the effect of the reproduction mode on *width, envelopment*, and *spaciousness* is preserved across programs.

For other attributes (e.g., *elevation* and *distance*), certain patterns can be observed that distinguish the classical from the pop music selections. This is also true for preference (Fig. 2): While matrix upmixing (ma) was preferred over stereo (st) for pop music, it made it worse for the classical programs. Conversely, while increasing the stereo base angle (ws) was beneficial for classical music, it was detrimental for pop music. Bech (1998) showed that wider base angles yield higher perceived quality; however, this investigation only included angles up to ±30°. Increasing the angle to ±45° in the present study resulted in a perceived elevation of the sound sources (cf. Fig. 4), which could be the reason for the lower preference for ws in the pop music. Such an elevation effect as a function of loudspeaker base angle has been studied by Damaske (1969), and can be explained by the spectral changes introduced (Bloom, 1977), a phenomenon closely related to Blauert's (1997, Chap. 2) "boosted bands." This constitutes a plausible explanation for the high correlation observed between the attributes *elevation* and *brightness* (Fig. 4).

Finally, two observations can be made across musical genres. First, mono and phantom mono were the least preferred formats for all four types of program material. This is likely to be due to the low values on most of the spatial attributes: *width, envelopment* and *spaciousness*, as well as *clarity* and *naturalness*. Second, the overall preference for stereo reproduction was quite high in all four types of program material: For only one of the excerpts (Steely Dan) was the original five-channel reproduction preferred over stereo. This may be explained by the subjects' familiarity with two-channel stereo reproduction, or, in the case of classical music, by the only subtle changes introduced by downmixing. Zieliński *et al.* (2003) and Zieliński *et al.* (2005) reported that the perceived quality of material containing dry sources both in the front and in the surround channels (*foreground-foreground*) is more impaired by downmixing than is material containing predominantly reverberation in the surround channels (*foreground-background*). The result of the present study that the stereo downmix was less preferred than the original *only* for the foreground-foreground material (Steely Dan) supports this hypothesis. The present study, however, suggests that the original is not always judged to be of highest quality, if the subjects are not explicitly instructed to assign the maximum rating to the original (as in, e.g., Zieliński *et al.*, 2005).

C. Predicting preference

Predicting listener preference from specific subjective attributes and, ultimately, from objective measures, is one of the ongoing challenges in research on sound quality. It was not the ambition of this exploratory study to develop a general sound quality model; however, the relation between specific auditory attributes and overall preference established in this paper provides some insight in which sensations might play a role when assessing the overall quality of reproduced sound.

The four recordings were grouped into two musical genres, resulting in two models: one for classical music [Eq. (7)] and one for pop music [Eq. (8)], which accounted for 94% and 84% (respectively) of the variance in the preference

scale values. The similarities between the classical and pop genres in Table V and in Eqs. (7) and (8) are encouraging, as they suggest that similar sensations might have played a similar role in the preference judgments across program materials.

In several studies (e.g., Toole, 1985; Letowski, 1989; Zacharov and Koivuniemi, 2001), the overall sound quality is conceived as consisting of timbral and spatial quality. Recently, Rumsey *et al.* (2005) provided experimental evidence that global quality judgments can be predicted by judgments on timbral and spatial fidelity scales in the context of multichannel audio reproduction. In the present study, the elicited attributes were reduced to two principal components, which might be described as primarily spatial and timbral, respectively: *width, envelopment, distance* and *spaciousness* loaded on one of the components, while *brightness* and *elevation* (attributed to spectral changes) loaded on the other one. These results support the notion that both timbral and spatial auditory attributes are important predictors of overall listener preference.

It is not possible from the present data to determine whether the collinearity of certain auditory attributes results from a common underlying sensation, or whether distinct sensations are involved, but covary, in the context of the selected stimuli. Therefore, the relation between single attributes and overall preference must be interpreted with care. Considering the exploratory nature of this study, and the limited number of stimuli, it will be incumbent upon future research to gain a clearer picture of the functional relations between overall preference and the underlying (more specific) auditory attributes in the context of multichannel sound.

V. CONCLUSIONS

In summary, the following conclusions can be drawn:

(1) By applying probabilistic choice models to binary paired-comparison judgments, it is possible to scale auditory attributes in complex sounds, while revealing inconsistencies related to multidimensionality.

(2) Consistent judgments (with respect to transitivity) were obtained from nonexpert listeners on overall preference as well as on more specific attributes.

(3) The preference judgments were highly reliable both across repetitions and when retesting after six months, indicating that listeners have a clear and stable concept of what they prefer to listen to.

(4) Perceptual similarities were observed between materials; those were more pronounced within musical genres (classic and pop) than across.

(5) For the centered listening position investigated in the present study, stereo downmix was found to be among the most preferred formats, while mono was generally the least preferred.

(6) Preference could be predicted using two principal components derived from the attribute scales: one related to the spatial characteristics of the sounds, the other related to their spectral characteristics.

ACKNOWLEDGMENTS

This research was carried out as part of the "Centerkontrakt on Sound Quality" that establishes participation in and funding of the "Sound Quality Research Unit" (SQRU) at Aalborg University. The participating companies are Bang & Olufsen, Brüel & Kjær, and Delta Acoustics & Vibration. Further financial support comes from the Ministry for Science, Technology, and Development (VTU), and from the Danish Research Council for Technology and Production (FTP). The authors would like to thank Flemming Christensen and Søren Legarth for their help in translating the instructions to the subjects into Danish, Geoff Martin for his qualified advice on the program material selection, and Wolfgang Ellermeier and Søren Bech for their comments on an earlier version of the manuscript. Furthermore, we are grateful to the four anonymous reviewers for their comments.

APPENDIX: ATTRIBUTE DEFINITIONS

These definitions of the attributes were part of the written instructions (in Danish) given to the test subjects prior to the scaling task.

Which of these two sounds is *wider*? Imagine the area occupied by the sound sources (e.g., the instruments). For every pair of sounds, you should indicate for which of the sounds this area is wider.

Which of these two sounds is *more elevated*? Some sounds might appear to be positioned at the same level as your ears. Some others might be lower (closer to the floor) or higher (toward the ceiling). Indicate which of the two sounds you perceive as being higher in space.

Which of these two sounds is *more spacious*? A sound is said to be spacious when you have a good impression of the space in which it is played. Try to imagine this space, it can be a small room for example, or a large hall. Select the sound in which the impression of space is greater.

Which of these two sounds is *more enveloping*? A sound is enveloping when it wraps around you. A very enveloping sound will give you the impression of being immersed in it, while a nonenveloping one will give you the impression of being outside of it.

Which of these two sounds is *further ahead*? Some sounds might appear to be closer to you, whereas others seem farther away. If one of the sounds appears to be behind you, then choose the one that is farther ahead (in front).

Which of these two sounds is *brighter*? A sound is bright when it has emphasized treble, and dark when the emphasis is on the bass (or lacking treble). As an example, a female voice is usually brighter than a male voice.

Which of these two sounds is *clearer*? The clearer the sound, the more details you can perceive in it. Choose the sound that appears clearer to you.

Which of these two sounds is *more natural*? A sound is natural if it gives you a realistic impression, as opposed to sounding artificial.

[1]Zieliński *et al.* (2003) make the distinction between foreground/foreground and foreground/background material in order to denote whether or not there are distinct sources in the surround channels.

[2]This is consistent with Rumsey's (2002) proposal that a "macroattribute" (such as *envelopment*) consists of several "microattributes" (such as *individual-source envelopment* and *ensemble envelopment*).

Baird, J. C. (**1997**). *Sensation and Judgment* (Lawrence Erlbaum, Mahwah, NJ).

Bech, S. (**1998**). "The influence of stereophonic width on the perceived quality of an audiovisual presentation using a multichannel sound system," J. Audio Eng. Soc. **46**, 314–322.

Berg, J. and Rumsey, F. (**2006**). "Identification of quality attributes of spatial audio by repertory grid technique," J. Audio Eng. Soc. **54**, 365–379.

Blauert, J. (**1997**). *Spatial Hearing* (MIT Press, Cambridge, MA).

Bloom, P. J. (**1977**). "Creating source elevation illusions by spectral manipulation," J. Audio Eng. Soc. **25**, 560–565.

Böckenholt, U. (**2001**). "Hierarchical modeling of paired comparison data," Psychol. Methods **6**, 49–66.

Bradley, R. A. and Terry, M. E. (**1952**). "Rank analysis of incomplete block designs: I. The method of paired comparisons," Biometrika **39**, 324–345.

Carroll, J. D., and De Soete, G. (**1991**). "Toward a new paradigm for the study of multiattribute choice behavior," Am. Psychol. **46**, 342–351.

Choisel, S. and Wickelmaier, F. (**2006a**). "Extraction of auditory features and elicitation of attributes for the assessment of multichannel reproduced sound," J. Audio Eng. Soc. **54**, 815–826.

Choisel, S. and Wickelmaier, F. (**2006b**). "Relating auditory attributes of multichannel sound to preference and to physical parameters," *120th Convention of the Audio Engineering Society*, Paris, France, 20–23 May, preprint 6684.

Damaske, P. (**1969**). "Richtungsabhängigkeit von Spektrum und Korrelationsfunktionen der an den Ohren empfangenen Signale (Directional dependence of the spectrum and the correlation function of the signals received at the ears)," Acustica **22**, 191–204.

David, H. A. (**1988**). *The Method of Paired Comparisons* (Oxford University Press, New York).

Ellermeier, W., Mader, M., and Daniel, P. (**2004**). "Scaling auditory unpleasantness according to the BTL model: Ratio-scale representation and psychoacoustical analysis," Acust. Acta Acust. 90, 101–107.

Gabrielsson, A. and Sjögren, H. (**1979**). "Perceived sound quality of sound-reproducing systems," J. Acoust. Soc. Am. **65**, 1019–1033.

Gerzon, M. A. (**1985**). "Ambisonics in multichannel broadcasting and video," J. Audio Eng. Soc. **33**, 859–871.

Guastavino, C. and Katz, B. F. G. (**2004**). "Perceptual evaluation of multidimensional spatial audio reproduction," J. Acoust. Soc. Am. **116**, 1105–1115.

ISO 389-1 (**1998**). "Reference zero for the calibration of audiometric equipment – Part 1: Reference equivalent threshold sound pressure levels for pure tones and supra-aural earphones," ISO, Geneva, Switzerland.

ITU-R BS.1116 (**1997**). "Methods for the subjective assessment of small impairment in audio systems including multichannel sound systems," International Telecommunications Union, Geneva, Switzerland.

ITU-R BS.1534 (**2003**). "Method for the subjective assessment of intermediate quality level of coding systems," International Telecommunications Union, Geneva, Switzerland.

ITU-R BS.775-1 (**1994**). "Multichannel stereophonic sound system with and without accompanying picture," International Telecommunication Union, Geneva, Switzerland.

ITU-T P.800 (**1996**). "Methods for subjective determination of transmission quality," International Telecommunications Union, Geneva, Switzerland.

Jesteadt, W. (**1980**). "An adaptive procedure for subjective judgments," Percept. Psychophys. **28**, 85–88.

Letowski, T. (**1989**). "Sound quality assessment: Concepts and criteria," *87th Convention of the Audio Engineering Society*, New York, USA, 18–21 October, preprint 2825.

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

Luce, R. D. (**1959**). *Individual Choice Behavior: A Theoretical Analysis* (Wiley, New York).

Mattila, V.-V. (**2002**). "Descriptive analysis and ideal point modelling of speech quality in mobile communication," *113th Convention of the Audio Engineering Society*, Los Angeles, USA, 5–8 October, preprint 5704.

May, K. O. (**1954**). "Intransitivity, utility, and the aggregation of preference patterns," Econometrica **22**, 1–13.

McCullagh, P. and Nelder, J. A. (**1989**). *Generalized Linear Models* (Chapman and Hall, London).

Nakayama, T., Miura, T., Kosaka, O., Okamoto, M., and Shiga, T. (**1971**). "Subjective assessment of multichannel reproduction," J. Audio Eng. Soc. **19**, 744–751.

Rossi, P. E., Gilula, Z., and Allenby, G. M. (**2001**). "Overcoming scale usage heterogeneity: A Bayesian hierarchical approach," J. Am. Stat. Assoc. **96**, 20–31.

Rumsey, F. (**1998**). "Subjective assessment of the spatial attributes of reproduced sound," in *Proceedings of the AES 15th International Conference: Audio, Acoustics & Small Spaces*, pp. 122–135.

Rumsey, F. (**2001**). *Spatial Audio* (Focal Press, Oxford).

Rumsey, F. (**2002**). "Spatial quality evaluation for reproduced sound: Terminology, meaning, and a scene-based paradigm," J. Audio Eng. Soc. **50**, 651–666.

Rumsey, F., Zieliński, S. K., Kassier, R., and Bech, S. (**2005**). "On the relative importance of spatial and timbral fidelities in judgments of degraded multichannel audio quality," J. Acoust. Soc. Am. **118**, 968–976.

Stevens, S. S. (**1946**). "On the theory of scales of measurement," Science **103**, 677–680.

Toole, F. E. (**1985**). "Subjective measurements of loudspeaker sound quality and listener performance," J. Audio Eng. Soc. **33**, 2–32.

Tversky, A. (**1969**). "Intransitivity of preferences," Psychol. Rev. **76**, 31–48.

Tversky, A. (**1972**). "Elimination by aspects: A theory of choice," Psychol. Rev. **79**, 281–299.

Tversky, A. and Sattath, S. (**1979**). "Preference trees," Psychol. Rev. **86**, 542–573.

Wickelmaier, F. and Choisel, S. (**2005**). "Selecting participants for listening tests of multichannel reproduced sound," *118th Convention of the Audio Engineering Society*, Barcelona, Spain, 28–31 May, preprint 6483.

Wickelmaier, F. and Schmid, C. (**2004**). "A Matlab function to estimate choice model parameters from paired-comparison data," Behav. Res. Methods Instrum. Comput. **36**, 29–40.

Zacharov, N. and Koivuniemi, K. (**2001**). "Audio descriptive analysis & mapping of spatial sound displays," in *Proceedings of the 2001 International Conference on Auditory Displays*, Espoo, Finland, pp. 95–104.

Zieliński, S. K., Rumsey, F., and Bech, S. (**2003**). "Effects of down-mix algorithms on quality of surround sound," J. Audio Eng. Soc. **51**, 780–798.

Zieliński, S. K., Rumsey, F., Kassier, R., and Bech, S. (**2005**). "Comparison of basic audio quality and timbral and spatial fidelity changes caused by limitations of bandwidth and by down-mix algorithms in 5.1 surround audio systems," J. Audio Eng. Soc. **53**, 174–192.

Zimmer, K., Ellermeier, W., and Schmid, C. (**2004**). "Using probabilistic choice models to investigate auditory unpleasantness," Acust. Acta Acust. 90, 1019–1028.

The binaural temporal window in adults and children

Joseph W. Hall III,[a] Emily Buss, and John H. Grose
Department of Otolaryngology/Head and Neck Surgery, University of North Carolina School of Medicine, Chapel Hill, North Carolina 27599

(Received 21 March 2006; revised 19 September 2006; accepted 30 October 2006)

This study investigated the binaural temporal window in adults and children 5–10.5 years of age. Detection thresholds were estimated for a brief, interaurally out-of-phase ($S\pi$) 500 Hz pure tone signal masked by bandpass, 100–2000 Hz Gaussian noise. In one set of conditions, the masker was consistently either in phase (*No*) or out of phase ($N\pi$). In another set of conditions, the masker changed abruptly in interaural phase ($NoN\pi$ or $N\pi No$), and threshold was estimated at a range of delays with respect to the phase transition. Masked thresholds were also obtained in further conditions where the masker interaural phase was steady and the signal was of long duration. Age effects obtained with dynamic maskers could be accounted for by positing that children have a binaural temporal window with a relatively prolonged leading edge or that the children position the binaural temporal window relatively late with respect to the signal. Modeling of the reduced masking-level difference shown by children for a brief $S\pi$ signal presented in a steady *No* or $N\pi$ masker was more consistent with late placement of a symmetrical binaural temporal window than a binaural temporal window having a relatively prolonged leading edge. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2400673]

PACS number(s): 43.66.Pn, 43.66.Dc, 43.66.Ba, 43.66.Rq [AK] Pages: 401–410

I. INTRODUCTION

The present study investigated the binaural temporal window (Kollmeier and Gilkey, 1990; Culling and Summerfield, 1998; Holube *et al.*, 1998; Bernstein *et al.*, 2001) in adults and school-aged children. The binaural temporal window refers to the temporal epoch during which the auditory system integrates information related to binaural difference cues. Experiments on the binaural temporal window are often performed in the context of the masking-level difference (MLD) paradigm (Hirsh, 1948), where the signal and the masker are presented to the two ears with different interaural phase characteristics. In the paradigm used here Kollmeier and Gilkey, 1990; Holube *et al.*, 1998), the binaural temporal window is measured using a brief, interaurally out-of-phase signal ($S\pi$) as a function of the temporal position of the signal with respect to an abrupt interaural phase transition of the noise masker. This transition is either from in phase to out of phase ($NoN\pi$) or from out of phase to in phase ($N\pi No$). The time constant of the binaural temporal window is estimated, in part, from the steepness with which the detection thresholds of the brief signals change with respect to the temporal occurrence of the signal relative to the interaural phase transition of the masking noises. Estimates of binaural time constants using this method are longer than typical estimates of monaural time constants under analogous stimulus conditions (Kollmeier and Gilkey, 1990; Holube *et al.*, 1998), consistent with the proposal that the binaural system is "sluggish" (Grantham and Wightman, 1978; Grantham and Wightman, 1979).

Although we are not aware of any previous comparison of the binaural temporal windows in adults and school-aged children, several studies have examined monaural temporal processing in children. For example, studies have found that monaural temporal gap detection thresholds are often elevated in children (Irwin *et al.*, 1985; Wightman *et al.*, 1989). Children also have relatively high monaural thresholds for the detection of a pure tone signal in narrow bands of noise that are sinusoidally amplitude modulated (Grose *et al.*, 1993). The temporal modulation transfer function (TMTF) paradigm (Viemeister, 1979) has also been used to study monaural temporal resolution in children (Hall and Grose, 1994). In this method, sensitivity to the presence of amplitude modulation is determined as a function of the modulation rate. The TMTF results indicated that although thresholds for the detection of modulation were higher in children than in adults, they were *uniformly* higher across low and high modulation rates, with the result that derived monaural time constants did not vary across the age range tested (four years to adult). The TMTF results therefore indicated that although children were less sensitive to the presence of modulation than adults, there was no difference in monaural temporal resolution, per se.

One motivation for the present work was to provide basic, new information on the development of temporal resolution for binaural hearing. A more specific motivation concerns previous results that have been obtained on MLDs in children. Whereas the MLD for a pure-tone signal in a wideband masker appears to be adult-like by age 5–6 years (Hall and Grose, 1990), studies have indicated that 5–10-year-old children have reduced MLDs when the masker is a narrowband noise (Grose *et al.*, 1995; Grose *et al.*, 1997). Narrowband noise maskers possess prominent envelope fluctuations (e.g., Bos and de Boer, 1966), and recent studies have indi-

[a]Author to whom correspondence should be addressed. Electronic mail: jwh@med.unc.edu

cated that the $S\pi$ thresholds for tones presented in narrowband noise are determined largely by information coincident with masker envelope minima (Grose and Hall, 1998; Hall *et al.*, 1998; Buss *et al.*, 2003), where the binaural difference cues are the largest (Buss *et al.*, 2003). Although the *monaural* signal-to-noise ratio is also favorable in the envelope minima of narrowband noise maskers, listeners are not able to exploit these epochs in monaural detection, but instead apply equal weight across envelope maxima and minima (Buss *et al.*, 1996). Listeners are able to take advantage of the favorable signal-to-noise ratios available in the masker envelope minima of monaural narrowband noise stimuli only when the narrowband noise is multiplied by a low-frequency modulator (e.g., Carlyon *et al.*, 1989; Grose *et al.*, 1993) or when comodulated flanking noise bands are present (e.g., Hall *et al.*, 1984). Thus, for a single, narrowband Gaussian noise masker, listeners are able to take advantage of the good signal-to-noise ratios associated with masker envelope minima for binaural but not for monaural detection.

A developmental study by Hall *et al.* (2004) suggested that the relatively small MLDs shown by children in narrowband masking noise may be due to a reduced ability to take advantage of the binaural information occurring in the relatively brief masker envelope minima. The ability to take advantage of the temporal epochs containing the most favorable binaural detection cues (masker envelope minima) may depend upon the binaural temporal window, both in terms of its duration (a shorter duration associated with better acuity), and in terms of the temporal relation between the center of the window and the optimal time to listen for the signal (optimal temporal alignment associated with better acuity). The present study examined the binaural temporal window in adults and in children aged 5–10.5 years.

II. METHODS

A. Listeners

All listeners had pure-tone detection thresholds of 20 dB Hearing Level (HL) or better at octave frequencies from 250 to 8000 Hz (ANSI, 1996). None had a history of chronic ear disease, and none had a known history of otitis media within a three-year period preceding testing. Thirteen children were recruited into the study. Two of these children were dropped from the study due to high test-retest variability (more than 15 dB variation among threshold estimates). The remaining 11 listeners (seven females and four males) ranged in age from 5 to 10.5 years, with a mean age of 8.0 years (standard deviation 1.5 years). There were 12 adult listeners (nine females and three males), ranging in age from 18 to 43 years, with a mean age of 26.6 years (standard deviation 9.2 years). All listeners were paid for participation and provided data in four sessions lasting no more than 1 h each.

B. Stimuli

The signal was a 500 Hz pure tone, ramped on and off with 5 ms $\cos^2$ ramps and 10 ms of steady state. All signals were presented in $S\pi$ phase. Maskers were Gaussian noise samples, bandpass filtered 100–2000 Hz, with onset and offset ramps imposed by Finite Impulse Response (FIR) filters.

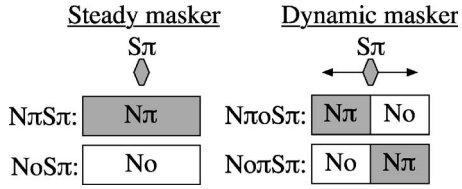


FIG. 1. Schematic (not to scale) of the stimuli used in the *steady masker* and *dynamic masker* conditions of the main experiment. In the *steady masker* conditions, the signal was presented in the temporal center of the masker. In the *dynamic masker* conditions, the signal was presented at a range of delays relative to the abrupt interaural phase transition occurring at the temporal center of the masker.

The total duration of each masker sample was 950 ms. The masker was presented at a level of 40 dB/Hz sound pressure level (SPL). Stimulus conditions are illustrated schematically in Fig. 1. In the *steady masker* conditions, the masker was either *No* or $N\pi$ throughout its presentation. In the *dynamic masker* conditions, the phase changed abruptly in the temporal center of the masker.

Listeners completed the *steady masker* conditions first. In these conditions, the signal was coincident with the temporal center of the masker. In the *dynamic masker* conditions, all signal delays were defined in terms of the relation between the masker transition point and the temporal center of the signal. Listeners were randomly assigned to two groups: one group completed the $NoN\pi$ conditions before the $N\pi No$ conditions, and the other group completed these conditions in the reverse order.

Because testing time was limited with the children, procedures for estimating thresholds in the *dynamic masker* conditions were designed to maximize efficiency. To this end, the function associated with each set of conditions was first broadly characterized by estimating thresholds at four signal delays: one at each estimated asymptote (upper and lower), one near the masker phase transition, and one well into the region of improved thresholds. Based on the pattern of results obtained for these four delays, two additional delays were identified and tested, with the goal of increasing the number of thresholds obtained in the most steeply sloping portion of the function.

For the $NoN\pi$ condition, the four initial delays were $-320, -120, -20$, and 280 ms. The two additional delays were based on comparison of (1) the midpoint between the thresholds for the −320 and −20 ms delay conditions and (2) the threshold for the −120 ms delay condition. If the threshold for the −120 ms delay condition was lower than the midpoint (indicating that appreciable changes in threshold were confined to delays near the abrupt transition), then the two additional delays were −70 and 30 ms (relatively short delays, placing the signal near the abrupt transition). This was the case for all of the adults and for 10 of the 11 children. For the remaining child (age=10.5 years), the threshold for the −120 ms delay condition was higher than the midpoint between the −320 and −20 ms delay conditions (indicating an appreciable change in threshold over a longer time period), and the two additional delays were −220 and 180 ms (relatively long temporal separation relative to the abrupt transition).

The same general procedure was followed for the $N\pi No$ conditions. The four initial delays were −320, −20, 80, and 280 ms. If the threshold for the 80 ms delay condition was lower than the midpoint between the 280 and −20 ms delay conditions, then the two additional thresholds were obtained for the relatively short delays of −70 and 30 ms. As with the $NoN\pi$ conditions, this occurred for all of the adults and for 10 of the 11 children. For the remaining child (age =6.5 years), the threshold for the 80 ms delay condition was higher than the midpoint between the 280 and −20 ms delay conditions, and the two additional delays were −220 and 180 ms.

C. Threshold estimation procedure

The task used a three-alternative forced-choice procedure with the signal level adjusted in a three-down one-up track estimating the 79% correct point on the psychometric function. The signal level was adjusted in steps of 4 dB for the first two reversals and then in steps of 2 dB for the remaining six reversals. Thresholds were computed as the average level at the last six reversals. All thresholds were obtained in blocks, by condition. Two estimates were obtained in each condition, with a third estimate obtained only in cases where the first two varied by 3 dB or more. Listening intervals were marked visually using animation on a video monitor. Over the course of a threshold run, a cartoon picture was unmasked, in the style of a jigsaw puzzle, with one piece revealed following each correct response. This cartoon was completely unmasked and performed a 2 s animation at the end of the threshold run. All listeners used this interface.

D. Modeling the binaural temporal window

Thresholds were fitted with a double-sided exponential window, using the *fmins* function in MATLAB (MathWorks). This window acts upon and integrates interaural correlation associated with the masking stimulus. The double-sided exponential window was described by Kollmeier and Gilkey (1990) and takes the form of

$$\omega(t) = \frac{\exp(t/\tau_1)}{(\tau_1+\tau_2)} \quad \text{for } t \leqslant 0 = \frac{\exp(-t/\tau_2)}{(\tau_1+\tau_2)} \quad \text{for } t > 0, \tag{1}$$

where t is time and $\tau 1$ and $\tau 2$ are the time constants associated with the lagging and leading edges of the window, respectively. Following Kollmeier and Gilkey (1990), masking at the output of this window was estimated based on the equalization and cancellation model (Durlach, 1963), with

$$L(t) = L_M - 10\log((K+1)/[K-r(t)]), \tag{2}$$

where L_M is the monaural threshold, K is internal noise, and $r(t)$ is the time-varying interaural correlation of the windowed input. The parameter K was estimated based on the MLD in *steady masker* condition, while the parameter L_M was allowed to vary freely. This reflects an assumption that the maximum MLD in the *dynamic masker* conditions is equal to the MLD obtained in the *steady masker* conditions, but that all thresholds may suffer from the presence of the masker phase transition, those based on binaural cues and those based on monaural cues alike. As in previous studies using this fitting technique (Kollmeier and Gilkey, 1990; Holube *et al.*, 1998), three parameters were allowed to vary: time constants for the lagging and leading edges of the window ($\tau 1$ and $\tau 2$) and an estimate of overall processing efficiency (L_M). As in Kollmeier and Gilkey (1990), it was assumed that the binaural temporal window was centered on the onset of the $S\pi$ signal in the $NoN\pi$ masker and on the offset of the $S\pi$ signal in the $N\pi No$ masker. This assumption was based on the idea that "off-time" listening would allow improvement of the effective signal-to-noise ratio at the output of the binaural temporal window. The data for the $NoN\pi$ conditions and the $N\pi No$ conditions were fitted simultaneously.

TABLE I. Mean $N\pi S\pi$ and $NoS\pi$ thresholds (dB SPL) and derived MLDs for the *steady masker* conditions, with results shown separately for the two age groups. Standard deviations are in parentheses.

	$N\pi S\pi$	$NoS\pi$	MLD
Adult	67.5	51.0	16.5
	(0.7)	(1.6)	(1.7)
Child	70.4	57.6	12.8
	(1.4)	(3.1)	(3.5)

III. RESULTS AND DISCUSSION

A. Developmental differences for the steady masker

Although the main aim of the present study was to investigate binaural temporal resolution for maskers with dynamically varying interaural phase, MLDs were also obtained for *No* and $N\pi$ *steady maskers* in order to provide a basis for estimating K (see Methods Sec. II). Table I summarizes findings for the *steady masker* conditions. As indicated in Table I, the children had higher thresholds than the adults in the $N\pi S\pi$ and the $NoS\pi$ conditions, and also had smaller MLDs. A repeated measures analysis of variance showed that the $N\pi S\pi$ thresholds were higher than the $NoS\pi$ thresholds ($F_{1,21}=650; p<0.001$), that the children had higher thresholds than the adults ($F_{1,21}=65.5; p<0.001$), and also that there was a significant interaction between threshold and group ($F_{1,21}=10.6; p<0.005$). This interaction reflects the fact that the developmental effect (higher thresholds for the children) was greater for the $NoS\pi$ threshold than for the $N\pi S\pi$ threshold, resulting in a smaller MLD for the children. The smaller MLD of the children in the relatively wideband masking noise used here might appear to be at odds with results reported in the studies of Hall and Grose (1990) and Grose *et al.* (1997), where it was found that children 5–6 years of age and older had adult-like MLDs for masking noise bandwidths wider than approximately 300 Hz. However, both Hall and Grose (1990) and Grose *et al.* (1997) used a long duration (400 ms) signal, so it is possible that the discrepancy between the present and past findings is related to signal duration. This possibility will be considered further below.

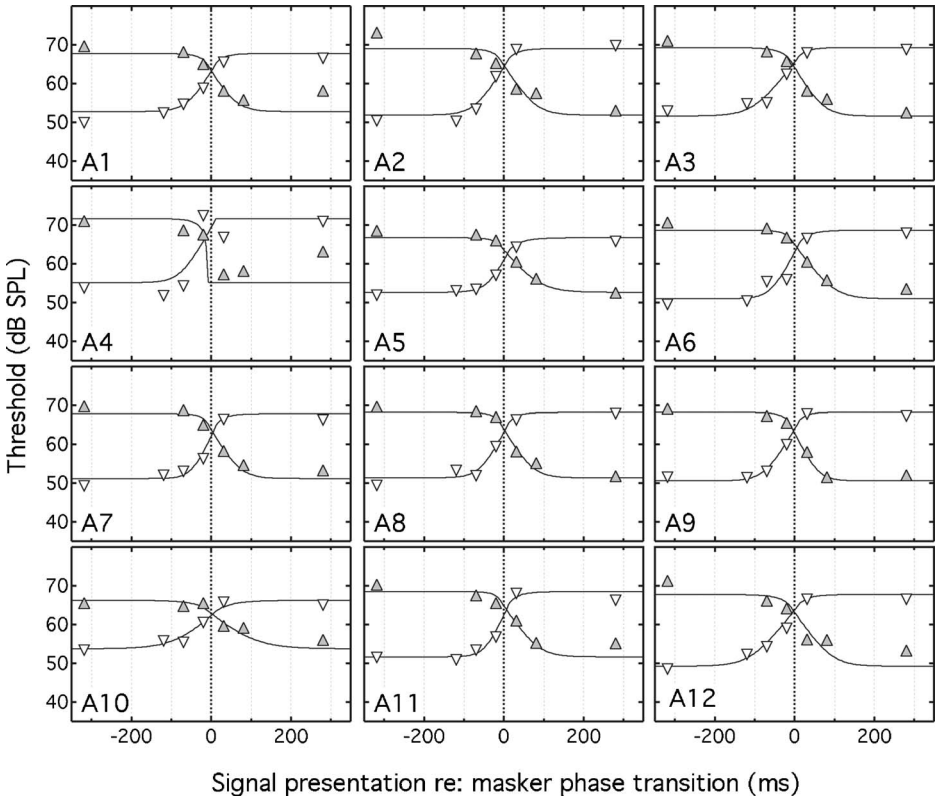


FIG. 2. Mean $S\pi$ thresholds for each adult (A1–A12) listener are shown by the unfilled, downward pointing triangles for the $NoN\pi$ conditions, and by the filled, upward pointing triangles for the $N\pi No$ conditions. Thresholds are plotted as a function of the delay between the signal and the interaural phase transition of the masker. The solid lines show modeled fits based on a double-sided, exponential temporal window. The dotted vertical line represents the timing of the masker phase transition.

B. Developmental differences for the dynamic maskers

1. Time constants and off-time listening

Individual results for the adults and children are shown in Figs. 2 and 3, respectively, with the $S\pi$ signal threshold plotted as a function of delay of the temporal center of the signal with respect to the masker transition point. The unfilled, downward pointing triangles represent data for the $NoN\pi$ conditions and the filled, upward pointing triangles represent data for the $N\pi No$ conditions. The columns of Table II under the "onset/offset fit" heading show the individual and median values for $\tau 1$ and $\tau 2$ and the percent

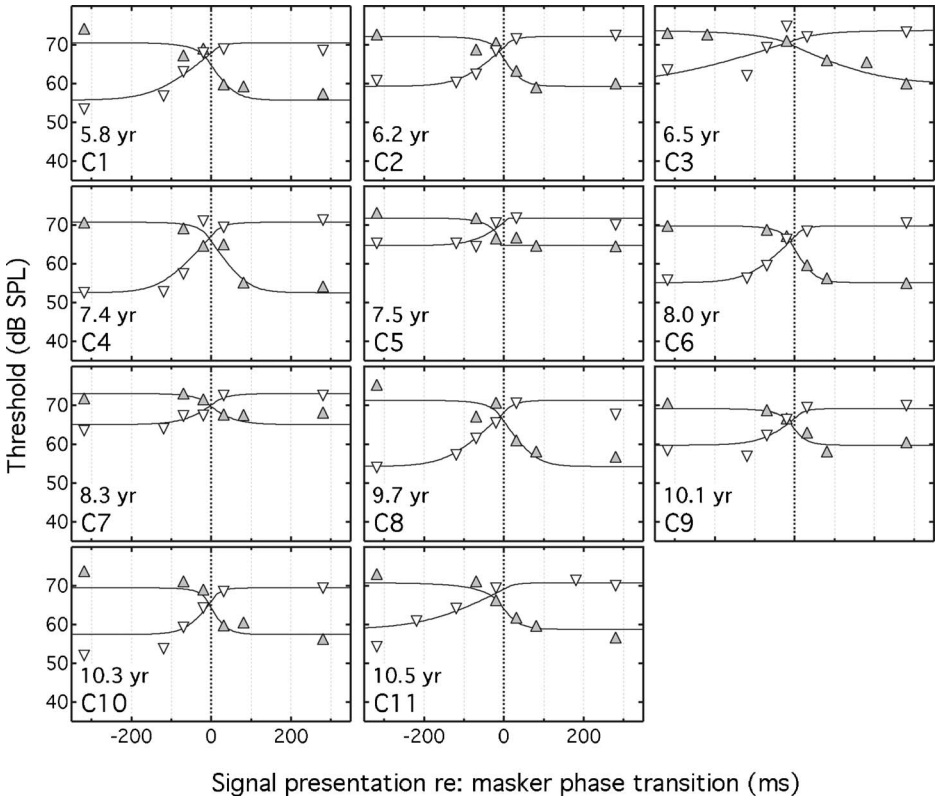


FIG. 3. Mean $S\pi$ thresholds for each child (C1–C11) listener are shown, following the plotting conventions of Fig. 2. Also shown is the child's age in years.

TABLE II. Values of $\tau 1$, $\tau 2$(ms) and percent of variance accounted for the onset/offset approach are shown on the left part of the table, and values of τ (ms), SHIFT (ms) and percent of variance accounted for in the fitted shift approach are shown on the right side of the table. Adults (A1–A12) are shown at the top of the table and children (C1–C11) are shown at the bottom of the table. AX refers to adult median values for τ and mean value for percent of variance accounted for. CX refers to like values for children. Numbers in parentheses are the lower and upper quartiles around the median.

	Onset/offset fit			Fitted shift fit					
					NπNo			NπNo	
Listener	$\tau 1$	$\tau 2$	%Var	τ	SHIFT	% var	τ	SHIFT	%var
A1	24	24	77	32	60	98	28	−4	99
A2	26	25	93	40	60	99	16	0	99
A3	24	31	90	34	42	98	36	−28	99
A4	−1	24	52	140	−78	68	6	52	81
A5	33	20	98	32	12	99	40	−18	100
A6	29	21	87	34	18	99	28	−42	99
A7	23	18	95	32	44	100	24	−24	99
A8	24	21	94	26	18	99	32	−30	97
A9	20	24	97	22	14	99	20	−12	99
A10	47	42	91	52	16	99	36	−18	98
A11	27	18	88	40	58	99	20	−10	98
A12	29	32	86	26	14	99	28	−38	100
AX	25	24	87	33	18	96	28	−18	97
	(24/29)	(21/26)		(30/40)	(14/47)		(20/33)	(−28/−8)	
C1	24	47	91	16	26	91	26	50	100
C2	22	41	75	38	8	98	30	0	99
C3	83	121	43	70	−4	100	38	16	99
C4	28	36	58	36	54	96	6	56	99
C5	0	42	80	32	60	100	28	24	95
C6	21	39	94	34	40	100	34	8	100
C7	35	40	74	140	74	74	40	−2	97
C8	27	42	77	20	26	97	46	30	99
C9	20	40	78	38	34	97	12	46	93
C10	19	27	69	116	78	90	22	56	99
C11	23	75	81	46	22	98	64	48	98
CX	23	41	75	38	34	95	30	30	98
	(20/28)	(39/45)		(35/58)	(24/57)		(24/39)	(12/49)	

variance accounted for by the data fits (the "fitted shift fit" results in the table will be considered in the following section). As can be seen in Table II, the percent of variance accounted for by the data fits varied considerably among listeners, ranging from 52% to 98% in the adults and from 43% to 94% in the children. In the statistics below that are related to parameter estimates, the median rather than the mean is used to characterize central tendency and nonparametric tests are used to test for parameter differences. This approach was adopted because evaluations of kurtosis and skewness associated with the parameters of interest were often not consistent with a normal distribution, particularly for the children. An alpha criterion of $p<0.05$ was adopted for statistical significance.

The estimated time constants associated with the lagging and leading edges of the binaural temporal window ($\tau 1$ and $\tau 2$) were similar for the adults, with a median value of 25 ms for $\tau 1$ and 24 ms for $\tau 2$ (see Table II). These results are consistent with those of Kollmeier and Gilkey (1990), where $\tau 1$ and $\tau 2$ were relatively symmetrical, both ranging between approximately 17 and 40 ms. A Wilcoxon signed ranks test indicated no significant difference between $\tau 1$ and $\tau 2$ for the adults tested here ($Z=0.76$; $p=0.45$). For children, the median values of $\tau 1$ and $\tau 2$ were 23 and 41 ms, respectively. The signed ranks test indicated that the difference between the $\tau 1$and $\tau 2$ values of the children was significant ($Z=2.9; p=0.003$). A Mann-Whitney U test indicated that $\tau 1$ did not differ between adults and children ($U=53.0; p=0.45$), but that $\tau 2$ was longer for children than for adults ($U=9.0; p<0.001$).

2. Evaluation of the possibility that children place the binaural temporal window late

Although the above analysis indicates that $\tau 2$ differed between adults and children, it is not clear that the duration of the time constant is the best way to characterize the difference in data patterns between the adults and children. As noted above, Kollmeier and Gilkey (1990) assumed that their adult listeners were able to improve the effective signal-to-noise ratio by listening off time and attempted to take this into account by centering the binaural temporal window at the beginning of the $S\pi$ signal in the $NoN\pi$ conditions and the end of the signal in the $N\pi No$ conditions. It is possible that children might not be as adept as adults in approaching optimal temporal placement of the binaural temporal win-

dow. One simple possibility that is qualitatively consistent with the data pattern observed here is that, whereas adults tend to optimize the position of the binaural temporal window (as suggested by Kollmeier and Gilkey), children have a general tendency to listen late. With respect to optimal off-time listening, the penalty associated with listening late might be relatively minor for an $S\pi$ signal in an $N\pi No$ masker because listening late would tend to improve the signal-to-noise ratio for a signal occurring near the masker phase transition. This interpretation is consistent with the finding that τ1 did not differ significantly for adults and children in the present study. A more significant penalty to listening late would be expected for an $S\pi$ signal in an $NoN\pi$ masker because listening late in this case would tend to worsen the effective signal-to-noise ratio, particularly for a signal near the masker phase transition. This would create a difference between adults and children, as adults are assumed to listen "early" in order to improve the effective signal-to-noise ratio in this condition.

In order to evaluate the idea that adults and children have similar binaural time constants but that children place the binaural temporal window relatively late in time, two additional data fitting approaches were pursued. The first employed a fitting procedure very similar to that of Kollmeier and Gilkey (1990), but with separate fits to the $NoN\pi$ and $N\pi No$ data and different free parameters. The parameter K was estimated based on the MLD in *steady masker* condition, while the parameters L_M, the time constant (τ) and placement of the temporal window (SHIFT) were allowed to vary freely. A single value of τ and single value of SHIFT were fitted to the $N\pi No$ data set, and a single value of τ and a single value of SHIFT were fitted to the $NoN\pi$ data set. With the single value of τ, the leading and lagging edges of the temporal window were forced to be the same. The SHIFT parameter allowed the temporal center of the binaural temporal window to vary, instead of being fixed, at values assumed to be near optimal. SHIFT represents the time difference between the center of the window and the center of the signal, with positive SHIFTs associated with a temporal window position after the center of the signal, and negative shifts associated with a temporal window position before the center of the signal. This "fitted shift" approach will be contrasted with the onset/offset shift approach used above, where the temporal position of the binaural temporal window was assumed to be either at the onset or at the offset of the signal (depending on the dynamic masker phase condition). If adults and children have similar binaural temporal window time constants, but differ in terms of the temporal placement of the binaural temporal window (adults place the window to enhance the effective signal-to-noise ratio but children place the window relatively late), the following outcomes were expected from the fitted shift approach: (1) Neither the $NoN\pi$ value of τ nor the $NoN\pi$ value of τ would differ between adults and children; (2) Adults would show SHIFT values that were negative for the $NoN\pi$ condition and positive for the $N\pi No$ condition; Children would show SHIFT values that were positive for both $NoN\pi$ and $N\pi No$. The results of the data fits were consistent with these expectations (see fitted shift fit columns in Table II). The variance accounted for in the fitted shift approach was considerably greater than found in the onset/offset fit approach (see Table II). The primary reason for this is that, whereas the $NoN\pi$ and $N\pi No$ data functions were fitted simultaneously in the onset/offset fit approach, each function was fitted separately in the fitted shift approach. The variance accounted for in the fitted shift approach was better than 90% for all listeners except for one adult (A4) and one child (C7) (see Table II) for whom the variance accounted for was considerably less. These two listeners were excluded from the statistical tests that follow (although the pattern of statistical significance was the same regardless of this exclusion). A Mann-Whitney test indicated that the children and adults did not differ for τ based on the $N\pi No$ data ($U=42.5; p=0.39$) or τ based on the $NoN\pi$ data ($U=50.5; p=0.76$); furthermore, Wilcoxon signed ranks tests indicated that the two values of τ did not differ from each other significantly either for the adults ($Z=1.4; p=0.15$) or for the children ($Z=1.2; p=0.24$). For the adults, estimates of SHIFT differed significantly for the two masker conditions ($Z=2.9; p=0.003$), with a positive SHIFT derived from the $NoN\pi$ conditions and a negative SHIFT derived from the $N\pi No$ conditions (see Table II). The SHIFT values for the children did not differ significantly ($Z=0.05; p=0.96$), with a positive median value for both the $NoN\pi$ conditions and the $N\pi No$ conditions (see Table II). The fact that the fitted SHIFT differed in sign for the adults is consistent with Kollmeier and Gilkey's assumption that adults can shift the binaural temporal window in a direction that achieves a better effective signal-to-noise ratio. The fact that the fitted values of SHIFT were both positive for the children is consistent with an interpretation that children do not shift the binaural temporal window in a way that improves the effective signal-to-noise ratio, but instead listen late in all conditions.

There are at least two caveats that should be considered with respect to the fitted shift approach. One concerns the assumption that the listener can position the temporal window to enhance the effective signal-to-noise ratio. Although this assumption appears to be reasonable, at least for adult listeners, it is likely that listeners would not use a *constant* SHIFT across all signal delays in either the $NoN\pi$ or the $N\pi No$ conditions. It would seem more likely that the SHIFT would be greater relatively near the masker phase transition and, perhaps, negligible for longer delays, where the signal is well removed from the masker phase transition. This is a consideration not only for the case where SHIFT is a free parameter, but also in the onset/offset shift approach used here and by Kollmeier and Gilkey (1990). The second caveat concerns a penalty to be paid in terms of reduced effective signal level when the center of the binaural temporal window is shifted away from the temporal center of the signal. In the Kollmeier and Gilkey fitting procedure, K is derived from the *steady masker* condition where it can be assumed (at least for adults) that the binaural temporal window is centered on the signal and, therefore, the signal is minimally attenuated by the window. For the dynamic masker phase conditions, where a shift in the position of the binaural temporal window may improve the effective signal-to-noise ratio, it is important to consider the effective attenuation of signal level that

may result from such a shift. Consideration of this issue may be particularly important in children, where an assumption of optimal off-time listening would appear to be very questionable. For example, in the $NoN\pi$ conditions, the median SHIFT for the children using the fitted shift fitting procedure was 34 ms, and this SHIFT was in a temporal direction *opposite* to that of the SHIFT thought to be optimal in these conditions. These observations raise the question of whether the modeling of a free SHIFT parameter should take into account a reduction in the effective signal level due to the difference between the temporal position of the signal and the temporal position of the binaural temporal window. This question is difficult to answer with certainty, but it could be argued that the value of K used in modeling the $NoN\pi$ and $N\pi No$ data of the children already takes into account the difference between the temporal position of the signal and the temporal position of the binaural temporal window. If the premise is correct—that children place the binaural temporal window late in all conditions—then the effective reduction in signal amplitude resulting from late placement of the binaural temporal window is reflected in the *steady masker* condition from which the value of K was derived. Indeed, placing the binaural temporal window late with respect to the signal temporal position in the *steady masker* condition could be one explanation for the finding that children showed smaller MLDs than adults in this condition. Recall that previous studies had shown comparable MLDs for adults and children for long-duration signals presented in relatively broadband noise, and, in these cases, placement of the binaural temporal window would have less of an impact on thresholds because the long duration signal would introduce multiple opportunities distributed in time in which to detect the signal. This possibility will be considered in more detail, below. In summary, the validity of the parameter estimates from fitted shift analysis is questionable due to additional effects of signal attenuation due to misalignment of the signal and the binaural temporal window. However, the results of the fitted shift analysis are broadly consistent with the assumption of Kollmeier and Gilkey (1990) that adults place the temporal window in a temporal direction consistent with optimal placement, and with an interpretation that children place the temporal window late.

An additional data fitting approach for the *steady masker* condition further explored the question of whether the results of the children in the dynamic masker conditions might be better accounted for in terms of a relatively long $\tau 2$ or late placement of a symmetrical (adult-like) binaural temporal window. This approach examined whether the reduced MLD of the children found in the *steady masker* condition was more consistent with the assumption of a prolonged $\tau 2$ or late placement of the binaural temporal window. This approach assumed that the MLDs of adults and children are similar in conditions where the parameters of the binaural temporal window have minimal influence on the MLD (in agreement with finding that adults and children have similar MLDs for long-duration signals in wideband noise), and that the reduced MLDs of children for brief signals result from developmental differences in the binaural temporal window. Note that the assumption of similar MLDs in adults and children does not require similar binaural detection thresholds, but rather similar differences between the monaural and binaural thresholds. The crux of the data fitting approach was to pass an $S\pi$ signal and *No* noise like those used in the *steady masker* condition, above, through a binaural temporal window representative of those fitted to the data of the adults using the onset/offset shift approach. The parameters of $\tau 1$ and $\tau 2$ were set to 25 and 24 ms, respectively, the median values fitted for the adult listeners. Effects were then examined of either prolonging the $\tau 2$ of the temporal window or increasing the relative delay between the center of the binaural temporal window and the temporal center of the signal, in order to gain insight into the question of whether the reduced MLD of the children in the *steady masker* could be better accounted for by a prolonged $\tau 2$ or late temporal placement of an adult-like temporal window. This approach assumed that, in the *steady masker* case, the optimal binaural temporal window placement is at the temporal center of the signal, as there are no transient masker features to corrupt base line (*No*) correlation.The first step was to model the correlation at the output of the binaural temporal window with a signal-plus-noise stimulus that corresponded to the median $NoS\pi$ threshold achieved by the adults in the *steady masker* condition (40 dB/Hz SPL Gaussian masker filtered from 100 to 2000 Hz; 500 Hz $S\pi$ signal with 5 ms ramps and 10 ms steady state, set to the level of 51 dB SPL). A total of 10 000 masker samples, each of 1000 ms duration were generated. The stimuli were FIR filtered to a bandwidth of 78 Hz, the approximate width of the monaural auditory filter at 500 Hz (Moore and Glasberg, 1983). Each masker sample was copied and the signal was added at the center of one copy and subtracted from the other, resulting in a pair of arrays characterizing the *steady masker*, $NoS\pi$ condition. The interaural correlation at the output of the binaural temporal window centered on the signal was then computed as the integral of the instantaneous correlation weighted by the shape of the binaural temporal window, using integration limits of −200 to +200 ms. The mean correlation so determined was 0.981. Repeating this procedure using values of $\tau 1$ and $\tau 2$ estimated from child data (23 and 41 ms, respectively), the signal level required to produce a correlation of 0.981 was elevated by approximately 1.5 dB. Whereas this analysis suggests that the relatively longer $\tau 2$ estimated for the children would produce a reduction in the MLD of 1.5 dB, the observed reduction of the MLD shown by the children in the *steady masker* condition was approximately 3.5 dB. Thus the agreement between the modeled reduction of the *steady masker* MLD and the actual reduction was not particularly close. Further modeling indicated that the value of $\tau 2$ required to result in a 3.5 dB reduction of the MLD was approximately 75–80 ms.

A similar analysis was then performed to determine what delay between the center of the signal and the center of the binaural temporal window would be needed in order to account for a 3.5 dB reduction in the MLD in the *steady masker* condition. To examine this question, the adult estimates of $\tau 1$ and $\tau 2$ were maintained (25 and 24 ms, respectively) and the center of the binaural temporal window was progressively advanced with respect to the center of the sig-

nal. The analysis indicated that the advance associated with a $S\pi$ threshold increase of 3.5 dB was approximately 25–30 ms. This value agrees well with the values suggested for the children in the fitted shift approach used above (see Table II). Thus, whereas the children's prolonged $\tau 2$ of 41 ms estimated in the onset/offset shift approach would not appear to account for the reduced MLD of the children in the *steady masker* condition, the effect can be accounted for by late (25–30 ms) placement of the binaural temporal window. This interpretation will be considered further in Sec. III, below.

3. Consideration of possible confusion effects

Kollmeier and Gilkey (1990) noted that there appeared to be a general deleterious effect in the $N\pi NoS\pi$ condition such that even when the $S\pi$ signal was presented well before the masker phase transition, the obtained threshold was approximately 2 dB higher (poorer) than that obtained in a steady $N\pi$ masker. Although the source of this effect was not clear, Kollmeier and Gilkey speculated that the masker interaural phase transition might result in a sensation that could be confused with the signal. It was of interest to determine whether the magnitude of this effect was different between adults and children in the present study. We therefore examined the difference between the $S\pi$ threshold in the *steady masker* to the $S\pi$ threshold in the $N\pi No$ masker at the −320 ms delay. For our adult listeners, the threshold for the −320 ms delay condition was higher than that in the *steady masker* condition by an average of 2.2 dB ($sd=1.6$ dB), and for the children this effect averaged 1.6 dB ($sd=1.8$ dB). The difference between the adults and children was not significant ($t_{21}=0.89; p=0.38$).We also examined the difference between the $S\pi$ threshold in the *steady masker* and the $S\pi$ threshold in the $N\pi No$ masker at the 280 ms delay. For our adult listeners, the threshold in the 280 ms delay condition was higher than that in the *steady masker* condition by an average of 2.1 dB ($sd=1.7$ dB), and for the children this effect averaged 0.9 dB ($sd=1.6$ dB). The difference between the adults and children was again not significant ($t_{21}=1.8; p=0.09$). These results suggest that it is reasonable to conclude that any general, deleterious effect associated with the masker transition was no worse for children than for adults.

IV. SUPPLEMENTARY CONDITIONS EXAMINING THE MLD FOR BRIEF AND LONG-DURATION SIGNALS

As discussed above, one way to account for the reduced MLD in wideband noise obtained by the children in the present study is by late placement of the binaural temporal window relative to the signal. If signal energy is present for only a brief time, even a relatively small error in the placement of the binaural temporal window could have a material negative consequence for binaural signal detection. However, if signal energy is present over several hundred milliseconds, as was the case in the studies of Hall and Grose (1990) and Grose *et al.* (1997), a relatively small (e.g., 25–30 ms) delay in the placement of the window would be of little consequence. This interpretation is consistent with the fact that the previous studies using long-duration signals found no developmental difference for the MLD, but the present study using a short-duration signal found a smaller MLD in children. This interpretation is undermined to some extent by the fact that it is based upon results that were obtained in different studies using different sets of listeners. Furthermore, the previous studies compared *NoSo* and $NoS\pi$ thresholds to compute the MLD, whereas the present study compared $N\pi S\pi$ and $NoS\pi$ thresholds to compute the MLD. Given the theoretical importance of the finding that the MLDs of children in relatively wideband noise are adult-like for long duration signals but are reduced for brief signals, we examined supplementary conditions to determine the effect of signal duration within a single set of listeners.

TABLE III. Mean $N\pi S\pi$ and $NoS\pi$ thresholds (dB SPL) and derived MLDs for the two age groups. Data are shown for the long- and brief-duration signals of the supplementary conditions. Standard deviations are in parentheses.

	Long signal			Brief signal		
	$N\pi S\pi$	$NoS\pi$	MLD	$N\pi S\pi$	$NoS\pi$	MLD
Adult	56.1	43.0	13.1	68.1	53.3	14.8
	(1.0)	(1.8)	(1.6)	(1.2)	(1.6)	(2.1)
Child	58.7	46.7	12.0	70.4	59.6	10.8
	(2.2)	(2.8)	(2.5)	(2.3)	(4.0)	(3.2)

A. Listeners

All listeners had pure-tone detection thresholds of 20 dB HL or better at octave frequencies from 250 to 8000 Hz (ANSI, 1996). None had a history of chronic ear disease, and none had a known history of otitis media within a 3-year period preceding testing. Nine children were recruited (four females and five males), ranging in age from 5 to 10.5 years, with a mean age of 7.9 years (standard deviation 1.9 years). There were ten adult listeners (six females and four males), ranging in age from 21 to 44 years, with a mean of 31.5 years (standard deviation 8.5 years). All listeners were paid for participation and provided data in one session lasting approximately 1 h each. None of the children and only one of the adults had participated in the main experiment.

B. Stimuli and threshold estimation

The masker was a continuous Gaussian noise, bandpass filtered from 100 to 2000 Hz, and presented at a level of 40 dB/Hz SPL. The signal was a 500 Hz pure tone, ramped on and off with 5 ms $\cos^2$ ramps. The steady-state duration of the signal was either 10 or 400 ms. The signal was $S\pi$ and the masker was either *No* or $N\pi$. The threshold estimation procedure and visual interface providing interval and feedback information were the same as for the main experiment (see above).

C. Results and discussion

Table III summarizes findings for both the brief and long-duration signal conditions. The long-duration signal results will be considered first. A repeated measures analysis of

variance showed that the $N\pi S\pi$ thresholds were higher than the $NoS\pi$ thresholds ($F_{1,17}=688.7; p<0.001$), that the children had higher thresholds than the adults ($F_{1,17}=15.3; p=0.001$), and there was no significant interaction between threshold and group ($F_{1,17}=1.3; p=0.27$). The lack of a significant interaction indicates that the MLD magnitude did not differ significantly between adults and children. This uniformity in the MLD magnitude across the age range tested here is consistent with the previous results obtained by Hall and Grose (1990) and Grose *et al.* (1997) for *NoSo* and $NoS\pi$ stimuli and masking noise bandwidths of 300 Hz or wider. A repeated measures analysis of variance on the data for the brief signal showed that the $N\pi S\pi$ thresholds were higher than the $NoS\pi$ thresholds ($F_{1,17}=327; p<0.001$), that the children had higher thresholds than the adults ($F_{1,17}=14.6; p=0.001$), and that there was a significant interaction between threshold and group ($F_{1,17}=7.6; p=0.01$). This interaction reflects the fact that the MLD for a brief signal was smaller in children than in adults. Thus, in contrast to the results for the long-duration signal, the adults showed a larger MLD than the children for the brief signal (see Table III), a result that replicates the significant developmental difference obtained in the main experiment.

Overall, the results of the supplementary conditions confirm that whereas children and adults have similar MLDs for a relatively wideband masker when the signal is of long duration, children have smaller MLDs than adults when the signal duration is brief. One interpretation that is consistent with this finding is that shifts in the binaural temporal window, such as those derived in the fitted shift procedure described above, are not restricted to dynamic masker phase conditions.

V. GENERAL DISCUSSION

The developmental findings of the main experiment are consistent with an interpretation that children have either an asymmetrical binaural temporal window with a relatively long $\tau 2$ or late placement of a symmetrical binaural temporal window. Although the analyses performed here do not rule out the possibility of an asymmetrical binaural temporal window in children, an interpretation based upon late placement of a symmetrical temporal window was favored on the basis of two data fitting/modeling approaches. The first approach indicated that if the temporal position of the binaural temporal window is allowed to vary, the shift is in opposite directions for the $NoN\pi$ and $N\pi No$ conditions (consistent with the optimal off-time listening suggested by Kollmeier and Gilkey (1990)) for adults, but it is in the same (late) direction for children. This approach indicated no significant difference between the $NoN\pi$ or $N\pi No$ time constants for either adults or children. The second approach indicated that the reduced MLDs of the children in the steady masker condition are not well accounted for by the prolonged $\tau 2$ derived in the onset/offset fitting approach, but are well accounted for by a 25–30 ms lag in the placement of the binaural temporal window. Late placement of the binaural temporal window should result in a reduced MLD for a brief signal, where the signal-to-noise ratio is good for only a short time, but should not result in a reduced MLD for a long-duration signal, where the signal-to-noise ratio is good over an extended time. The data of the supplementary conditions reported above were consistent with this interpretation, with children showing adult-like MLDs for a long-duration signal but MLDs that were smaller than those of adults for a brief signal.

The present interpretation that children are inefficient in the placement of the binaural temporal window is also relevant to previous findings indicating that children show reduced MLDs for long-duration signals presented in narrowband masking noise (Grose *et al.*, 1995; Grose *et al.*, 1997). The results of a previous study (Hall *et al.*, 2004) suggested that the reduced MLDs of children for narrowband noise maskers were related to a poor ability to take advantage of the binaural information occurring in the masker envelope minima, where the signal-to-noise ratio is most favorable. One interpretation of the present results is that children are not as adept as adults in weighting the temporal epochs associated with the most favorable binaural detection cues. This interpretation is also consistent with the reduced MLDs of children for narrowband masking noise, where optimal performance hinges upon the weighting of the good binaural cues that are present during the relatively brief masker envelope minima.

There are interesting parallels between the developmental results that have been obtained on monaural and binaural temporal processing. For example, the finding by Grose *et al.* (1995) that children were relatively poor in listening in the envelope minima of a monaural, amplitude-modulated narrowband noise is analogous to the finding that children have a reduced ability to exploit binaural information in the envelope minima of a narrowband noise (Hall *et al.*, 2004). A further correspondence between these monaural and binaural findings is that neither of the effects may be driven by an essential deficit in temporal acuity: monaural TMTF results indicated no developmental difference in the monaural time constant (Hall and Grose, 1994), and some aspects of the present results are consistent with an interpretation that there is no developmental difference in the time constant of the binaural temporal window. We plan to investigate possible developmental parallels between monaural and binaural temporal resolution further by examining the monaural temporal window using a method that is analogous to the one used here to investigate the binaural temporal window. One possible approach would be that used by Kollmeier and Gilkey (1990), where the detection of a brief monaural signal is obtained as a function of its temporal relation to an abrupt 15 dB *level* transition in the monaural masker.

Deficits in auditory perception that occur despite apparent acuity in basic auditory functions, such as temporal and frequency resolution, are sometimes "accounted for" in terms of processing efficiency. The concept of poor processing efficiency has some utility in that it can guide the search for explanation away from the peripheral encoding of sound and toward the central analysis of that encoding. However, a significant limitation to the utility of the concept of processing efficiency is that the nature of any apparent inefficiency is often unspecified. The results of the present study may represent a step forward in this regard, as they suggest that an

important component of inefficiency in the binaural hearing of children may be related to a reduced ability to optimize temporal weighting in the analysis of binaural sound sequences where the signal-to-noise ratio changes dynamically. Such dynamic changes can occur in narrowband noise maskers due to the fact that the inherent, pronounced fluctuations of the masking stimulus cause the signal-to-noise ratio to vary markedly over time. In the present study, such dynamic changes occurred in a wideband noise due to the imposition of an abrupt transition in the masker interaural phase. The ability to weight dynamic sequences of binaural information in an optimal way may be associated with a relatively protracted auditory development.

VI. CONCLUSIONS

1. The developmental differences found here for the dynamic masker conditions are consistent either with an interpretation that children have a binaural temporal window with a relatively long leading edge ($\tau 2$) or with an interpretation that children place the binaural temporal window relatively late with respect to the timing of the signal. Fitting procedures that allowed the temporal position of the binaural temporal window to vary and analyses that related the pattern of data for the dynamic maskers to the pattern of data for the steady masker were more consistent with the interpretation that children place the binaural temporal window late.

2. The interpretation that children are relatively poor in optimizing the temporal weighting of dynamically changing binaural information is also consistent with (a) previous findings suggesting that children have a reduced ability to exploit the robust binaural cues available in the masker envelope minima when a long-duration signal is presented in a narrowband noise; and (b) the present finding that children show adult-like MLDs for long-duration signals in wideband noise but reduced MLDs for brief signals in wideband noise.

ACKNOWLEDGMENTS

This work was supported by NIH, RO1 DC00397. We thank Madhu B. Dev and Heidi Reklis for assistance in running subjects and technical support. We thank Michael Akeroyd and an anonymous reviewer for many helpful suggestions on a previous version of this manuscript. We also thank the associate editor, Armin Kohlrausch, for thoughtful suggestions that led to improvements in our approach to data analysis and interpretation.

ANSI (**1996**). "ANSI S3.6-1996, "Specification for audiometers," (American National Standards Institute, New York).

Bernstein, L. R., Trahiotis, C., Akeroyd, M. A., and Hartung, K. (**2001**). "Sensitivity to brief changes of interaural time and interaural intensity," J. Acoust. Soc. Am. **109**, 1604–1615.

Bos, C. E., and de Boer, E. (**1966**). "Masking and discrimination," J. Acoust. Soc. Am. **39**, 708–715.

Buss, E., Hall, J. W., and Grose, J. H. (**2003**). "The masking level difference for signals placed in masker envelope minima and maxima," J. Acoust. Soc. Am. **114**, 1557–1564.

Buus, S., Zhang, L., and Florentine, M. (**1996**). "Stimulus-driven, time-varying weights for Comodulation Masking Release," J. Acoust. Soc. Am. **99**, 2288–2297.

Carlyon, R. P., Buss, S., and Florentine, M. (**1989**). "Comodulation masking release for three types of modulators as a function of modulation rate," Hear. Res. **42**, 37–46.

Culling, J. F., and Summerfield, A. Q. (**1998**). "Measurement of the binaural temporal window using a detection task," J. Acoust. Soc. Am. **103**, 3540–3553.

Durlach, N. I. (**1963**). "Equalization and cancellation theory of binaural masking-level differences," J. Acoust. Soc. Am. **35**, 1206–1218.

Grantham, D. W., and Wightman, F. L. (**1979**). "Detectability of a pulsed tone in the presence of a masker with time-varying interaural correlation," J. Acoust. Soc. Am. **65**, 1509–1517.

Grantham, W., and Wightman, F. L. (**1978**). "Detectability of varying interaural temporal differences," J. Acoust. Soc. Am. **63**, 511–523.

Grose, J., Hall, J., and Dev, M. (**1997**). "MLD in children: Effects of signal and masker bandwidths," J. Speech Lang. Hear. Res. **40**, 955–959.

Grose, J. H., Hall, J. W., and Gibbs, C. (**1993**). "Temporal analysis in children," J. Speech Hear. Res. **36**, 351–356.

Grose, J. H., and Hall, J. W., III (**1998**). "Masker fluctuation and the masking-level difference," J. Acoust. Soc. Am. **103**, 2590–2594.

Grose, J. H., Hall, J. W., and Mendoza, L. (**1995**). "Developmental effects in complex sound processing," in *10th International Symposium on Hearing*, edited by G. A. Manley, G. M. Klump, C. Koppl, H. Fastl, and H. Oekinghaus (World Scientific, Singapore), pp. 97–104.

Hall, J. W., Buss, E., Grose, J. H., and Dev, M. B. (**2004**). "Developmental effects in the masking-level difference," J. Speech Lang. Hear. Res. **47**, 13–20.

Hall, J. W., and Grose, J. H. (**1990**). "The masking-level difference in children," J. Am. Acad. Audiol **1**, 81–88.

Hall, J. W., and Grose, J. H. (**1994**). "Development of temporal resolution in children as measured by the temporal modulation transfer function," J. Acoust. Soc. Am. **96**, 150–154.

Hall, J. W., Grose, J. H., and Hartmann, W. M. (**1998**). "The masking-level difference in low-noise noise," J. Acoust. Soc. Am. **103**, 2573–2577.

Hall, J. W., Haggard, M. P., and Fernandes, M. A. (**1984**). "Detection in noise by spectro-temporal pattern analysis," J. Acoust. Soc. Am. **76**, 50–56.

Hirsh, I. J. (**1948**). "Influence of interaural phase on interaural summation and inhibition," J. Acoust. Soc. Am. **20**, 536–544.

Holube, I., Kinkel, M., and Kollmeier, B. (**1998**). "Binaural and monaural auditory filter bandwidths and time constants in probe tone detection experiments," J. Acoust. Soc. Am. **104**, 2412–2425.

Irwin, R. J., Ball, A. K., Kay, N., Stillman, J. A., and Bosser, J. (**1985**). "The development of auditory temporal acuity in children," Child Dev. **56**, 614–620.

Kollmeier, B., and Gilkey, R. (**1990**). "Binaural forward and backward masking: Evidence for sluggishness in binaural detection," J. Acoust. Soc. Am. **87**, 1709–1719.

Moore, B. C. J., and Glasberg, B. R. (**1983**). "Suggested formulae for calculating auditory filter bandwidths and excitation patterns," J. Acoust. Soc. Am. **74**, 750–753.

Viemeister, N. F. (**1979**). "Temporal modulation transfer functions based upon modulation thresholds," J. Acoust. Soc. Am. **66**, 1364–1380.

Wightman, F., Allen, P., Dolan, T., Kistler, D., and Jamieson, D. (**1989**). "Temporal resolution in children," Child Dev. **60**, 611–624.

Individual differences in the masking level difference with a narrowband masker at 500 or 2000 Hz

Emily Buss,[a] Joseph W. Hall III, and John H. Grose
Department of Otolaryngology/Head and Neck Surgery, University of North Carolina School of Medicine, Chapel Hill, North Carolina 27599

(Received 2 August 2005; revised 31 October 2006; accepted 1 November 2006)

The masking level difference (MLD) for a narrowband noise masker is associated with marked individual differences. This pair of studies examines factors that might account for these individual differences. Experiment 1 estimated the MLD for a 50 Hz wide band of masking noise centered at 500 or 2000 Hz, gated on for 400 ms. Tonal signals were either brief (15 ms) or long (200 ms), and brief signals were coincident with either a dip or peak in the masker envelope. Experiment 2 estimated the MLD for both signal and masker consisting of a 50 Hz wide bandpass noise centered on 500 Hz. Signals were generated to provide only interaural phase cues, only interaural level cues, or both. The pattern of individual differences was dominated by variability in NoSπ thresholds, and NoSπ thresholds were highly correlated across all conditions. Results suggest that the individual differences observed in Experiment 1 were not primarily driven by differences in the use of binaural fine structure cues or in binaural temporal resolution. The range of thresholds obtained for a brief NoSπ tonal signal at 500 Hz was consistent with a model based on normalized interaural correlation. This model was not consistent for analogous conditions at 2000 Hz. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2400849]

PACS number(s): 43.66.Pn, 43.66.Dc, 43.66.Mk [AK] Pages: 411–419

I. INTRODUCTION

The masking level difference (MLD) is the detection advantage obtained for some stimuli when the signal and the masker are presented to the two ears with different interaural characteristics (Hirsh, 1948). In the typical example of an MLD, thresholds for a diotic stimulus, with the signal and masker in phase at the two ears (NoSo), are compared to thresholds for a diotic masker and a signal that is out of phase at the two ears (NoSπ). Thresholds in the NoSπ condition can be substantially lower than those in the NoSo condition, particularly at low signal frequencies. Large individual differences have been reported in the MLD when the maskers are spectrally narrow (e.g., Bernstein *et al.*, 1998).

There are several lines of evidence suggesting that inherent fluctuations of narrowband maskers may affect the magnitude of the MLD. Grose and Hall (1998) estimated temporal weighting functions which showed that binaural cues impacted performance differently depending on the masker envelope: Portions of the signal coincident with masker envelope dips were given greater perceptual weight in the NoSπ detection task than other portions of the signal coincident with peaks or intermediate values. No such preferential dip weighting was observed for the NoSo condition. A follow-up study measured detection thresholds for a brief pure tone signal, with 15 ms $\cos^2$ ramps, and a 50 Hz wide band of Gaussian masking noise, both centered on 500 Hz. The brief signal was coincident with either a masker envelope dip or peak. Thresholds in the NoSo condition did not vary reliably with signal placement, while those in the NoSπ condition were on the order of 10 dB lower in a dip than a peak, a result that was referred to as a "dip advantage." In other words, the MLD was larger for a signal presented coincident with a masker envelope dip (Buss *et al.*, 2003; Hall *et al.*, 2004). These results are consistent with the conclusion that detection of a long duration Sπ signal is based on the interaural cues coincident with modulation dips in an No masker.

As is typical in experiments on the MLD, the studies described above used low-frequency narrowband stimuli. Therefore, stimulus fine structure almost certainly played an important role in these results (e.g., Zurek and Durlach, 1987). While the MLD is typically larger at low than high frequencies, substantial MLDs can also be obtained at frequencies above those for which interaural changes in fine structure are thought to be useful in binaural hearing, a finding that is attributed to the use of interaural envelope difference cues (McFadden and Pasanen, 1978). It has been suggested (Bernstein and Trahiotis, 1996) that the utilization of interaural fine structure and interaural envelope differences may rely on the same underlying cue—detection of a change in interaural correlation. The finding of a larger MLD with a narrowband masker at low as compared to high frequencies has been argued to reflect the large contribution of fine structure cues at low frequencies (van de Par and Kohlrausch, 1997).

Addition of a brief Sπ signal to an No narrowband masker dip or peak can introduce binaural fine structure and envelope differences, depending on the interactions between the signal and masker stimulus components. This is illustrated in Fig. 1 with stimuli from Buss *et al.* (2003). In the top row of panels, solid grey waveforms show representative masker samples for conditions in which a brief signal is co-

[a]Author to whom correspondence should be addressed. Telephone: (919) 966-8019. Electronic mail: ebuss@med.unc.edu

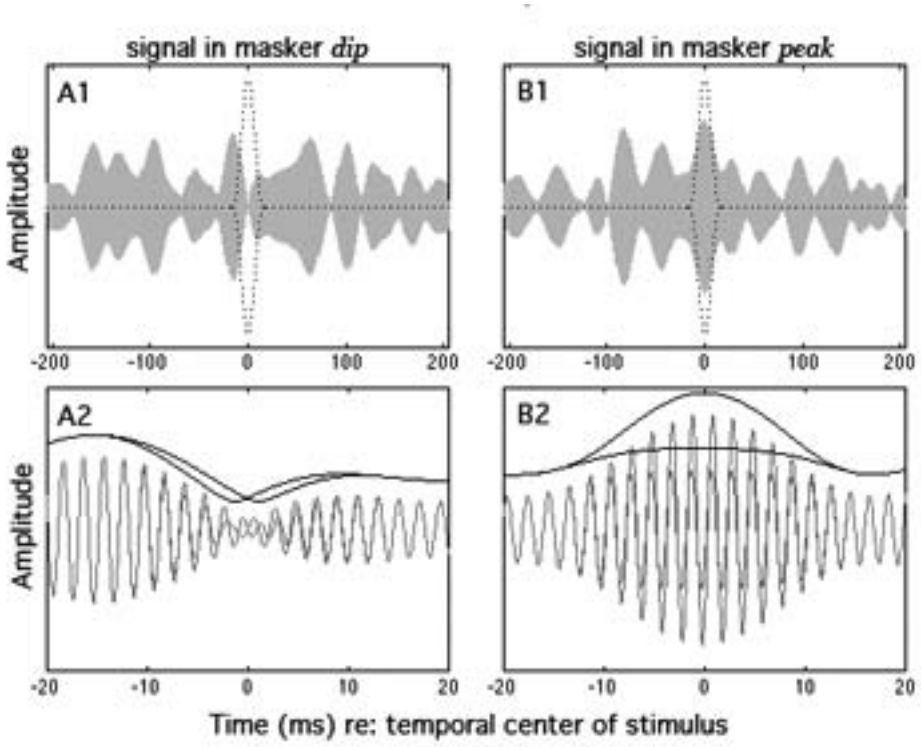


FIG. 1. Illustration of interaural cues available at threshold for a brief signal in a narrowband masker in the NoSπ condition, based on stimuli and thresholds from Buss *et al.* (2003). The left column of panels corresponds to a signal presented in a masker envelope dip and the right column to a signal presented in a masker envelope peak. The top row shows amplitude of a representative masker as a function of time, with dotted lines indicating the temporal position (but not the level) of an added signal. The bottom row of panels is a magnified view of the stimuli presented to the left and right ears (overlaid traces) with a diotic masker (as shown in the top row) and an $S\pi$ signal at threshold (60 dB for the dip condition and 70 dB SPL for the peak condition). Grey lines indicate the temporal fine-structure and dark lines show the raised envelope.

incident with a masker envelope dip (A1) and a masker envelope peak (B1). The dotted lines indicate the temporal placement (but not the amplitude) of the brief tonal signal. For a 72 dB SPL masker, thresholds in the NoSπ condition are on the order of 60 dB SPL in the *dip* condition and 70 dB SPL in the *peak* condition (Buss *et al.*, 2003). Panels A2 and B2 show a magnified view of the stimulus presented to the left and right ear in the NoSπ condition (overlaid traces), with a signal at the associated threshold level. The dark lines indicate the temporal envelopes of the left and right ear stimuli, which have been raised for visual clarity.

Two features of these stimuli are noteworthy. First, the fine structure of the composite signal-plus-masker waveforms presented to the left and right ears is approximately out-of-phase in the center of the masker dip (panel A2) and only slightly phase delayed in the masker peak (panel B2). Generating 100 random samples of the masker and adding a signal at threshold (60 dB for dips or 70 dB for peaks), interaural phase approaches 180° in the middle of the masker dip for all 100 samples, while the maximum interaural phase difference was approximately 30° in the *peak* condition. Summarized in this way, the interaural fine structure cue at threshold could be characterized as less salient in the *peak* than the *dip* condition. Recall that the detection of a long-duration low-frequency Sπ signal in a narrowband No masker is thought to be based largely on fine structure cues (van de Par and Kohlrausch, 1997), and that NoSπ detection for such a long duration tone is argued to be based on cues in the masker dips (Buss *et al.*, 2003). This implies that NoSπ thresholds for a brief signal in the *dip* condition should likewise rely on interaural fine structure cues. If fine structure cues in a peak are not as salient and informative as those in a dip, this opens up the possibility that interaural envelope cues may play an increased or dominant role in the NoSπ, *peak* condition. If this speculation is correct, then repeating the experiment at a higher frequency, where fine structure is no longer available, should eliminate or greatly reduce the dip advantage by virtue of eliminating the interaural fine structure cues in the dip, while leaving unchanged the envelope cues in the peak. Experiment 1 tested this prediction; in addition, performance was compared across signal conditions and frequencies in an attempt to better characterize individual differences obtained in the narrowband masker MLD.

A second feature of note in the stimuli shown in Fig. 1 is the fact that the temporal epoch containing optimal information about a brief signal is by its nature brief, suggesting that fine temporal resolution in deriving a cue to the presence of a signal could result in better performance. There is a wide range of estimates of binaural temporal resolution in the literature. Some data suggest that the binaural system is sluggish and integrates information over hundreds of milliseconds (e.g., Grantham and Wightman, 1979), while other data suggest more precise resolution (e.g., Bernstein *et al.*, 2001). Much of this variability is likely due to the different cues characterizing the signal interval across experimental paradigms (Holube, Kinkel, and Kollmeier, 1998). Grantham (1984) suggested that binaural temporal resolution may be different for interaural time and interaural level cues, further complicating the issue. If observers differ in the temporal resolution with which they are able to process brief interaural cues, then this could affect their ability to make use of brief cues and this factor could underlie the individual differences observed for the narrowband MLD. Experiment 2 estimated narrowband MLD for fluctuating and stable binaural cues to test this possibility.

II. EXPERIMENT 1

Experiment 1 tested the hypothesis that the dip advantage previously demonstrated at 500 Hz would not be found at 2000 Hz. Further, it was hypothesized that the individual differences obtained with long-duration 500 Hz stimuli would be related to those found for brief signals coincident with masker dips. This result would support previous claims that detection with a long duration signal is dominated by cues coincident with dips at 500 Hz (Buss *et al.*, 2003). Further, if the underlying detection cues associated with interaural fine structure and interaural envelope differences are the same, then the pattern of individual differences should be consistent across the 500 and 2000 Hz signal frequencies.

A. Methods

1. Observers

Observers were 28 adults, 18–49 years of age, with thresholds 15 dB HL or better at octave frequencies 250–8000 Hz, and no reported history of ear disease. Of this group, 14 had previously participated in psychoacoustic studies. Five observers had experience listening to MLD stimuli. These five observers spanned the range of individual differences in NoSπ performance. For example, in the NoSπ condition for a brief 500 Hz signal positioned in a masker envelope dip, one of these five had the lowest threshold and another had the second highest threshold with respect to the entire group of 28 observers.

An additional three observers were omitted from the study because of excessive threshold variability, defined as 10 dB or more variation across thresholds in a single condition. It should be noted that these cases of threshold variability did not resemble simple practice effects, in that no clear trends towards improvement were evident. All other data are reported below.

2. Stimuli

The signal was either a 500 Hz or a 2000 Hz pure tone, with random starting phase. In the *peak* and *dip* conditions, the signal was of brief duration, ramped on and off with 15 ms $\cos^2$ ramps (no steady state), and temporally centered in the masker. In the *long* condition, the signal was 200 ms in duration and was gated on with 15 ms $\cos^2$ ramps, the onset occurring 100 ms after masker onset.

Maskers were 50 Hz wide bands of noise centered on the signal frequency (either 500 or 2000 Hz), with a level of 72 dB SPL and duration of 409 ms, including 15 ms $\cos^2$ ramps. Bands of noise were generated digitally in the frequency domain as an array of 2^{14} points, with independent Gaussian draws assigned to the real and imaginary components at points within the 50 Hz passband. This array was transformed into the time domain via IFFT. At the DAC rate of 24.4 kHz used here, the resulting waveform was approximately 670 ms in duration and could be repeated seamlessly (no discontinuity between the ending and starting points of the array). Three copies of the noise array were concatenated. For the *peak* condition, the maximum envelope value of the middle sample was identified, and the extended array was truncated in such a way as to place that point in the temporal center of a 409 ms stimulus. Similarly, for the *dip* condition the minimum was identified and placed in the temporal center of a 409 ms stimulus. A new masker, based on independent samples of Gaussian noise, was generated prior to every stimulus presentation, both within and across trials. Thresholds were estimated for NoSo and NoSπ in all three conditions. These stimuli are functionally identical to those used by Buss *et al.* (2003) and those discussed in the introduction.

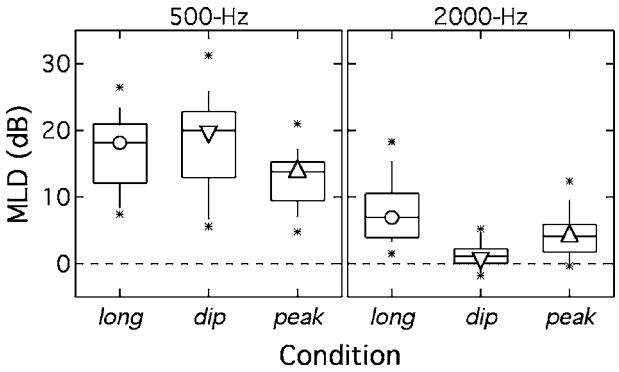


FIG. 2. The distribution of MLDs is plotted for 500 Hz (left panel) and 2000 Hz (right panel) conditions. Abscissa labels and symbols indicate condition. The median of each distribution is indicated with a symbol, the span between the 25th and the 75th percentiles is indicated with a box, the 10–90th percentile with bars, and the maximum and minimum of each distribution with stars.

3. Procedure

Stimuli were presented using deeply inserted earphones (Etymotic: ER-2). Detection thresholds were estimated using a three-alternative forced-choice, three-down one-up track estimating the 79% point on the psychometric function (Levitt, 1971), with feedback provided visually. In this procedure, three masker intervals were presented with a temporal separation of 500 ms. The signal was added to the masker in one randomly chosen interval, and the observer's task was to indicate which interval contained the signal. The signal level was adjusted in steps of 4 dB until two track reversals were obtained and then in steps of 2 dB for the remaining six reversals. Threshold estimates were computed as the average signal level at the last six track reversals. Conditions were run in blocks, with the order of blocks randomized across observers. A minimum of three threshold estimates were obtained in each condition, and a fourth was obtained if the first three spanned a range of 3 dB or more. The average of 3 to 4 estimates is reported.

B. Results

The distribution of MLDs is shown in Fig. 2. Panels indicate data for the 500 Hz (left) and 2000 Hz (right) conditions. Symbols indicate the median, boxes indicate the 25th-to-75th percentile span, bars indicate the 10th-to-90th percentile span, and stars show the maximum and minimum MLDs. Conditions are indicated on the abscissa. As suggested in this figure, there were substantial individual differences in some of the conditions. For example, the median MLD for a long duration signal at 500 Hz was 18.1 dB, with

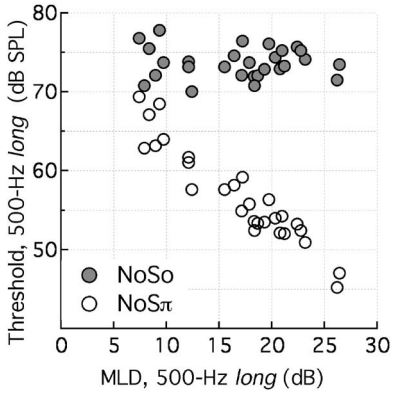


FIG. 3. For each observer, NoSo and NoSπ thresholds in the 500 Hz *long* condition are plotted in dB SPL as a function of the MLD in the 500 Hz *long* condition, in dB.

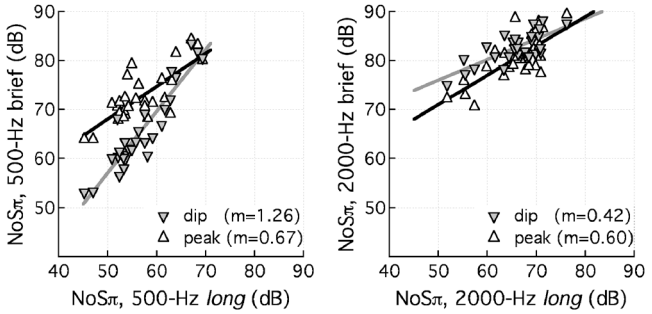


FIG. 4. NoSπ thresholds for brief signals are plotted as a function of NoSπ thresholds in the *long* condition. The left panel indicates data for the 500 Hz conditions, and the right for the 2000 Hz conditions. Upward-pointing triangles indicate *peak* data, downward-pointing triangles indicate *dip* data, and lines indicate the best linear fit to the data (grey=dip; black=peak). The *m*-values indicated in the lower right corner of each panel show the slopes of these linear fits.

values for individuals ranging from 7.5 to 26.5 dB. The median MLD for a long duration signal at 2000 Hz was 6.9 dB, with values for individuals ranging from 1.6 to 18.3 dB.

The MLD data (NoSo-NoSπ) were submitted to a repeated measures ANOVA, with two levels of FREQUENCY (500 and 2000 Hz) and three levels of CONDITION (*long, dip*, and *peak*). There was a main effect of FREQUENCY ($F_{1,27}=277.9$, $p<0.0001$), reflecting the fact that MLDs were larger at 500 than 2000 Hz. There was also a significant main effect of CONDITION ($F_{1,27}=5.42$, $P<0.05$) and a significant interaction of FREQUENCY and CONDITION ($F_{1,27}=41.53$, $p<0.001$). *Post hoc* contrasts indicated that at 500 Hz the MLDs associated with the *long* and the *dip* conditions were not different ($p=0.17$), but that the MLDs in both of these conditions were significantly greater than the MLD in the *peak* condition ($p<0.005$). At 2000 Hz, the MLDs were significantly different in all three conditions (*long*>*peak*>*dip*; $p<0.005$). Two-tailed t-tests with Bonferroni adjustment for multiple tests confirmed that the MLD was greater than zero for all six conditions (2 freq×3 cond) at $\alpha=0.01$. In addition to these within-subjects effects, there were substantial individual differences ($F_{1,27}=223.8$, $p<0.0001$).

Closer examination of these MLD data in terms of the constituent NoSo and NoSπ thresholds suggests that individual differences are dominated by across-observer variation in NoSπ (as opposed to NoSo) thresholds. Figure 3 shows constituent thresholds in the 500 Hz, *long* condition, plotted as a function of the 500 Hz, *long* MLD. Individual differences in the MLD for a long duration, 500 Hz tone were highly correlated with NoSπ thresholds ($R=-0.95$, $p<0.0001$), but not with NoSo thresholds ($R=-0.06$, $p=0.75$). This indicates that individual differences in the MLD are driven by variability in the NoSπ condition. Across all six conditions of Experiment 1, correlations between NoSo thresholds and associated values of the MLD ranged from $R=-0.02$ to $R=-0.44$. In contrast, correlations between NoSπ thresholds and associated values of the MLD were much higher, ranging from $R=-0.84$ (2000 Hz, *peak*) to $R=-0.96$ (500 Hz, *dip*). The one exception was the 2000 Hz, *dip* condition, where the correlation was only $R=-0.59$. The relatively weak correlation in this condition may be due to the reduced inter-observer variance and small values of the MLD.

Figure 4 shows the thresholds for brief NoSπ signals as a function of the NoSπ thresholds for the long duration signal. At 500 Hz (left panel), those observers with the highest (worst) thresholds in the *long* condition showed comparable thresholds in the *dip* and *peak* conditions. In contrast, those with the lowest (best) thresholds in the *long* condition tended to show the biggest dip advantage. The relationship between performance in the brief and long conditions is best illustrated in terms of the slope of the regression line that best fits each set of data. At 500 Hz the slope of the best-fitting line was significantly greater than one for the *dip* condition and less than one for the *peak* condition (at $\alpha=0.005$). In contrast, at 2000 Hz there is a "peak advantage" for better-performing observers. Slopes of the best-fitting lines were less than one for both *dip* and *peak* conditions (at $\alpha=0.005$). The slopes were compared by performing a regression analysis on the difference between thresholds in the *dip* and *peak* condition for each observer, as a function of threshold in the *long* condition for that observer. The slope of the regression line for this difference variable was significantly less than zero, indicating that the dip/peak difference decreased significantly with increasing thresholds in the *long* condition ($F_{1,26}=4.39$, $p<0.05$).

As suggested by Fig. 4, NoSπ thresholds in the brief signal conditions are positively correlated with those in the associated *long* condition. For the 500 Hz data, those correlations are $R=0.92$ and $R=0.77$ for *dip* and *peak* conditions, respectively, and the correlation between thresholds in the *dip* and *peak* conditions is $R=0.80$. Corresponding correlations for 2000 Hz data are $R=0.75$, $R=0.73$, and $R=0.83$. This suggests some degree of uniformity in individual differences observed across conditions within frequency. Further, NoSπ thresholds in the *long* duration condition are correlated across the 500 and 2000 Hz signal frequencies, with $R=0.78$, suggesting that the underlying source of individual differences is not frequency-specific.

C. Discussion

The MLD in the *long* condition was highly variable across observers, with mean (and standard deviations) of

17.2 (5.8) and 8.1 (4.5) dB at 500 and 2000 Hz, respectively. These results are comparable to those of Bernstein *et al.* (1998), who reported means of 15.8 (4.7) and 5.8 (3.1) at 500 and 4000 Hz, respectively. As noted by Bernstein *et al.*, individual differences in the MLD for a long duration signal are more strongly associated with NoSπ (as compared to NoSo) thresholds.

The pattern of MLDs in the 500 Hz *dip* and *peak* conditions replicates the results of Buss *et al.* (2003), showing the dip advantage. The pattern of MLDs was quite different at 2000 Hz, where results indicate a slight peak advantage. As in the previous study, MLDs in the 500 Hz *dip* condition were comparable to those in the *long* condition, suggesting that much of the MLD in the *long* condition could be based on use of binaural cues present in the dip. In contrast, at the 2000 Hz frequency the MLD in the *long* condition exceeded that in either the *dip* or the *peak* condition. This result was not predicted. It is possible that neither the *dip* nor the *peak* condition represents the optimal placement for a signal at 2000 Hz. If the optimal signal placement were at an intermediate point in the masker envelope, associated with more pronounced interaural envelope cues, then binaural release would be larger in the *long* condition where such cues were available (as opposed to the *dip* or *peak* condition). This result may also be associated with the finding of greater temporal integration of binaural information at high frequencies where temporal fine-structure is not available as a cue (Bernstein and Trahiotis, 1999).

As expected, there were extensive individual differences at both frequencies. At 500 Hz, the 'dip advantage' for NoSπ thresholds was largest for those observers with lowest thresholds in the 500 Hz *long* condition. Analogously, the peak advantage for NoSπ thresholds at 2000 Hz was largest for those observers who attained the lowest thresholds in the 2000 Hz *long* condition. Whereas it is difficult to assess the relative importance of fine structure versus envelope cues in the peak condition at 500 Hz, the pattern of results obtained here is consistent with an interpretation that good performance in the *dip* condition at 500 Hz is based on differences in sensitivity to interaural fine structure cues. In contrast, good performance in the *peak* condition at 2000 Hz must be associated with cues related to interaural envelope differences. The fact that better-performing observers demonstrate a more pronounced dip/peak difference at both frequencies suggests a range of performance in utilization of both cues.

III. EXPERIMENT 2

One possible interpretation of the results of Experiment 1 is that the individual differences in the MLD can be explained in terms of differences in binaural temporal resolution. If the binaural cue coincident with a masker envelope dip is superior to the binaural cue coincident with a masker envelope peak, then thresholds for a long duration tone should be determined in part by the extent to which the detection is based on cues derived from those brief epochs. Degradation in the temporal specificity with which those cues are derived would be associated with degradation in those cues. Alternatively, the individual differences in the pure-tone MLD may be driven by a factor that is largely independent of the binaural temporal resolution.

Experiment 2 assessed the role of temporal resolution in the use of interaural phase and level cues, and compared the pattern of individual differences with that observed in Experiment 1. Narrowband stimuli were constructed in such a way as to minimize the possible effects of binaural temporal resolution on the detection process. For some of the stimuli used here, the magnitude of the interaural cue resulting from signal-masker interaction was constant across the entire duration of a 200 ms signal presentation. With consistency of the binaural cue across masker dips and peaks, binaural temporal resolution should not be a relevant factor for signal detection. Thus, if binaural temporal resolution were not an important factor contributing to the individual differences observed in Experiment 1, we would expect to see a similar pattern of individual differences in the present experiment.

A. Methods

1. Observers

A subset of 12 observers from Experiment 1 participated. Effort was made to ensure that the observers recruited for this study represented a range of binaural performance, as estimated in Experiment 1. Represented in this group were the observers with the lowest and the highest MLD in the 2000 Hz *long* condition and the observers with lowest and the second highest MLD in the 500 Hz *long* condition.

2. Stimuli

As in Experiment 1, maskers were 50 Hz wide bands of noise centered on the signal frequency (500 Hz), presented at 72 dB SPL. In contrast to the stimulus generation procedures of Experiment 1, maskers in this experiment were generated via multiplication in quadrature in the time domain (for discussion of this method in binaural psychoacoustics, see Amenta *et al.*, 1987). Two independent samples of 25 Hz, lowpass Gaussian noise were generated, $n_A(t)$ and $n_B(t)$. Each noise was multiplied by a 500 Hz tone, where one tone was 90° phase-advanced relative to the other. The two products were summed to produce a bandpass Gaussian noise masker sample, M

$$M(t) = n_A(t) \cdot \cos(\omega t) + n_B(t) \cdot \sin(\omega t).$$

Like the masker, the signal (S) was also a band of Gaussian noise, 50 Hz wide, centered on 500 Hz and generated by multiplication in quadrature. In the ILD condition, the signal was identical to the masker, $S(t)=M(t)$. In the IPD condition, the noise samples used to generate the masker were also used to generate the signal, but the tones multiplied by those samples to generate the signal were 90° phase-shifted relative to those used in generation of the masker

$$S(t) = n_A(t) \cdot \cos\left(\omega t + \frac{\pi}{2}\right) + n_B(t) \cdot \sin\left(\omega t + \frac{\pi}{2}\right).$$

In the RAN condition the signal was generated based on independent samples of lowpass noise. That is

$$S(t) = n_C(t) \cdot \cos(\omega t) + n_D(t) \cdot \sin(\omega t),$$

where $n_A \neq n_C$ and $n_B \neq n_D$. Thus, in the RAN condition the signal and masker were independent samples of bandpass Gaussian noise. In all three signal conditions, stimuli in the NoSo condition were diotic, with the left and right ear defined as $M(t)+\alpha S(t)$, where α controls the amplitude of the signal. In the NoSπ condition, the stimulus delivered to the left ear was defined as $M(t)+\alpha S(t)$, while the stimulus delivered to the right ear was defined as $M(t)-\alpha S(t)$.

As in the *long* conditions of Experiment 1, the signal was 200 ms in duration and temporally centered in the masker. In contrast to Experiment 1, the signal was ramped on and off with 25 ms (as compared to 15 ms) $\cos^2$ ramps, a change that was instituted to guard against spectral artifacts.[1]

For NoSπ signal presentation, these conditions are associated with an interaural level cue only (ILD), an interaural phase cue only (IPD) or both types of cues (RAN). The approach of isolating IPD and ILD cues is similar to that taken by Jeffress and McFadden (1970). It should also be noted that IPD and ILD cues were constant for the duration of the signal in the associated conditions, but cues were fluctuating over the course of the signal presentation in the RAN condition.

Thresholds for both NoSo and NoSπ were estimated in all three conditions. In the NoSo ILD condition, the signal and the masker were identical and added in phase. In contrast, the signal and masker added in random phase or orthogonal phase in the NoSo RAN and IPD conditions, respectively. As a consequence, a given signal level resulted in a larger increase in energy in the NoSo ILD condition as compared to the RAN and IPD conditions.

3. Procedures

Procedures were similar to those used in Experiment 1. The task was a three-alternative forced-choice with feedback provided visually. The signal level was adjusted in a three-down one-up track estimating 79% correct. Conditions were run in blocks, with the order of blocks randomized across observers. A minimum of three threshold estimates was obtained in each condition, and a fourth was obtained if the first three spanned a range of 3 dB or more.

B. Results

The MLDs from Experiment 2 are shown in Fig. 5, plotted following the same convention as used in Fig. 2. The median MLD for the RAN condition was 19.8 dB. This closely resembles MLD in the 500 Hz *long* condition of Experiment 1, supporting the implicit assumption that the cues underlying the MLD for the tonal signal of Experiment 1 can be characterized using a narrowband signal. The median MLD for the IPD condition was 16.3 dB, and that for the ILD condition was 4.5 dB.

Thresholds in the NoSo condition were relatively consistent across the three conditions when expressed in terms of ΔL. Median thresholds in the RAN and IPD conditions were 73.1 and 73.6 dB SPL, respectively. Because masker and signal added in random or orthogonal phase, the level of a masker-plus-signal at threshold was approximately 3.3 dB higher than the 72 dB level of a masker alone. The median threshold in the ILD condition was 64.9 dB SPL. Because masker and signal added in phase, the level of a masker-plus-signal at threshold was approximately 3.1 dB higher than a masker alone. These results are generally in agreement with those of Bos and de Boer (1966) for a similar stimulus bandwidth.

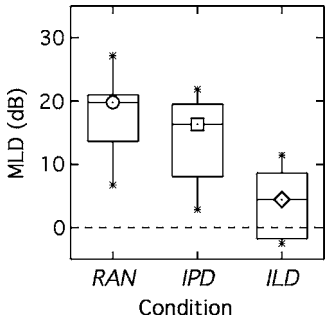


FIG. 5. The distributions of MLDs in Experiment 2 using narrowband noise signals are plotted following the conventions of Fig. 2. Due to differences in signal stimuli, NoSo thresholds were not constant across conditions. As a result, these values of MLD do not solely reflect differences in the NoSπ threshold.

Because NoSo thresholds were not constant across conditions when expressed in terms of the level of the signal, differences in the MLD across conditions do not solely reflect differences in NoSπ thresholds. The median NoSπ threshold was 55.8 dB in the RAN condition, 60.7 dB in the IPD condition, and 62.6 dB in the ILD condition. A repeated measures ANOVA resulted in a significant effect of condition ($F_{2,22}=47.18$, $p<0.0001$). *Post hoc* contrasts revealed that NoSπ thresholds in the RAN condition were lower than those in the IPD condition ($p<0.001$), and thresholds in the IPD condition were lower than those in the ILD condition ($p<0.05$).

As in conditions with pure tone signals from Experiment 1, MLDs with narrowband signals are associated with marked individual differences, and the variability in the MLD was more highly correlated with NoSπ than NoSo thresholds (mean $R=-0.91$ and $R=-0.26$, respectively). Individual differences in NoSπ thresholds were highly correlated with those in Experiment 1. Figure 6 shows NoSπ threshold in the narrowband noise conditions of Experiment 2 as a function of associated thresholds in the 500 Hz *long* condition of Experiment 1. Correlations between NoSπ thresholds in the three conditions tested here and the six conditions of Experiment 1 were all significant ($p<0.005$, one-tailed) and ranged between $R=0.72$ and $R=0.95$, with an average correlation of approximately $R=0.85$ for each of the three narrowband signal conditions. These correlations are comparable to those among NoSπ thresholds from Experiment 1.

C. Discussion

The pattern of individual differences obtained with narrowband signals was quite similar to that measured in Experiment 1 with a pure tone signal, both in the *long* condition and the two conditions utilizing a brief signal (the *dip* and

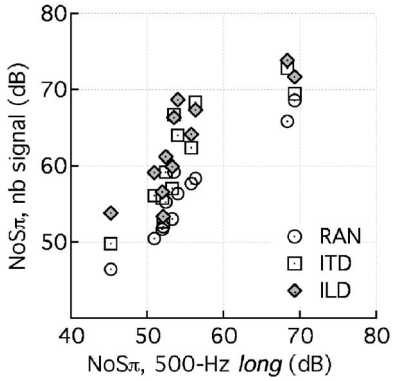


FIG. 6. NoSπ thresholds for narrowband signals of Experiment 2 are plotted as a function of NoSπ thresholds in the 500 Hz *long* condition of Experiment 1. Dotted circles indicate RAN thresholds, squares indicate ITD thresholds, and diamonds indicate ILD thresholds.

peak conditions). A subset of the stimuli in Experiment 2 was designed to minimize the possibility that binaural temporal resolution could have contributed to the detection of the signal. Under these conditions, the pattern of individual differences was quite similar to that found in Experiment 1, with NoSπ thresholds correlating highly across the experiments. Considered together, these results suggest that it is unlikely that binaural temporal resolution plays a material role in the individual differences that characterize the MLD for narrowband masking noise. Instead, the results suggest that there are "good performers" and "poor performers," a categorization that applied in a similar way in all binaural (NoSπ) conditions of Experiments 1 and 2, irrespective of the temporal fluctuation of the binaural cue. Comparable results in MLD, interaural time discrimination and interaural intensity discrimination conditions have been reported previously (e.g., Koehnke *et al.*, 1986).

The approach taken here bears some resemblance to that of Bernstein *et al.* (1998). In that study, individual differences in sensitivity were compared for narrowband MLD and for interaural time and level differences. Individual differences across conditions were weakly correlated ($R \approx 0.4$) when compared to the current data ($R \approx 0.8$). One possible reason for this discrepancy is that Bernstein *et al.* used an interaural difference paradigm that featured a 400 Hz wide stimulus and compared it to NoSπ thresholds for a 50 Hz wide masker, whereas the current experiments used a 50 Hz bandwidth for both paradigms. Given the closer correspondence of the conditions used in the current study, it seems likely that individual differences in NoSπ signal detection are closely related to those in the discrimination of interaural phase (or time) and level under comparable conditions.

IV. GENERAL DISCUSSION

The results of Experiment 1 support the hypothesis that the dip advantage observed previously with 500 Hz stimuli does not occur for comparable stimuli at 2000 Hz, where temporal fine-structure information is no longer useful for binaural hearing. Rather, a peak advantage at 2000 Hz was observed, presumably due to interaural envelope cues in masker peaks. Similar individual differences occurred in both the long duration and short duration NoSπ conditions. Results of Experiment 2 suggest that the individual differences obtained in Experiment 1 are not related to differences in the temporal acuity with which fluctuating IPD and ILD cues are utilized. These results are consistent with an interpretation that limitations in binaural temporal resolution are not responsible for the individual differences in NoSπ thresholds for a narrowband noise masker.

One very common way to characterize the binaural cue underlying the MLD is in terms of the reduction in interaural correlation associated with addition of a signal (e.g., Durlach *et al.*, 1986; for discussion see Bernstein and Trahiotis, 1996). Previous data for a brief 500 Hz tonal signal coincident with masker envelope dips and peaks were well characterized in terms of correlation at the output of a simplified model of the auditory periphery (Buss *et al.*, 2003). That model closely followed the approach developed by Bernstein, Trahiotis and colleagues (e.g., Bernstein *et al.*, 1999; Bernstein and Trahiotis, 1996), but also included a rectangular temporal window centered on the signal. A similar approach was pursued here to see if the individual differences observed could be characterized in terms of differential sensitivity to changes in interaural correlation. In contrast to the previous modeling efforts, a double exponential temporal window was adopted here, similar to the one that has been fitted to binaural temporal window data (e.g., Kollmeier and Gilkey, 1990).

Threshold estimates as a function of differential criterion levels of correlation were obtained with the following procedures, executed in MATLAB. Stimuli were generated using the same algorithm as used in Experiment 1. The brief $S\pi$ signal was added to the diotic masker at a range of S/N ratios, corresponding to signal levels of 50–95 dB SPL, in either the *dip* or the *peak* condition. These stimuli were then submitted to a binaural normalized correlation model which included the following stages: (1) exponential compression of the envelope, with an exponent of 0.23;[2] (2) half-wave, square-law rectification; (3) low-pass filtering;[3] (4) application of the double exponential window centered on the signal;[4] and (5) calculation of the normalized correlation (Bernstein and Trahiotis, 1996). In order to capture possible effects of off-time listening, results were computed for a family of nine temporal window placements, equally spaced between −15 and 15 ms relative to the temporal center of the signal; the correlation estimate associated with each sample was the minimum across these nine possibilities. This procedure was repeated for 1000 masker samples. The first 25 correlation functions in each condition are shown in Fig. 7.

The signal level associated with a criterion correlation spanning 0.45–0.95 (the range associated with observer's thresholds in the 500 Hz *dip* condition) was then determined based on the correlation functions for *dip* and *peak* conditions at both 500 and 2000 Hz. At 500 Hz, results were quite consistent with the pattern of results obtained experimentally: There was a dip advantage overall and the dip/peak differences grew with increases in the criterion level of correlation (with larger values thought to characterize performance of the better-performing subjects).

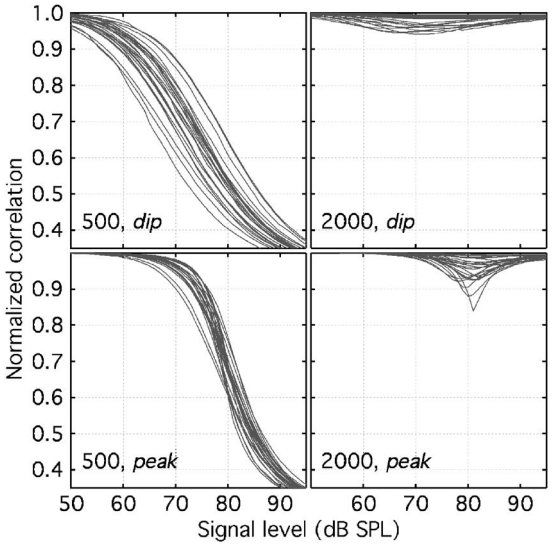


FIG. 7. Normalized correlation for a range of signal levels is plotted for the first 25 masker+signal tokens used to predict thresholds. Data for the 500 Hz stimuli appear in the left column and those for 2000 Hz data in the right; the top row shows results for the *dip* conditions, and the bottom for the *peak* conditions.

This approach was less consistent with 2000 Hz data. Consistent with experimental results, the signal level associated with a given criterion level of correlation was higher for 2000 Hz than 500 Hz conditions. However, for the 2000 Hz peak condition there was no indication of a binaural advantage (i.e., estimated signal levels exceeded NoSo thresholds) in all but two cases. In contrast, thresholds were lower in the NoSπ than the NoSo condition for 27 of 28 observers. Perhaps most inconsistent with empirical data, at a correlation of 0.98, corresponding to a signal level at threshold in the 500 Hz *dip* condition for the two most sensitive observers, thresholds in the 2000 Hz *dip* condition were underestimated by over 10 dB. The reason for this underestimation can be seen in Fig. 7. A criterion correlation of 0.98 corresponds to a signal level of approximately 60 dB SPL in the 2000 Hz *dip* condition and over 70 dB SPL in the 2000 Hz *peak* condition, and thus a dip advantage. This failure to predict a peak advantage at 2000 Hz is quite consistent and remains relatively unchanged with the use of a broader temporal window or a window with less pronounced central emphasis (e.g., a rectangular or a Hanning window).

Thus, constant criterion correlation at threshold appears to capture some trends in the data, particularly at 500 Hz, but appears to be inconsistent with the modest MLD and peak advantage obtained at 2000 Hz. This result is in contrast to those of Bernstein and Trahiotis (1999), where a range of binaural data with long-duration signals was successfully modeled based on constant criterion correlation, including data collected with a 50 Hz wide band of Gaussian noise centered on 500 Hz. This discrepancy suggests that some aspect of the model may not capture the transient high-frequency cues examined in the current study.

The current approach assumes that any difference in binaural hearing between 500 and 2000 Hz is due to the elimination of cues based on fine-structure at the higher frequency. This is a common assumption in the binaural literature (e.g., Zurek and Durlach, 1987), and one that has received empirical support. For example, the MLD for a narrowband masker is comparable at low and high frequencies when the low-frequency fine-structure is preserved in the envelope at high frequencies (van de Par and Kohlrausch, 1997). There is some evidence that temporal processing of binaural information at low and high frequencies may differ in detail, however. For example, sensitivity to dynamic changes in interaural intensity may be better at high than at low frequencies (Grantham, 1984). Temporal integration of binaural information has also been argued to differ fundamentally for frequencies where fine-structure information is available and frequencies where binaural differences are envelope-based (Bernstein and Trahiotis, 1999).

Another feature of the present approach that might be problematic is the assumption that correlation in the dips and peaks are equally useful as detection cues. If, instead, forward masking or some other level-dependent factor rendered correlation in the dips a less effective cue, the result would be an underestimate of thresholds in the *dip* condition. Using wider stimulus bandwidths than those employed here, forward masking has been shown to play a role in the MLD, with NoSπ thresholds following the same decline as a function of temporal separation for masker offset as NoSo thresholds (Deatherage and Evans, 1969; Yama, 1992). Further investigations are currently underway to explore the possible role of forward masking on the ability to make use of binaural information in the dips of narrowband maskers.

Regardless of whether the NoSπ data of Experiment 1 can ultimately be modeled in terms of correlation at threshold, results of Experiment 2 suggest that the predominant source of individual differences does not lie in differential ability to resolve brief interaural correlation cues or in differential sensitivity to IPD and ILD. Across the 12 listeners who participated in both Experiments 1 and 2, correlations between NoSπ thresholds in the 500 Hz *long* condition and those in the narrowband signal conditions associated with steady interaural cues (ILD and IPD) were $R=0.83$. This is comparable to the correlation of $R=0.95$ for the fluctuating-cue RAN condition. Because the signal in the ILD condition of Experiment 2 did not introduce interaural correlation cues other than those at the signal onset and offset, this consistency of individual differences may reflect a more basic aspect of binaural auditory processing. This is consistent with the suggestion that there are good binaural listeners and poor binaural listeners (Bernstein *et al.*, 1998), and that the factor underlying performance is common to both IPD and ILD (Koehnke *et al.*, 1986).

V. CONCLUSIONS

1. The MLD for a brief 500 Hz pure tone signal in a 50 Hz wide band of noise is larger if that signal is coincident with a masker envelope dip as opposed to an envelope peak. This result has been attributed to a dip advantage for utilizing the binaural cue in the NoSπ condition. The dip/peak difference is reversed at 2000 Hz, where there is a slight peak advantage.

2. A similar pattern of individual differences in NoSπ thresholds was obtained for brief and for long duration tones, as well as for 500 and 2000 Hz signal frequencies. This is consistent with the idea that there are good binaural listeners and poor binaural listeners.

3. Thresholds for detection of a narrowband signal associated with nonfluctuating interaural phase or intensity difference cues demonstrated the same pattern of individual differences as seen with pure tones (and fluctuating cues). This result suggests that temporal resolution is not a dominant factor contributing to individual differences for an $S\pi$ tone in No narrowband noise.

4. Individual differences in criterion levels of interaural correlation are consistent with 500 Hz NoSπ thresholds for pure tones presented coincident with masker dips and peaks. The 2000 Hz pure tone data are less consistent with this approach; a cue based on interaural correlation, at least as computed here, is consistent with an elevation in NoSπ thresholds at 2000 Hz, but not with a significant release from masking (MLD) and not with the peak advantage observed empirically.

ACKNOWLEDGMENTS

This work was supported by NIH Grant No. R01 DC00397. A partial report of these data was presented at the 149th meeting of the Acoustical Society of America [Buss, Hall, and Grose (2005), **117**, 2563]. The authors are grateful to Associate Editor Armin Kohlrausch, Bill Hartman, and an anonymous reviewer for helpful comments on earlier versions of this manuscript.

[1]Previous work suggests that spectral cues do not play a significant role in determining thresholds for the brief tonal signal in the paradigm used in Experiment 1 (Buss *et al.*, 2003). While this is likely to also be the case for the 50 Hz bandpass signal in Experiment 2, temporal specificity was not an issue in this paradigm and 25 ms ramps were adopted as a conservative measure to guard against the influence of spectral cues.

[2]Results were not very sensitive to the value of this exponent. Lopez-Poveda *et al.* (2003) have reported different compressive functions at 500 and 2000 Hz—values of 0.23 and 0.28, respectively. Using these different exponents did not materially change the results reported.

[3]Filter parameters are those described by Bernstein and Trahiotis (1996).

[4]This window was taken from Kollmeier and Gilkey (1990). The mean of the time constant measured in that study (τ=25 ms) was adopted.

Amenta, C. A., III, Trahiotis, C., Bernstein, L. R., and Nuetzel, J. M. (**1987**). "Some physical and psychological effects produced by selective delays of the envelope of narrow bands of noise," Hear. Res. **29**, 147–161.

Bernstein, L. R., and Trahiotis, C. (**1996**). "On the use of the normalized correlation as an index of interaural envelope correlation," J. Acoust. Soc. Am. **100**, 1754–1763.

Bernstein, L. R., and Trahiotis, C. (**1999**). "The effects of signal duration on NoSo and NoSπ thresholds at 500 Hz and 4 kHz," J. Acoust. Soc. Am. **105**, 1776–1783.

Bernstein, L. R., Trahiotis, C., Akeryod, M. A., and Hartung, K. (**2001**). "Sensitivity to brief changes of interaural time and interaural intensity," J. Acoust. Soc. Am. **109**, 1604–1615.

Bernstein, L. R., Trahiotis, C., and Hyde, E. L. (**1998**). "Inter-individual differences in binaural detection of low-frequency or high-frequency tonal signals masked by narrowband or broadband noise," J. Acoust. Soc. Am. **103**, 2069–2078.

Bos, C. E., and de Boer, E. (**1966**). "Masking and discrimination," J. Acoust. Soc. Am. **39**, 708–715.

Buss, E., Hall, J. W., III, and Grose, J. H. (**2003**). "The masking level difference for signals placed in masker envelope minima and maxima," J. Acoust. Soc. Am. **114**, 1557–1564.

Deatherage, B. H., and Evans, T. R. (**1969**). "Binaural masking: Backward, forward, and simultaneous effects," J. Acoust. Soc. Am. **46**, 362–371.

Durlach, N. I., Gabriel, K. J., Colburn, H. S., and Trahiotis, C. (**1986**). "Interaural correlation discrimination: II. Relation to binaural unmasking," J. Acoust. Soc. Am. **79**, 1548–1557.

Grantham, D. W., and Wightman, F. L. (**1979**). "Detectability of a pulsed tone in the presence of a masker with time-varying interaural correlation," J. Acoust. Soc. Am. **65**, 1509–1517.

Grantham, D. W. (**1984**). "Discrimination of dynamic interaural intensity differences," J. Acoust. Soc. Am. **76**, 71–76.

Grose, J. H., and Hall, J. W., III (**1998**). "Masker fluctuation and the masking-level difference," J. Acoust. Soc. Am. **103**, 2590–2594.

Hall, J. W., III, Buss, E., Grose, J. H., and Dev, M. B. (**2004**). "Developmental effects in temporal resolution of binaural hearing," J. Speech Lang. Hear. Res. **47**, 13–20.

Hirsh, I. J. (**1948**). "The influence of interaural phase on interaural summation and inhibition," J. Acoust. Soc. Am. **20**, 536–544.

Holube, I., Kinkel, M., and Kollmeier, B. (**1998**). "Binaural and monaural auditory filter bandwidths and time constants in probe tone detection experiments," J. Acoust. Soc. Am. **104**, 2412–2425.

Jeffress, L. A., and McFadden, D. (**1970**). "Differences of interaural phase and level in detection and lateralization," J. Acoust. Soc. Am. **4**, 1169–1179.

Koehnke, J., Colburn, H. S., and Durlach, N. I. (**1986**). "Performance in several binaural-interaction experiments," J. Acoust. Soc. Am. **79**, 1558–1562.

Kollmeier, B., and Gilkey, R. H. (**1990**). "Binaural forward and backward masking: Evidence for sluggishness in binaural detection," J. Acoust. Soc. Am. **87**, 1709–1719.

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

Lopez-Poveda, E. A., Plack, C. J., and Meddis, R. (**2003**). "Cochlear nonlinearity between 500 and 8000 Hz in listeners with normal hearing," J. Acoust. Soc. Am. **113**, 951–960.

McFadden, D., and Pasanen, E. G. (**1978**). "Binaural detection at high frequencies with time-delayed waveforms," J. Acoust. Soc. Am. **63**, 1120–1131.

van de Par, S., and Kohlrausch, A. (**1997**). "A new approach to comparing binaural masking level difference at low and high frequencies," J. Acoust. Soc. Am. **101**, 1671–1680.

Yama, M. (**1992**). "Effects of temporal separation and masker level on binaural analysis in forward masking," J. Acoust. Soc. Am. **91**, 327–335.

Zurek, P. M., and Durlach, N. I. (**1987**). "Masker-bandwidth dependence in homophasic and antiphasic tone detection," J. Acoust. Soc. Am. **81**, 459–464.

Sound localization with a preceding distractor[a)]

Norbert Kopčo, Virginia Best, and Barbara G. Shinn-Cunningham[b)]
Hearing Research Center, Boston University, Boston, Massachusetts 02215

(Received 1 March 2006; revised 18 October 2006; accepted 18 October 2006)

Experiments explored how a distractor coming from a known location influences the localization of a subsequent sound, both in a classroom and in an anechoic chamber. Listeners localized a target click preceded by a distractor click coming from a location fixed throughout a run of trials (either frontal or lateral). The stimulus onset asynchrony (SOA) between distractor and target was relatively long (25–400 ms); control trials presented the target alone. The distractor induced bias and variability in target localization responses even at the longest SOA, with the specific pattern of effects differing between the two rooms. Furthermore, the presence of the distractor caused target responses to be displaced away from the distractor location in that run, even on trials with no distractor. This contextual bias built up anew in each run, over the course of minutes. The different effects illustrate that (a) sound localization is a dynamic process that depends on both the context and on the level of reverberation in the environment, and (b) interactions between sequential sound sources occur on time scales from hundreds of milliseconds to as long as minutes. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2390677]

PACS number(s): 43.66.Pn, 43.66.Qp, 43.55.Br, 43.66.Mk [AK] Pages: 420–432

I. INTRODUCTION

Everyday settings typically contain multiple, uncorrelated sound sources coming from different locations. In order to respond appropriately to events, listeners in such settings often must estimate the locations of the sound sources. Even in environments with only a single acoustic source, this task is computationally demanding because the brain must compute source location from the acoustic signals received at the two ears. Multiple factors influence localization of single sources, from the way in which sound propagates in the environment to the way in which information is processed by the listener (see also Middlebrooks and Green, 1991). Our understanding of the factors that influence localization in scenes with *multiple* sound sources is very limited.

Previous studies of how a simultaneous masker influences target localization show that perceived target location can either be "attracted towards" or "repulsed away from" the masker location, depending on the stimulus characteristics and configuration (e.g., Butler and Naunton, 1962; Good and Gilkey, 1996; Heller and Trahiotis, 1996; Braasch and Hartung, 2002; Best *et al.*, 2005). However, two stimuli do not have to overlap in time in order to interact perceptually. For example, a stimulus that immediately precedes a target can act as a masker that interferes with the detection of the target (a phenomenon known as "forward masking," see, e.g., Kollmeier and Gilkey, 1990) or as an adaptor that introduces biases in the perceived location of the target (Thurlow and Jack, 1973; Kashino and Nishida, 1998; Duda *et al.*, 1999; Carlile *et al.*, 2001; Phillips and Hall, 2005).

In the "precedence effect," a sound arriving shortly after a preceding sound has little influence on perceived location, to the point that listeners are even poor at detecting changes in the spatial cues in the second sound (for a review, see Litovsky *et al.*, 1999). This phenomenon is often invoked to explain why listeners are able to localize sounds relatively accurately in reverberant space (Hartmann and Rakerd, 1999). The neural mechanism underlying the precedence effect is thought to suppress spatial information contained in later-arriving sounds ("lag discrimination suppression") as well as reduce the likelihood of hearing later sounds as new, discrete events ("echo suppression"). Moreover, the precedence effect builds up over time, such that the likelihood of perceiving reflections as unique events decreases with repetition (Clifton and Freyman, 1997). These effects have been characterized experimentally by presenting pairs of "lead" and "lag" stimuli in a simulated or real anechoic environment. In such conditions, echo suppression is observed for lead-lag delays of up to 10 or 20 ms for impulsive sounds (although the suppression can last up to 50–100 ms for ongoing sounds such as speech and music; Zurek, 1987). Of course, localization suppression of a second sound in a reverberant space may last longer than the suppression seen in typical precedence effect studies, as the reverberant energy from a preceding sound may help suppress localization cues in the second sound as well as reduce the saliency of the second sound onset (e.g., see Roberts *et al.*, 2004).

There are hints that spatial perception of a sound source can be affected by another sound source even when the two sources are separated considerably in time. For example, the minimum audible angle (MAA) paradigm (Mills, 1958) involves the sequential presentation of two stimuli in order to measure the smallest detectable change in source angle from one stimulus to the next (the MAA). Several studies have shown that the MAA depends on the stimulus onset asynchrony (SOA) between the first and the second stimulus for SOAs of up to 150 ms (see, e.g., Perrott and Pacheco, 1989;

[a)]Portions of this work were presented at the 2005 Mid-Winter meeting of the Association for Research in Otolaryngology.

[b)]Author to whom correspondence should be addressed. Electronic mail: shinn@bu.edu

Strybel and Fujimoto, 2000). Furthermore, two preliminary studies of localization with a preceding auditory cue (Kopčo *et al.*, 2001; Kopčo and Shinn-Cunningham, 2002) showed that the cue can have a complex effect on localization of a subsequent target for temporal separations up to hundreds of milliseconds.

Finally, several neurophysiological studies show that a preceding stimulus modulates the neural response to a target stimulus in a spatially dependent manner (Yin, 1994; Litovsky and Yin, 1998; Fitzpatrick *et al.*, 1999; Reale and Brugge, 2000). This modulation is observed for SOAs of tens of milliseconds in the inferior colliculus and hundreds of milliseconds in the auditory cortex, comparable to the time scales observed in many psychophysical studies.

The goal of the current study was to begin to characterize long-lasting spatial interactions between successive sound sources and the factors affecting these interactions. The experiments measured how the perceived lateral angle of a single-click target stimulus is influenced by an identical stimulus presented before the target from a different azimuthal location. A large range of time scales was investigated, with SOAs ranging from 25 to 400 ms. Several factors are likely to influence performance on this task, acting across different time scales (see Sec. IV A for more discussion). For example, relatively peripheral interactions are likely to contribute with a strength that decreases with increasing SOA, whereas more central factors may affect performance in ways that depend more weakly on SOA (e.g., affecting the strategy a listener employs in responding under different circumstances).

Two experiments were performed: one in a small classroom (Experiment 1) and the other in an anechoic chamber (Experiment 2). We attribute any differences in performance across the experiments to the presence of reverberant energy, either because of interference between the reverberant energy from the distractor and the direct sound energy from the target (acoustic interactions) and/or because of the effect of reverberant energy from the distractor on processing of the location information from the target (neural interactions).

II. GENERAL METHODS

A. Subjects

Seven subjects (three female and four male) participated in Experiment 1. All subjects had normal hearing as confirmed by audiometric screening, with ages ranging from 23 to 32 years. Four listeners (one female and three male) had prior experience in psychoacoustic experiments (including authors NK and VB). The four experienced listeners were the only participants in Experiment 2.

B. Stimuli and setup

Distractor and target stimuli each consisted of a single click (rectangular envelope of 2 ms duration) presented at 67 dB sound pressure level (SPL) (A-weighted maximum rms value in a 2 ms running window at the location of the listener's head). The measured spectrum in third-octave bands was flat from 2 to 4 kHz, with a 6 dB per octave roll-off outside the band. The stimuli were generated by a

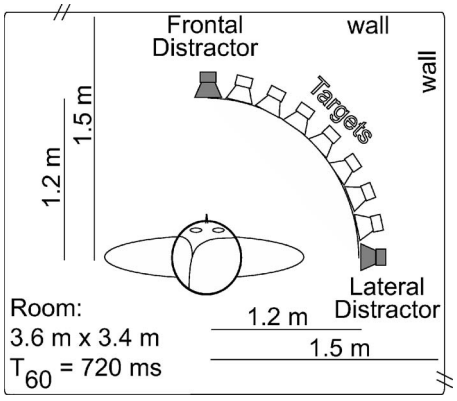


FIG. 1. Diagram of the orientation and location of the listener and the loudspeakers in the classroom used in Experiment 1. The same setup was used in the anechoic chamber in Experiment 2 although the (acoustically transparent) chamber walls were at different distances. In the figure the listener is facing the left-most loudspeaker and the targets are on his/her right. In half of runs, the listener was oriented to face the right-most loudspeaker and the targets were on his/her left.

PC-controlled Tucker Davis Technology System 3, amplified by a Crown D-75A amplifier, and played by one of nine matched Bose Acoustimass cube loudspeakers selected by a serial-port-controlled eight-relay output module (KITSRUS K108). Analysis of recordings of the stimuli showed differences between loudspeakers of no more than 5 dB in any third-octave band over 700–16,000 Hz. The stimulus onset asynchrony (SOA) between the distractor and the target click was set to 25, 50, 100, 200, or 400 ms.

A Polhemus FastTrak electromagnetic tracker was used to measure the location of the listener's head, the approximate location of the loudspeakers, and the listener's responses. The listeners indicated the perceived direction of the target source by pointing a stick with the Polhemus electromagnetic sensor attached at its end and pressing an attached button. Note that the listeners were allowed to point in any direction, including outside the range of target directions. Moreover, they were specifically instructed to point in the perceived direction of the sound. However, given the subjects' familiarity with the experimental setup, they may have limited their responses to fall within the actual speaker range even if they perceived the target as outside of this range.

The nine loudspeakers were equally spaced along a quarter circle of diameter 1.2 m with the listener at the center (Fig. 1). The loudspeakers were fixed on stands 1.5 m above the floor, approximately at the level of the listener's ears. The listener was seated on a chair that could be rotated so that they faced the left-most or the right-most loudspeaker, with the loudspeaker array either in their right or left frontal quadrant (respectively). The left-most and right-most loudspeakers were used only to present the distractor stimuli. The remaining seven loudspeakers were used to present target stimuli. An additional loudspeaker directly behind the listener played instructions to the listener during the experiment.

C. Listening environments

Experiment 1 was conducted in an empty, quiet rectangular classroom measuring 3.4 m × 3.6 m × 2.9 m (h), with a

small entrance space measuring 1.5 m × 1.6 m × 2.9 m. The room was carpeted, with hard walls and acoustic tiles covering the ceiling. The reverberation times in octave bands centered at 500, 1000, 2000, and 4000 Hz were 613, 508, 512, and 478 ms, respectively. The background acoustic noise was at approximately 39 dB SPL (A-weighted). As illustrated in Fig. 1, the loudspeaker array was set up in a corner of the room with the left-most and the right-most loudspeakers 30 cm from and facing away from the two nearest walls. The listener was seated 1.5 m from these walls.

Experiment 2 was conducted in an anechoic chamber at the Dept. of Psychology of the University of Massachusetts, Amherst. The chamber measures 4.9 m × 4.1 m × 3.12 m and its walls, floor, and ceiling are lined with 0.72 m foam wedges. The subjects were seated near the center of the chamber as illustrated in Fig. 1; the only difference in setup from Experiment 1 was that there were no reflective walls.

D. Experimental procedure

Each experiment consisted of four 30 min blocks, separated by breaks. Within each block, the listener performed four runs, one for each combination of listener orientation (facing the left- most or the right-most loudspeaker) and distractor location (from the left-most or the right-most loudspeaker). The order of the runs within each block was random, and differed from subject to subject. Each run contained 168 trials [seven (target loudspeaker locations) × six (five SOAs + no distractor) × four (repeats)]. Within each run, one repeat of each of the six conditions (five SOAs and the no-distractor control) was presented in random order before any condition was repeated, so that each run logically could be broken down into four subruns.

At the beginning of each run the subject was instructed to rotate the chair to face the predetermined loudspeaker, sit on the chair, and put his/her head on the headrest. After calibration measurements (see below), the listener was instructed to close his/her eyes and remain still for the remainder of the run. The listener was told which loudspeaker would present the distractor in the run, and a sample stimulus (a pair of clicks, one from the distractor loudspeaker and one from one of the randomly selected target loudspeakers) was presented before the experimental trials began.

A single trial consisted of a presentation of one stimulus, followed by the listener's response, after which there was a constant delay (approximately 0.5 s) before the stimulus for the next trial was presented. With this inter-trial delay, the subject had no difficulty in reorienting from indicating the previous location to preparing for the next stimulus. There was no limit on how fast the subject had to respond, so the pace of the experiment was controlled by the subject. On average, a trial took 2–3 s and a run took 5–6 min.

Each stimulus contained one target click presented from a randomly chosen target loudspeaker. In a majority of the trials (five out of six, i.e., 83%), a distractor click was also presented before the target; on the rest of the trials there was no distractor, but the target click was preceded by 400 ms of silence in addition to the standard inter-trial pause (this additional delay came about because the no-distractor stimulus was derived from the 400 ms SOA distractor stimulus by zeroing the distractor channel). The listener did not know a priori whether or not a given trial would contain a distractor, but could always tell whether one or two sources had been presented. In addition, although listeners were not instructed about the timing of the stimuli explicitly, they rapidly learned that no-distractor trials began with a 400 ms silence, which cued them to expect a target-only trial. After each stimulus, the listener pointed in the perceived direction of the target and pressed a button that caused the response to be recorded and the next trial to be initiated.

In Experiment 1 the subjects performed 1–2 blocks per day. In Experiment 2 the subjects performed the whole experiment in one day. In both experiments, the blocks were interleaved with blocks of another similar study; however, in each case there was at least 1 h of rest between consecutive blocks.

At the beginning and end of every run, subjects calibrated the electromagnetic tracker's coordinate system. With his/her eyes open, the subject was asked to point the electromagnetic pointer at the center of the two distractor loudspeakers and at the middle target loudspeaker. He/she was then asked to establish the location of the head by pointing at (in this order) his/her left ear, right ear, and nose. Because subjects actively and frequently performed this calibration, they were highly familiar with the layout of the loudspeakers.

At the beginning of Experiment 1, the experimental procedure was described to the listeners. In particular, the listeners were instructed to try to ignore the distractor, that the distractor and no-distractor trials would be interleaved within a run, and that they should point in the *perceived* direction of the target, regardless of what they knew about the experimental setup and the possible speaker locations. In addition, the purpose of the calibration measurements was explained and it was stressed that the listeners should not move their heads or open their eyes during a run (i.e., between the initial and final calibration measurement). After receiving these instructions, the listeners performed a brief practice session consisting of at least two runs.

E. Data analysis

All subject responses, recorded by the tracker in the form of Cartesian coordinates, were first projected into the plane defined by the recorded location of the three loudspeakers and the center of the listener's head. These recorded locations were determined by averaging the calibration measurements taken at the beginning and the end of each run. The lateral angle between the response direction and straight ahead (with respect to the center of the head) was calculated and stored as the response angle. Analysis showed negligible left-right differences, so the data measured with the subjects facing the right-most loudspeaker were mirror flipped and combined with the data measured when the subjects faced the left-most loudspeaker. This reduced the four spatial configurations to two: one with a frontal distractor and one with a lateral distractor. For each subject, 32 responses were col-

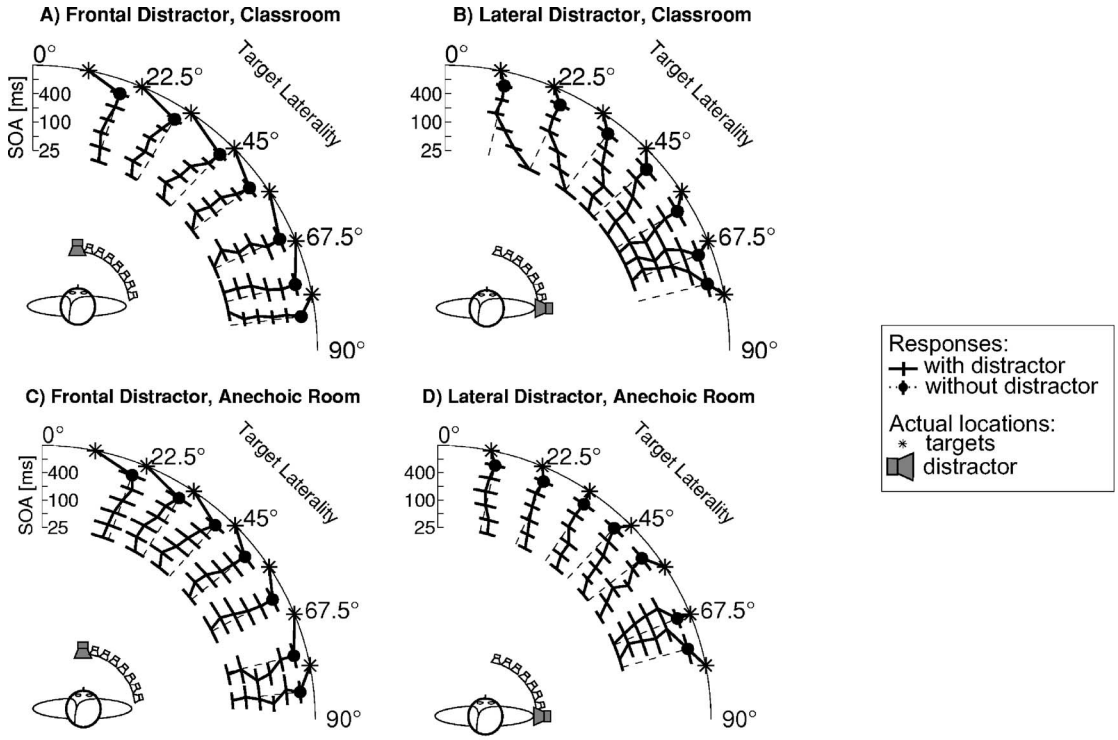


FIG. 2. Mean localization responses in the classroom (A, B) and the anechoic room (C, D). Each panel shows the across-subject mean and standard error in perceived target lateral angle as a function of actual target lateral angle for different SOAs, as well as in the no-distractor condition. A, C) Frontal distractor. B, D) Lateral distractor.

lected in total for each combination of configuration (frontal vs lateral distractor), SOA (five SOAs + no distractor), and target lateral angle (seven lateral angles).

III. RESULTS

A. Experiment 1: Classroom

The top two panels of Fig. 2 show the across-subject mean and standard error in the perceived target location as a function of the actual source lateral angle for frontal (panel A) and lateral (panel B) distractors (the bottom two panels show results for Experiment 2, discussed in Sec. III B). The asterisks in both panels show the actual lateral angles of the target loudspeakers. The radial solid lines connected to each asterisk (target location) show the average response angles for the different stimulus conditions. Each crossing of a radial line with a line segment shows a mean response angle, with the length of the segment showing the standard error in the mean response for a given target lateral angle and SOA. The outermost data points (filled circles) represent the control-trial responses with no distractor, followed by the responses obtained with an SOA of 400 ms. Further decreases in the radial distance correspond to gradually smaller SOAs, with the 25 ms SOA shown by the innermost ring of data. A dashed radial line starting at the no-distractor response location (filled circles) is shown to allow the effect of the distractor on the perceived target lateral angle to be easily assessed.

1. Contextual bias

One large effect evident in Fig. 2(A) is that for blocks involving the frontal distractor, localization responses are biased towards the side, shown by a clockwise displacement of all judgments (even on trials where the distractor is not present; compare asterisk locations to all mean response angles). For blocks involving the lateral distractor, the most lateral targets are biased towards the midline, while more frontal targets show a slight opposite bias, towards the side (Fig. 2(B); compare asterisks and filled circles). Overall, a consistent response bias is evident: responses are more lateral for the frontal distractor than for the lateral distractor. These biases are caused by the ensemble of trials presented in a given experimental run, not by the immediately preceding distractor, because they occur for all trials, including the no-distractor control trials. Note that there were no control runs made up entirely of control trials, so we cannot directly determine exactly what response bias is caused by the distractor presented in a particular run. However, we can directly compare the bias observed for trials in frontal- and lateral-distractor runs to assess the change in bias caused by changing the distractor location from frontal to lateral in different runs.

To quantify this context effect, we computed the difference in responses for the same target location and SOA for both frontal- and lateral-distractor runs. For each subject and target direction, the mean response in the lateral-distractor runs was subtracted from the mean response for the frontal-distractor runs. These results are shown in Fig. 3, averaged

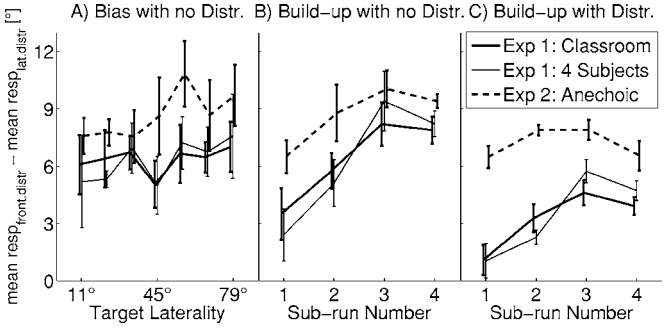


FIG. 3. Contextual effects observed in the no-distractor trials (A, B) and in the distractor trials (C). Performance for all seven subjects from Experiment 1 is shown by thick solid lines, for a subset of four subjects by thin solid lines, and for Experiment 2 by dashed lines. The statistic plotted in each panel is the across-subject mean and within-subject standard error[1] in the: A) difference between the perceived target location in the context of frontal vs lateral distractor as a function of target lateral angle; B) across-lateral-angle average difference in responses as a function of subrun within experimental run; C) across-lateral angle and across-SOA average difference in responses in trials with distractor as a function of the subrun number within experimental run.

across subjects. Panel A shows the overall magnitude of the change in response bias for the no-distractor control trials as a function of target laterality. Panel B shows the magnitude of the effect for the no-distractor trials as a function of time within a run, collapsed across target location. Panel C shows the magnitude of the effect for the distractor trials, collapsed across SOA and target location, as a function of time within a run. The solid thick lines in Fig. 3 plot the across-subject mean effect for the seven subjects who completed Experiment 1. The solid thin lines show the results in Experiment 1 averaged over the subset of four subjects who performed both Experiment 1 and Experiment 2. The dashed line shows the same results (for the same four common subjects) for Experiment 2 (these results are discussed below, in Sec. III B). Error bars in each panel show the within-subject standard error of the across-subject mean.[1]

In general, Fig. 3 shows that the four subjects who performed both Experiment 1 and Experiment 2 show the same pattern of results in Experiment 1 as the larger subject population (compare thick and thin lines in each panel).

Figure 3(A) quantifies the influence of the distractor on the no-distractor trials as the difference in the mean response on no-distractor trials for frontal- and lateral-distractor runs. For all target angles this difference is positive (thick and thin solid lines in Fig. 3(A)). In other words, distractors cause response bias within a run such that the no-distractor control trial responses are relatively closer to the median plane when the distractor is to the side and relatively farther from the median plane when the distractor is in front. Computed as the difference between the response bias in the lateral- and frontal-distractor runs, the contextual bias is roughly independent of target laterality (i.e., the solid lines in Fig. 3(A) are relatively flat).

Because the experimental runs with the frontal and lateral distractors were interleaved, any contextual shift caused by the distractors had to develop anew in each experimental run. Each run consisted of 168 trials: four repeats of each combination of target lateral angle and SOA condition. Trials within a run were ordered in subruns, such that all combinations of target angle and SOA were presented once before any were repeated. Thus, in each run, exactly four full sets of no-distractor responses (one for each of the seven target locations) were measured, one in each subrun. Because there was no large effect of target lateral angle on the contextual bias in the no-distractor responses (see Fig. 3(A)), data were combined across the target angle for each subject to estimate the contextual effect as a function of subrun.

Figure 3(B) shows the across-subject mean in the contextual difference as a function of subrun. In general, the contextual effect increased with subrun. Averaged across all seven subjects, the contextual effect grew from roughly 4 to 8° across the four subruns (solid thick line; one-way repeated measures analysis of variance (ANOVA): $F_{3,18}=7.64, p<0.005$). The buildup was also significant when considering only the four subjects who also completed Experiment 2 (solid thin line; one-way repeated-measures ANOVA: $F_{3,9}=17.52, p<0.0005$). Given that one experimental run took approximately 5 min, this result shows that the context effect built up over the course of minutes, orders of magnitude longer than the millisecond time scale of primary interest in this study.

Figure 3(C) plots the contextual bias within each subrun for the distractor trials (collapsed across SOA and target angle) to see if this buildup was general. Although the average contextual bias for the distractor trials is smaller than for the no-distractor trials, the contextual bias builds up over subruns in a way that is similar to that seen in the no-distractor trials (compare the solid thick and thin lines in Figs. 3(B) and 3(C); the difference in the change in contextual bias over time is roughly 2° or less across all subruns). The increase in contextual bias with subrun for the distractor trials is significant both for the full set of seven subjects (solid thick line; one-way repeated-measures ANOVA: $F_{3,18}=4.05, p<0.05$) and for the subset of four subjects who also performed Experiment 2 (solid thin line; one-way repeated-measures ANOVA: $F_{3,9}=9.95, p<0.005$).

Taken together, these results show that there is an unexpected effect of the distractor on localization of the target that builds up over time for both control and distractor trials. This suggests that, in the absence of any measurements of localization in blocks without any distractors, the most appropriate controls for judging the effect of an immediate distractor are the responses to the control trials within a block.

2. Effect of distractor on mean responses

Figures 2(A) and 2(B) show that the distractor causes different biases at different SOAs. However, inter-subject differences are large, as indicated by the standard errors in the means. Some of this inter-subject variability may be due to individual differences in the average response biases present across all trials in a run (including the no-distractor control trials). Therefore, differences between the responses for targets preceded by distractors and responses in the no-distractor control trials were computed for each subject individually (for each run). Having established that both control and distractor trials are similarly affected by the built-up

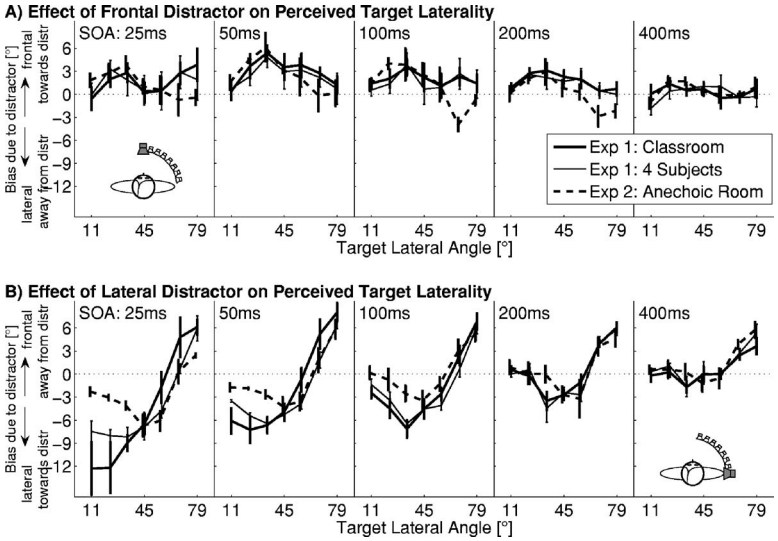


FIG. 4. Effect of the distractor on the perceived target lateral angle in the classroom and the anechoic room. Plotted is the across-subject mean and standard error in the difference between the perceived lateral angle with the distractor vs without the distractor. Each panel shows results for a different SOA. A) Frontal distractor. B) Lateral distractor.

contextual bias, these differences show the effects of the immediately preceding distractor on performance.

Figure 4 shows the within-subject change in localization response due to the presence of the distractor, averaged across subjects (error bars show the standard error in this within-subject difference). Positive values in Fig. 4 correspond to bias towards midline; negative values represent a bias to the side. The solid thick lines in Fig. 4 show results for the full set of seven subjects in Experiment 1. The solid thin lines show results for the subset of subjects who completed both experiments. Finally, the dashed line shows results for the same four subjects in Experiment 2 (discussed in Sec. III B).

As seen by the smaller size of the standard errors in Fig. 4 compared to Fig. 2, the effects of the distractor are much more consistent across subjects when computed relative to each subject's individual no-distractor response. In general, results for the full set of seven subjects who performed Experiment 1 are very similar to results for the subset of four subjects common to both experiments (compare solid thick and thin lines in Fig. 4), although the magnitude of some of the biases is smaller for the subset of four subjects (e.g., the thin solid lines tend to fall above the thick solid lines in the left edges of the left-most panel of Fig. 4(B)).

Several effects of the distractor on target localization can be observed in the reverberant classroom:

- For frontal targets, the lateral distractor causes a bias towards the side, an effect that decreases with increasing SOA. For instance, for an SOA of 25 ms (left-most panel of Fig. 4(B)), the left edge of the solid line shows a bias that is 12° on average; however, this effect is not present for an SOA of 400 ms (the right-most panel in Fig. 4(B)).
- Both the frontal and the lateral distractors cause targets located at intermediate source angles (near 45°) to be localized closer to the distractor. The average size of this effect is as large as 6° and decreases with increasing SOA. For instance, in the solid lines in Fig. 4(A), response bias is positive for intermediate target lateral angles at SOAs of 25 and 50 ms, but is negligible for an SOA of 400 ms. Similarly, in Fig. 4(B) there is a negative dip in the response bias for intermediate lateral angles that decreases with increasing SOA.
- The lateral distractor causes a bias in the localization of nearby targets towards the midline, an effect that is independent of the SOA. In Fig. 4(B), the right edge of each solid line is positive, falling at around 5°.

3. Standard deviation in responses

For each subject, the standard deviation in the responses was computed for each combination of target lateral angle, SOA (including the no-distractor condition), and distractor location. These standard deviations were subjected to a three-way repeated-measures ANOVA with factors of SOA, distractor angle, and target angle. For the full set of seven subjects, this analysis found a significant interaction between SOA and target location ($F_{30,180}=1.63, p<0.05$) as well as a significant main effect of SOA ($F_{5,20}=26.16, p<0.0001$). For the subset of four subjects who completed both experiments, the same ANOVA found a significant three-way interaction between SOA, distractor location, and target location ($F_{30,90}=1.59, p<0.05$) and a significant main effect of SOA ($F_{5,15}=18.40, p<0.0001$). All other interactions and main effects were not significant. This analysis suggests that response variability changes with SOA in a manner that depends upon distractor location and target location.

Figure 5 shows the across-subject means in the standard deviations as a function of target laterality for two SOAs (the shortest SOA of 25 ms and an intermediate SOA of 100 ms, shown by triangles and squares, respectively) and for the no distractor condition (shown by circles). Data for the other SOAs are left out to improve the figure legibility, but followed the trends illustrated by the data included in the figure. The error bars in Fig. 5 represent the within-subject standard errors of the means.[1] Panels A and B show the data from Experiment 1 for the frontal and lateral distractors, respectively. Data for all seven subjects are shown in the upper portion of panels A and B; data for the subset of four subjects

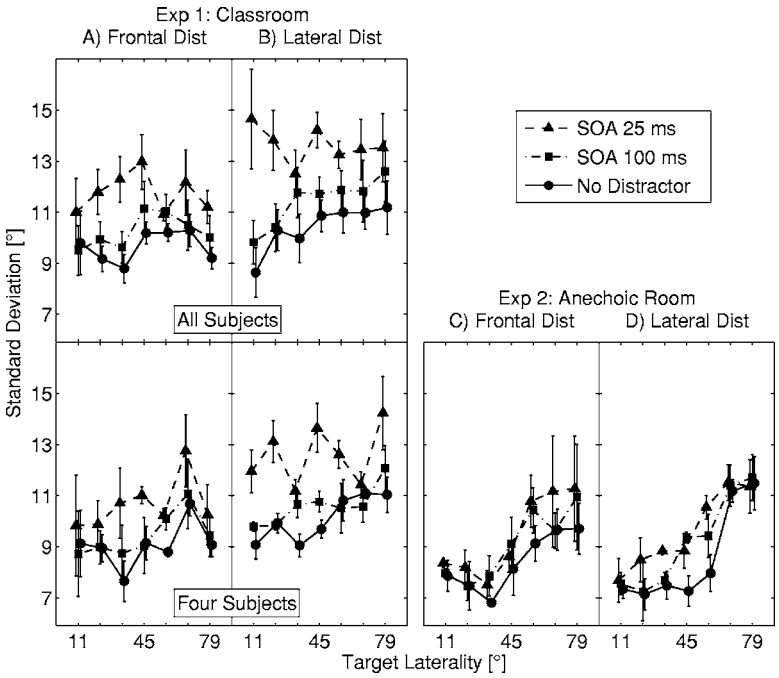


FIG. 5. Standard deviations in listeners' responses in the classroom (Exp 1, panels A and B) and the anechoic room (Exp 2; panels C and D) as a function of the target laterality for two example SOAs and the no-distractor condition (results for other SOAs follow the same trends, and are left off for visual clarity). The across-subject means in the standard deviations are shown (error bars show within-subject standard error of the mean). A, C) Frontal distractor. B, D) Lateral distractor. Upper portion of panels A and B: all seven subjects. Lower portion: four subjects who participated in both experiments.

who completed both experiments are in the lower portion. Panels C and D show the data from Experiment 2, discussed in Sec. III B.

Overall, the average variability in subject responses was larger for the full set of seven subjects than the subset of four subjects who completed both experiments (data in upper portion of panels A and B are above the corresponding data in the lower portion). However, all other patterns are similar for the two subject groups. In Experiment 1 (panels A and B), variability is larger with a lateral distractor than with a frontal distractor (data in Fig. 5(A) fall below the corresponding data in Fig. 5(B)). For both distractor positions, response variability is greatest at the shortest SOA and smallest when there is no distractor present (within Figs. 5(A) and 5(B), triangles fall above squares, which are above circles). The effect of SOA on response variability tends to be larger for the lateral distractor than for the frontal distractor (compare Figs. 5(A) and 5(B)). This difference did not reach statistical significance with the full set of subjects (i.e., neither the main effect of distractor location nor any of the interactions including distractor location reached significance), but the significant three-way interaction for the subset of four subjects supports this conclusion. For a frontal distractor, variability is essentially independent of target laterality (in Fig. 5(A), results are essentially flat). For a lateral distractor, variability is large and does not show any clear trend as a function of target laterality at the shortest SOA (triangles in Fig. 5(B)); however, at longer SOAs and in the absence of a distractor, variability tends to increase with increasing source laterality.

B. Experiment 2: Anechoic environment

The bottom two panels of Fig. 2 show the results of Experiment 2 in a format identical to the top two panels. As in Experiment 1, when the distractor is in front there is a consistent clockwise displacement of all the judgments relative to the true loudspeaker angle, when the distractor is in front (Fig. 2(C)). The lateral distractor causes a weaker, opposite bias relative to the true target location (Fig. 2(D)), but, as in Experiment 1, this effect is evident only for the target angles near the distractor.

1. Contextual bias

To quantify the context effect, we computed the difference in mean responses on no-distractor control trials between frontal- and lateral-distractor runs as in Experiment 1. Mean differences are shown in Fig. 3(A) (dashed lines; error bars show the within-subject standard error of the across-subject mean[1]). As in the classroom, this difference is positive at all target lateralities and roughly independent of target laterality. The context effect tended to be slightly greater in anechoic space than in the classroom (in Fig. 3(A), the dashed line is above the solid and the thin lines), although this difference did not reach statistical significance.

The dashed line in Fig. 3(B) shows the across-subject mean in the contextual difference as a function of the subrun in Experiment 2. As in Experiment 1, the contextual bias built up over the four subruns, growing from roughly 6 to 9° (one-way ANOVA with subrun as the factor; $F_{3,9}=5.15$, $p<0.05$).

Distractor trials were analyzed to see if the contextual bias is general (dashed lines in Fig. 3(C)). As a function of subrun, the contextual bias for the distractor trials in the anechoic room was relatively flat (consistent with this, there was no statistically significant effect of subrun on the contextual bias; $F_{3,9}=1.59, p=0.26$). Despite this, the average contextual bias observed across all SOAs and target angles is nearly as large for distractor trials as it is for the no-distractor trials in Experiment 2 (a two-way repeated-measures ANOVA with factors of subrun and distractor presence/absence found no significant interaction; $F_{3,9}=3.05$, $p=0.08$). This suggests that, as in Experiment 1, localization

of the no-distractor targets within the runs provides a control for assessing the effect of the immediately preceding distractor on target localization.

2. Effect of distractor on mean responses

As in Experiment 1, data were individually normalized by the no-distractor responses and re-plotted (dashed lines in Fig. 4).

Comparing anechoic and classroom results shows:

- The large lateral bias of frontal targets caused by the lateral distractor in the classroom was much smaller in the anechoic room (compare the left-most edges of the dashed lines to the solid thick and thin lines in Fig. 4(B)).
- Consistent with the classroom data, both the frontal (Fig. 4(A)) and the lateral (Fig. 4(B)) distractors biased localization of targets from intermediate lateral angles towards the location of the distractor. This effect is similar in magnitude in anechoic space and in the classroom and decreases with increasing SOA. For instance, the dashed-line peaks in the plots in Fig. 4(A) and the dashed-line troughs in Fig. 4(B) are comparable to the solid thick- and thin-line results for targets near 45°.
- As in the classroom, the lateral distractor induced a bias of nearby targets of about 5° towards the midline, independent of SOA. Specifically, both dashed and solid lines in all the panels of Fig. 4(B) are positive at their right edges (i.e., for the target at 79°).

3. Standard deviation in responses

For Experiment 2, standard deviations were computed in the same way as in Experiment 1. Panels C and D in Fig. 5 show the across-subject means in the standard deviations as a function of the target laterality in a format similar to panels A and B. Overall, response variability in anechoic space is smaller than in the classroom (results in panel C fall below those in panel A, and those in panel D fall below those in panel B). In anechoic space, unlike in the classroom, response variability is essentially the same for frontal and lateral distractors (compare results in Figs. 5(C) and 5(D)). Similar to what was seen in the classroom, response variability tends to decrease with increasing SOA (triangles tend to be highest and circles tend to be lowest within each panel), although the effect of SOA is smaller than in the classroom. In Experiment 2, there is a monotonic increase in response variability with increasing target laterality, an effect that is independent of distractor location or SOA (all lines increase from left to right in Figs. 5(C) and 5(D)). Supporting these conclusions, a three-way repeated-measures ANOVA with factors of SOA, target laterality, and distractor location found statistically significant main effects of SOA ($F_{5,15}=4.13$, $p<0.05$) and target laterality ($F_{6,18}=7.46, p<0.001$), but no other significant effects or interactions.

IV. DISCUSSION

A. Stages of auditory spatial processing

In trying to understand the effects of a preceding distractor on localization, it is useful to first outline various stages of auditory processing at which the presence of the distractor could affect the spatial representation of the target.

For sounds that overlap in time, adding before they enter the ear, simple acoustic interactions between distractor and target stimuli may impact the ability to compute target location. Similarly, even if they do not overlap acoustically, when two sounds occur sufficiently close in time (e.g., within 1 or 2 ms of one another), the responses they elicit in the cochlea will overlap and interfere with one another, directly affecting what target spatial information can be extracted (Tollin, 1998; Hartung and Trahiotis, 2001).

Over slightly longer time scales, there may be temporal interactions (e.g., interactions between excitation and inhibition) within the neural structures that extract spatial parameters from the signals at the two ears, such as the brainstem structures that are known to be sensitive to interaural time and level differences. Such interactions (e.g., at the level of the inferior colliculus) are hypothesized to underlie the precedence effect; for instance, a strong onset may cause a subsequent suppression of spatial cues (Yin, 1994; Litovsky and Yin, 1998; Fitzpatrick *et al.*, 1999; Reale and Brugge, 2000; Litovsky and Delgutte, 2002; Dizon and Colburn, 2006). Such mechanisms undoubtedly influence localization in situations involving sequences of stimuli such as those in the present study, especially in the classroom, where reverberant energy persists through the nominal time delay between the distractor and the target stimulus.

Although there is some debate about the nature of the neural representation of exocentric space formed by integrating the different spatial cues (e.g., Carlile *et al.*, 2001; Stecker *et al.*, 2005) it has often been assumed that dynamic interactions within such a representation can explain spatial interactions between stimuli (Thurlow and Jack, 1973; Kashino and Nishida, 1998; Duda *et al.*, 1999; Carlile *et al.*, 2001; Phillips and Hall, 2005). In particular, adaptation after effects are assumed to arise because a repeated stimulus causes long-lasting changes in neural responses that cause a perceptual warping of auditory space (Kashino and Nishida, 1998; Carlile *et al.*, 2001).

Finally, a preceding distractor may influence sound localization at a cognitive level. For example, a distractor may alter the distribution of spatial attention, causing changes in spatial perception, and/or influence the strategy adopted by a listener when judging the location of the target, causing changes in response patterns.

Although the various effects observed in the current experiment cannot be attributed to any particular processing stage with certainty, below we consider the processing levels most likely to be involved in the various phenomena we observed.

B. Standard deviation effects

Overall, response variability was larger in the classroom than in anechoic space, especially for lateral distractors and short SOAs (Fig. 5). It is possible that this effect was simply related to the presence of direct and reverberant energy from

the distractor and/or the target. For example, this energy may have added to the acoustic variability in the spatial cues for target location.

Binaural recordings from a KEMAR manikin in the classroom were taken using the stimuli presented to the listeners (see Appendix for details). These recordings confirmed that at short SOAs, reverberant energy from the distractor was present at the onset of the target. Acoustic analysis was performed to examine whether the effect this overlap has on the target spatial cues is consistent with the observed behavior. The analysis showed that the height of the running interaural cross-correlation peak decreases with decreasing SOA (see also Shinn-Cunningham *et al.*, 2005). Such a reduction in interaural correlation is associated with an increase in the width or blurring of the perceived location of a source (Blauert, 1997; de Vries *et al.*, 2001). Similarly, the reverberant energy of the distractor also tended to reduce the magnitude of interaural level differences (ILDs) at the onset of the target, causing the ILDs to be somewhat inconsistent with the interaural time differences (ITD) cues.

Both of these effects (a reduction in interaural correlation strength and alteration of the mean ILDs in the stimuli) may have caused the target location to be more diffuse and more difficult to localize. Thus, acoustic interactions between reflected energy from the distractor and direct energy of the target may at least partially account for the increase in response variability for sources in the classroom compared to in anechoic space. In addition, there was some, albeit weak, background noise present in the classroom that was not present in the anechoic chamber. Given that the response variability to the control trials is also greater in the classroom than in the anechoic setting, this noise may have further reduced the reliability of the spatial cues in the room and increased the response variability.

A precedence-effect-like suppression of spatial information in sounds following an abrupt onset or preceding sound may also contribute to the larger response variability in the classroom compared to in anechoic space. Although most precedence studies talk about mechanisms that operate only up to about 10 ms, suppression lasts longer for ongoing stimuli such as speech or music. Indeed, reflections from the distractor may directly suppress later-arriving sound location cues (e.g., in the onset of the target sound), or may effectively extend the suppression initiated by the distractor onset (see also Roberts *et al.*, 2004). All of these observations suggest that at least some portion of the greater response variability observed in the classroom compared to in anechoic space may be caused by a neural mechanism invoked by ongoing acoustic energy occurring just before the onset of the target (in this case, the reflected energy of the distractor).

It is worth noting, however, that neither the acoustic analysis nor consideration of precedence-like suppression can explain certain details of the response variability data. For example, neither explanation can account for why, in the classroom, response variability is larger for lateral distractors than for frontal distractors.

C. Biases

The most prominent effect of the distractor in these experiments was that in the classroom, the lateral distractor biased the perceived location of frontal targets towards the side at short SOAs. Given that this effect is much weaker in anechoic space, one parsimonious explanation for this bias is that the lateral distractor and its reflections interacted acoustically with the target. However, analysis of the spatial cues in the total signal reaching the listener at the target onset (see Appendix) did not reveal any biases in the interaural parameters that would predict the observed shift in perceived lateral position towards more lateral positions. In particular, there is no consistent bias in the ITD value corresponding to the peak in the interaural cross-correlation function at the target onset. While ILD cues do show a systematic bias, this bias is in the wrong direction to explain the observed behavioral bias. The magnitude of the ILDs in the acoustic signals is reduced by reverberant energy, an effect that, if anything, should result in perceived target positions being biased towards, not away from, the median plane.

It may be that the localization bias of frontal targets by a lateral distractor is related to the variability in the subject responses, discussed above. Specifically, the listeners might be biased to respond towards the middle of the response range whenever they are uncertain about the target location. Indeed, response variability is greatest for lateral distractors in the classroom, particularly when the SOA is short (e.g., see triangles plotted in the left edge of Fig. 5(B), compared to all other results in Fig. 5), the very conditions that produce the greatest response bias (Fig. 4(B), left edges of the first and second panels). This correspondence suggests that listeners are simply uncertain about the target location when a lateral distractor precedes targets near the median plane in the classroom, causing them to guess the target location, which, in turn, causes a bias towards the side. Because this effect depends strongly on SOA (disappearing at long SOAs), it is likely caused by a reduction in sensitivity to target spatial information due to distractor energy, either due to added noise in the representation of target spatial information (e.g., from acoustic interference) or a reduced response to the target cues (e.g., neural suppression of target spatial information, as in the precedence effect).

In both experiments, perceived location of targets located near 45° is biased towards the distractor location, whether the distractor is frontal or lateral (peaks in Fig. 4(A) and dips in Fig. 4(B)). This effect decreases with increasing SOA (i.e., decreases when going from left to right in Figs. 4(A) and 4(B)), which suggests that it is also caused by relatively low-level effects related to the dynamics in the representation of auditory spatial cues.

Biases in spatial representation have been observed previously in behavioral experiments, both in the face of adaptation to a repeated preceding sound (Kashino and Nishida, 1998; Carlile *et al.*, 2001) and in situations involving concurrent stimuli (Best *et al.*, 2005). However, these previous studies generally found that perceived target locations were biased away from rather than towards the distractor location. It is possible that since the distractor and target are identical

in the current experiments, they are more likely to be perceived as coming from the same sound source—a source whose perceived location depends on integrating spatial information from both the target and the distractor.[2] Further experiments should test whether stimuli from different directions are more likely to be perceived at some intermediate, integrated location when they are perceptually similar.

Another factor that may play a role in the localization bias of sources at intermediate locations is exogenous allocation of spatial attention, or "orienting" (Spence and Driver, 1994). In particular, the distractor may cause an involuntary shift of spatial attention toward its location, causing subsequent targets from other regions of space to be localized less reliably. Such an explanation is also supported by the observation that the biases were reduced at larger SOAs, where there is sufficient time to disengage attention from the distractor and reorient attention toward the target.

The lateral distractor causes a strong medial bias in responses to nearby targets (shown by the peaks for the far right target positions in the graphs and panels of Fig. 4(B)). However, in contrast with all of the effects discussed above, this effect is essentially independent of SOA for the values used in this study and occurs in both the classroom and the anechoic room. Such a lack of dependence on the details of the stimulus attributes (e.g., SOA, the level of reverberation in the environment) suggests that the nature of this bias is very different from the biases already considered. In particular, such an effect is unlikely to be caused by low-level interactions, which should become weaker with increasing SOA and should be affected by the level of reverberant energy in the environment. Instead, it is likely that this effect is caused by a more central mechanism, such as a change in the response strategy employed by the listener. For instance, for lateral targets (where localization is relatively poor), subjects may use the distractor as a perceptual "anchor," judging target location relative to the distractor location rather than in absolute coordinates (Hartmann and Rakerd, 1989; Litovsky and Macmillan, 1994; Recanzone *et al.*, 1998). Such a strategy may have caused subjects to overestimate the separation between the target and distractor, resulting in the observed bias. Although there was no corresponding bias for frontal targets in the frontal-distractor runs, employing relative judgments may not be an efficient strategy for localizing sources near the medial plane, where spatial auditory resolution can be an order of magnitude better than for sources to the side (Mills, 1958). The idea that subjects may alter their response strategy based on specific knowledge about the possible stimulus locations has been explored in several previous experiments, and substantial effects on localization response patterns were observed (Perrett and Noble, 1995).

Another important finding in this study was that the listeners' responses changed during the course of the experimental runs in a way that depended on the location of the distractor presented in that run. The resulting contextual bias caused responses to consistently be displaced away from the distractor location. To our knowledge, this is the first example of short-term (on the scale of minutes) spatial auditory plasticity induced without either visual feedback (Jack and Thurlow, 1973; Warren *et al.*, 1981; Recanzone, 1998) or a near-continuous exposure to a constant (adapting) auditory stimulus (Thurlow and Jack, 1973; Kashino and Nishida, 1998; Carlile *et al.*, 2001). This effect may be either bottom up, driven purely by the statistical distribution of the stimuli heard within a run (45% of which were coming from the distractor location), or top down, driven by the listener's knowledge of the distractor location and attempts to direct spatial attention away from it.[3]

Finally, it is important to note that although several potential accounts for the localization biases observed have been explored above, there may be other possibilities. Importantly, however, given the large differences in how the various effects varied with attributes such as SOA and environment, it is clear that no one mechanism can explain all of the effects observed.

V. SUMMARY AND CONCLUSIONS

Localization of a target click was affected by the presence of a preceding click in a number of different ways, over SOAs much larger than are typically thought to cause interstimulus interactions. Given the complex pattern of observed effects, several forms of distractor-target interaction, operating at different stages in the spatial auditory processing pathway, likely contribute to how a target is localized in runs containing distractors:

- In the reverberant classroom (but not in anechoic space), response variability and response bias are larger when sources are near the median plane and preceded by a lateral distractor (as opposed to the frontal distractor). This effect is particularly strong at short SOAs (up to 100 ms). The observed reduction in localization accuracy and reliability suggests that, under these conditions, listeners cannot access spatial information in the target. This effect is likely caused by some combination of acoustic interference between distractor and target and ongoing neural suppression of later-arriving target location information by the preceding distractor and its subsequent reflections.
- Perceived locations of targets at intermediate angles are biased towards the distractor in both environments, an effect that disappears with increasing SOA. This effect appears to be the result of interactions in the neural representation of space or of exogenous orienting caused by the distractor.
- The lateral distractor repulses nearby sources in both environments, independent of SOA. This effect is likely more central, e.g., from listeners adopting a "relative" rather than "absolute" localization strategy.
- A contextual bias occurs both in trials with and without a distractor, causing a shift in the perceived locations of the targets in the frontal-distractor runs compared to the lateral distractor runs. This bias is consistently away from the distractor position in the run and builds up over the course of minutes. The contextual bias may be caused by either bottom-up or top-down factors.

ACKNOWLEDGMENTS

We thank Rachel Keen and Rich Freyman from the University of Massachusetts, Amherst, for providing help and

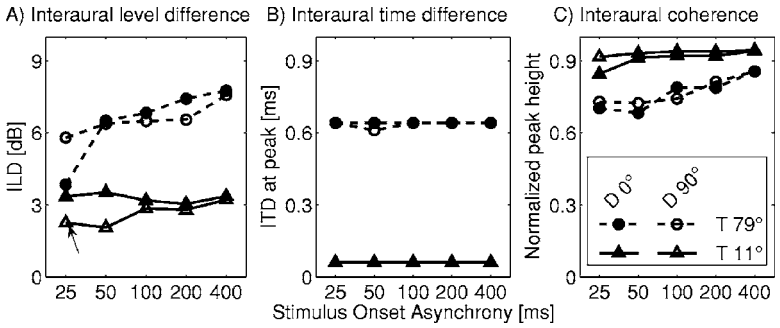


FIG. 6. Analysis of the acoustic effect of the distractor on the target for four combinations of the distractor/target locations, plotted as a function of the SOA. A) Interaural level difference. B) Interaural time difference (ITD) determined as a delay corresponding to the peak in the broadband interaural cross correlation. C) Height of the peak of the interaural cross correlation. The arrow indicates the condition in which the largest bias in the classroom, but not in the anechoic room, was observed.

access to their anechoic chamber. Armin Kohlrausch, Ruth Litovsky, and two anonymous reviewers provided helpful feedback on an earlier draft of this paper. This work was supported in part by the National Institute on Deafness and Other Communication Disorders Grant No. 5R01DC005778-03.

APPENDIX: ACOUSTIC ANALYSIS OF REVERBERANT STIMULI

Binaural recordings of several stimulus pairs used in the behavioral experiments in the classroom were obtained using a KEMAR manikin. A manikin with microphones fit in its ear canals (Etymotic Research ER 11 attached to a DB-100 Zwislocki Coupler) was placed on the chair in the center of the loudspeaker setup with its ears at approximately the level of the human subjects' ears. Signals from distractor/target pairs were recorded and digitized using a SR785 Dynamic Signal Analyzer. The digitized signals were then transferred to a PC for analysis. Four distractor/target stimulus pairs were chosen (distractor 0°/target 11°, distractor 0°/target 79°, distractor 90°/target 11°, and distractor 90°/target 79°), representing two distractor locations and targets both near to and far from the distractor positions. Recordings were made for SOAs of 25 and 400 ms, while results for other SOAs (50, 100, and 250 ms) were simulated from the 400 ms recording by shifting the target portion of the recording appropriately in time.

The recordings were analyzed to examine any consistent effects of the distractor (and its reflections) on the interaural parameters (interaural level differences or ILDs, interaural time differences or ITDs, and interaural coherence) present during the onset of the direct sound from the target.

ILD values were estimated by taking a 20 ms window, starting at the onset of the target click, and calculating the overall intensity difference in dB between the left and right ear signals using the formula $\mathrm{ILD}=20\times(\log_{10}\mathrm{rms}_{\text{right ear}}-\log_{10}\mathrm{rms}_{\text{left ear}})$. The onset of the target click was estimated from the recording by finding the first sample at which the signal passed a threshold (set by eye to be just above the level of the reverberant tail of the distractor). Panel A of Fig. 6 shows that the ILD is approximately 3 dB for targets located at 11° and ranged between 4 and 8 dB for targets located at 79°. The distractor reduced this ILD in an SOA-dependent matter in three out of the four spatial configurations (i.e., the ILD magnitude was smallest at the shortest SOAs).

The most prominent localization bias caused by a preceding distractor was an attraction of frontal targets by the lateral distractor for short SOAs in the classroom (solid lines in Fig. 4(B)). Analysis shows that the ILD is smaller for this configuration than ILD observed at longer SOAs (see the open triangle in panel A that is marked by the arrow). If anything, such a decrease in ILD magnitude predicts a localization bias *towards* the median plane, opposite the observed behavioral bias. Thus, the large lateral bias of frontal targets in the presence of a lateral distractor cannot be explained by distractor-induced changes in ILD.

ITD was estimated by calculating the normalized, running broadband interaural cross correlation of the left and right ear signals using a 5 ms analysis window with 0.6 ms cosine-squared ramps. Panel B of Fig. 6 plots the time delay at which the peak of the cross-correlation function occurs, and panel C plots the height of this normalized cross-correlation peak. Panel B shows that the ITD is approximately 640 μs for the target at 79° and approximately 60 μs for the target at 11°. There is no significant change in the ITD corresponding to the cross-correlation peak with changes in SOA, suggesting that the localization biases observed in the room were not due to shifts in the dominant ITD (a result consistent with previous analysis of how spatial cues are affected by reverberant energy; see Shinn-Cunningham *et al.*, 2005). However, the distractor reduced the interaural coherence of the target, an effect that decreased systematically with increasing SOA, especially for lateral targets (consider the open and filled circles in panel C, which increase from left to right).

[1]If all subjects show similar patterns of results, but differ in the overall magnitude of an effect, the standard deviation in the mean taken across subjects will tend to overestimate the variability in the results and underestimate the reliability of the observed effects. Such subject differences are automatically taken into account in many statistical techniques (e.g., in repeated-measures ANOVA analysis) in order to improve the power in the test. One way of accounting for differences is to only analyze changes in observations within subject, thereby reducing the across-subject variation (e.g., compare error bars in Figs. 2 and 4). However, such a representation makes it difficult to compare overall size of an effect. An alternative approach is to plot data using within-subject rather than across-subject standard deviations to show the reliability of the observed effects. Conceptually, plotting data in this way is identical to subtracting off overall differences in subject performance prior to computing the variability within a particular experimental condition. Practically speaking, such a representation is identical to assuming a statistical linear model in which each subject and each condition have effects (e.g., analogous to the assumptions in a repeated-measures ANOVA). Mathematically, the within-subject standard error of the mean can be computed as follows. Let X_{ij} denote the observations obtained for subject i observed in condition j, where each

condition corresponds to a particular, unique combination of the levels of the factors in an experiment (using this terminology, each experiment reported here had seven target lateralities $\times 6$ SOA conditions $\times 2$ distractor locations $=84$ conditions). For example, if analyzing standard deviations in responses, X_{12} corresponds to the standard deviation obtained for subject 1 in condition 2, where condition 2 denotes a particular combination of target azimuth, SOA, and distractor location. Let $\bar{X}_i$ represent the mean of the observations for subject i averaged across all (84) conditions, and let $\bar{X}$ represent the grand mean (i.e., the average across all subjects and all conditions). To account for between-subject differences in overall performance, individual differences in subject performance are removed by computing $Y_{ij}=X_{ij}-\bar{X}_i+\bar{X}$. By construct, the across-subject mean of Y_{ij} is equal to the across-subject mean of X_{ij}. However, the observations Y_{ij} remove the average difference between the observations for subject i and the grand mean by subtracting off $\bar{X}_i-\bar{X}$. The standard error of the mean of Y_{ij} gives the within-subject variability. As an example, the error bars in Fig. 5 use the within-subject standard errors, computed as described here. The (across-subject) standard errors of the mean are, on average, 1.59 times larger.

[2]Perceptual integration of target and distractor spatial information is likely to become stronger as the spatial separation of the distractor and target decreases (and it is more likely that the listener perceives distractor and target as coming from the same source at a fixed location). However, instead of growing stronger as the target location approaches the distractor location, the attraction observed here disappears. A possible explanation is that other factors come into play when target and distractor are close, counteracting the expected integration of spatial information.

[3]Given the presence of the contextual bias in the perceived target locations it would also be useful to know how the distractor location was perceived. This location was not measured directly. Listeners were highly familiar with the experimental setup and always knew the exocentric location of the distractor (both visually and by hearing an example). Although listeners were instructed to ignore the distractor and to only judge target location, casual reports from the listeners suggested that they were always aware of the distractor location during a run, both because they knew where the distractor would be within the run and because they heard multiple presentations coming from the distractor location within the run (i.e., on 83% of the trials). However, given the contextual bias observed in target localization, there is a reasonable chance that the perceived/memorized location of the distractor also shifted throughout a run. For instance, if the contextual bias is caused by a uniform shift in some internal neural representation of space (as suggested by the constant size of the shift seen in Fig. 3(A)), the perceived distractor location may have shifted in the same direction as the target. Alternatively, if the contextual bias caused by the repeated presentation of the distractor in a run occurs because listeners adopt the strategy of attending away from the distractor, the perceived distractor location may shift in the opposite direction. These possibilities could be tested by directly measuring the perceived distractor location throughout a run. However, whether or not the perceived location of the distractor itself shifted during an experimental run, the current results show that the repeated presentation of a distractor caused a number of systematic changes in how target stimuli were localized.

Best, V., van Schaik, A., Jin, C., and Carlile, S. (**2005**). "Auditory spatial perception with sources overlapping in frequency and time," Acta. Acust. Acust. **91**, 421–428.

Blauert, J. (**1997**). *Spatial Hearing* (MIT Press, Cambridge, MA).

Braasch, J., and Hartung, K. (**2002**). "Localization in the presence of a distracter and reverberation in the frontal horizontal plane. I. Psychoacoustical data," Acta. Acust. Acust. **88**, 942–955.

Butler, R. A., and Naunton, R. F. (**1962**). "Some effects of unilateral auditory masking upon the localization of sound in space," J. Acoust. Soc. Am. **34**, 1100–1107.

Carlile, S., Hyams, S., and Delaney, S. (**2001**). "Systematic distortions of auditory space perception following prolonged exposure to broadband noise," J. Acoust. Soc. Am. **110**, 416–424.

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. Gilkey, and T. Anderson (Erlbaum, New York), pp. 233–256.

de Vries, D., Hulsebos, E. M., and Baan, J. (**2001**). "Spatial fluctuations in measures for spaciousness," J. Acoust. Soc. Am. **110**, 947–954.

Dizon, R. M., and Colburn, H. S. (**2006**). "The influence of spectral, temporal, and interaural stimulus variations on the precedence effect," J. Acoust. Soc. Am. **119**, 2947–2964.

Duda, R. O., Avendano, C., and Algazi, V. R. (**1999**). "An adaptable ellipsoidal head model for the interaural time difference," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, Phoenix, Arizona, pp. II-965–968.

Fitzpatrick, D. C., Kuwada, S., Kim, D. O., Parham, K., and Batra, R. (**1999**). "Responses of neurons to click-pairs as simulated echoes: Auditory nerve to auditory encoding," J. Acoust. Soc. Am. **106**, 3460–3472.

Good, M., and Gilkey, R. H. (**1996**). "Sound localization in noise: The effect of signal-to-noise ratio," J. Acoust. Soc. Am. **99**, 1108–1117.

Hartmann, W. M., and Rakerd, B. (**1989**). "On the minimum audible angle—a decision theory approach," J. Acoust. Soc. Am. **85**, 2031–2041.

Hartmann, W. M., and Rakerd, B. (**1999**). "Localization of sound in reverberant spaces," J. Acoust. Soc. Am. **105**, 1149.

Hartung, K., and Trahiotis, C. (**2001**). "Peripheral auditory processing and investigations of the "precedence effect" which utilize successive transient stimuli," J. Acoust. Soc. Am. **110**, 1505–1513.

Heller, L. M., and Trahiotis, C. (**1996**). "Extents of laterality and binaural interference effects," J. Acoust. Soc. Am. **99**, 3632–3637.

Jack, C. E., and Thurlow, W. R. (**1993**). "Effects of degree of visual association and angle of displacement on the 'ventriloquism' effects," Percept. Mot. Skills **37**, 967–979.

Kashino, M., and Nishida, S. (**1998**). "Adaptation in the processing of interaural time differences revealed by the auditory localization aftereffect," J. Acoust. Soc. Am. **103**, 3597–3604.

Kollmeier, B., and Gilkey, R. H. (**1990**). "Binaural forward and backward masking: Evidence for sluggishness in binaural detection," J. Acoust. Soc. Am. **87**, 1709–1719.

Kopčo, N., Ler, A., and Shinn-Cunningham, B. G. (**2001**). "Effect of auditory cuing on azimuthal localization accuracy," J. Acoust. Soc. Am. **109**, 2377.

Kopčo, N., and Shinn-Cunningham, B. G. (**2002**). "Effect of cueing on sound localization accuracy in a room," in *25th Mid-Winter Meeting of the Association for Research in Otolaryngology*, St. Petersburg Beach, FL.

Litovsky, R. Y., Colburn, H. S., Yost, W. A., and Guzman, S. J. (**1999**). "The precedence effect," J. Acoust. Soc. Am. **106**, 1633–1654.

Litovsky, R. Y., and Delgutte, B. (**2002**). "Neural correlates of the precedence effect in the inferior colliculus: Effect of localization cues," J. Neurophysiol. **87**, 976–994.

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**, 753–758.

Litovsky, R. Y., and Yin, T. C.T. (**1998**). "Physiological studies of the precedence effect in the inferior colliculus of the cat. II Neural mechanisms," J. Neurophysiol. **80**, 1302–1316.

Middlebrooks, J. C., and Green, D. M. (**1991**). "Sound localization by human listeners," Annu. Rev. Psychol. **42**, 135–159.

Mills, A. W. (**1958**). "On the minimum audible angle," J. Acoust. Soc. Am. **30**, 237–246.

Perrett, S., and Noble, W. (**1995**). "Available response choices affect localization of sound," Percept. Psychophys. **57**, 150–158.

Perrott, D. R., and Pacheco, S. (**1989**). "Minimum audible angle thresholds for broadband noise as a function of the delay between the onset of the lead and lag signals," J. Acoust. Soc. Am. **85**, 2669–2672.

Phillips, D. P., and Hall, S. E. (**2005**). "Psychophysical evidence for adaptation of central auditory processors for interaural differences in time and level," Hear. Res. **202**, 188–199.

Reale, R. A., and Brugge, J. F. (**2000**). "Directional sensitivity of neurons in the primary auditory (AI) cortex of the cat to successive sounds ordered in time and space," J. Neurophysiol. **84**, 435–450.

Recanzone, G. H. (**1998**). "Rapidly induced auditory plasticity: The ventriloquism aftereffect," Proc. Natl. Acad. Sci. U.S.A. **95**, 869–875.

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.

Roberts, R. A., Besing, J. M., and Koehnke, J. (**2004**). "Effects of reverberation on fusion of lead and lag noise burst stimuli," Hear. Res. **187**, 73–84.

Shinn-Cunningham, B. G., Kopčo, N., and Martin, T. J. (**2005**). "Localizing nearby sound sources in a classroom: Binaural room impulse resonses," J. Acoust. Soc. Am. **117**, 3100–3115.

Spence, C. J., and Driver, J. (**1994**). "Covert spatial orienting in audition: Exogenous and endogenous mechanisms," J. Exp. Psychol. Hum. Percept. Perform. **20**, 555–574.

Stecker, G. C., Harrington, I. A., and Middlebrooks, J. C. (**2005**). "Location

coding by opponent neural populations in the auditory cortex," PLoS Biol. **3**, e78.
Strybel, T. Z., and Fujimoto, K. (**2000**). "Minimum audible angles in the horizontal and vertical planes: Effects of stimulus onset asynchrony and burst duration," J. Acoust. Soc. Am. **108**, 3092–3095.
Thurlow, W. R., and Jack, C. E. (**1973**). "Some determinants of localization-adaptation effects for successive auditory stimuli," J. Acoust. Soc. Am. **53**, 1573–1577.
Tollin, D. J. (**1998**). "Computational model of the lateralisation of clicks and their echoes," in *Proceedings of the NATO ASI on Computational Hearing*, edited by S. Greenberg, and M. Slaney, Il Ciocco, Italy, pp. 77–82.
Warren, D. H., Welch, R. B., and McCarthy, T. J. (**1981**). "The role of visual-auditory 'compellingness' in the ventriloquism effect: Implications for transitivity among the spatial senses," Percept. Psychophys. **30**, 557–564.
Yin, T. C. T. (**1994**). "Physiological correlates of the precedence effect and summing localization in the inferior colliculus of the cat," J. Neurosci. **l4**, 5170–5186.
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.

The acoustic and perceptual effects of two noise-suppression algorithms

Justin A. Zakis[a] and Christi Wise
Dynamic Hearing Pty Ltd, 2 Chapel Street, Richmond, Victoria 3121, Australia

(Received 10 May 2006; revised 2 November 2006; accepted 3 November 2006)

Internal noise generated by hearing-aid circuits can be audible and objectionable to aid users, and may lead to the rejection of hearing aids. Two expansion algorithms were developed to suppress internal noise below a threshold level. The multiple-channel algorithm's expansion thresholds followed the 55-dB SPL long-term average speech spectrum, while the single-channel algorithm suppressed sounds below 45 dBA. With the recommended settings in static conditions, the single-channel algorithm provided lower noise levels, which were perceived as quieter by most normal-hearing participants. However, in dynamic conditions "pumping" noises were more noticeable with the single-channel algorithm. For impaired-hearing listeners fitted with the ADRO® amplification strategy, both algorithms maintained speech understanding for words in sentences presented at 55 dB SPL in quiet (99.3% correct). Mean sentence reception thresholds in quiet were 39.4, 40.7, and 41.8 dB SPL without noise suppression, and with the single- and multiple-channel algorithms, respectively. The increase in the sentence reception threshold was statistically significant for the multiple-channel algorithm, but not the single-channel algorithm. Thus, both algorithms suppressed noise without affecting the intelligibility of speech presented at 55 dB SPL, with the single-channel algorithm providing marginally greater noise suppression in static conditions, and the multiple-channel algorithm avoiding pumping noises. © *2007 Acoustical Society of America.*
[DOI: 10.1121/1.2401656]

PACS number(s): 43.66.Ts, 43.71.Ky, 43.50.Ba [DOS] Pages: 433–441

I. INTRODUCTION

Electronic circuits generate noise, and this internally generated noise is injected into the audio signal of hearing aids and other communications devices. If amplified to a sufficiently high level, internal noise in an aid's lower frequency channels may mask soft speech sounds in adjacent higher frequency channels, especially since simultaneous masking effects tend to worsen at lower sound-pressure levels relative to normal hearing with sensorineural hearing loss (Martin and Pickett, 1970; Stelmachowicz *et al.*, 1987; Nelson and Schroder, 1997). Intermodulation products, which are spectral components generated by the aid at the sum and difference frequencies of pairs of simultaneously presented input frequencies, can also be created and act as a source of masking (Agnew, 1988). The constant hissing sound of internal noise can be annoying to hearing-aid users, particularly in quiet environments, and may even lead to the rejection of a hearing aid or possibly hearing aids in general. Internal noise may become a greater problem as aid usage among people with milder hearing losses increases due to the advent of open-ear fittings. Therefore, the reduction of the level of internally generated noise at the output of a hearing aid is an important issue that can positively affect speech intelligibility and hearing-aid acceptance and usage.

In a typical Digital Signal Processor (DSP) hearing aid, noise is injected into the audio signal by the microphone, preamplifier, analog-to-digital converter (ADC), and the output circuits that drive the receiver. Types of noise generated in electronic hearing-aid circuits include *shot, flicker, burst*, and *thermal* noise (Looney, 1993; Agnew, 1997). Thermal noise due to the random motion of air molecules is the dominant noise source in a typical low-noise, hearing-aid microphone from 20 Hz to 10 kHz (Thompson *et al.*, 2002). The rated equivalent input noise of hearing-aid microphones is usually around 25–30 dBA. It is usually the case that noise sources at the input of a hearing aid, such as the microphone, will dominate those at the output because they are amplified by the normal operation of the aid. Another type of circuit noise that is found in digital hearing aids is *quantization* noise, which is due to the approximation of exact analog signal values into a smaller set of digital values in the ADC. The quantization noise power is proportional to the square of the voltage difference represented by adjacent digital sample values. The quantization noise power is usually low since most modern hearing aids use 16-bit samples, which places the quantization noise approximately 107 dB below the maximum input voltage of the ADC (Looney, 1993).

The audibility of internal noise depends on its level after amplification by the aid relative to the hearing threshold levels (HTLs) of the aid user (Agnew, 1997), and whether it is masked by ambient background noise or a desired signal, such as speech. Modern digital hearing aids usually apply more gain in quiet conditions than in higher-SPL environments, which increases the level of internal noise in situations where it is most likely to be noticeable and annoying to aid users. In these situations, a noise-suppression algorithm may be effective in reducing the level of noise generated on the input side of the DSP (note that such an algorithm cannot reduce noise generated by circuits on the output side of the

[a]Electronic mail: jzakis@dynamichearing.com.au

DSP). A major limitation is that only a small fraction of the 1 to 5 MIPS of processing power that is typically available in a DSP hearing aid may be allocated to the noise-suppression algorithm. This limitation precludes the use of elaborate single-microphone, noise-suppression algorithms that are employed in telephony and other audio applications. Spectral subtraction and Wiener filtering have been evaluated in hearing aids with mixed results (Levitt *et al.*, 1993; Nilsson *et al.*, 2000), and may reduce speech intelligibility in quiet conditions (Nilsson *et al.*, 2000). For optimal performance these algorithms require an accurate estimate of the noise spectrum, which is difficult to obtain in complex, real-life conditions, where there may be multiple speakers with few nonspeech periods and nonstationary noise sources. The above problems can lead to the generation of acoustic artifacts, or even the suppression of speech if the speech-detection algorithm is not reliable. A less computationally demanding approach is the downward expansion of the audio signal below a threshold level in one or more frequency channels. Practical advantages of this approach are that it does not require the formation of a noise estimate, and if carefully designed it generates few artifacts. It is suitable for the suppression of internal hearing-aid noise and low-level environmental noise that aid users may find objectionable. The use of expansion below a threshold to suppress internal hearing-aid noise is not a new concept. However, there is very little literature on the perceptual effects of such algorithms in hearing aids, and the effect on speech intelligibility for listeners with impaired hearing.

In a recent article, Plyler *et al.* (2005) evaluated an expansion algorithm in a single-channel hearing aid that used attack and release times of 512 ms, an expansion ratio (ER) of 0.5:1, and an expansion threshold (ET) of 50 dB SPL. Above the threshold, the aid compressed the audio signal according to the recommendations of the NAL-NL1 procedure (Dillon, 1999). For the objective evaluation of the algorithm, the reduction in the sound-pressure level at the output of the aid and reductions in speech recognition were measured in quiet and in a background of speech-shaped white noise. Expansion significantly reduced speech understanding in quiet and in noise when the speech level was 40 or 50 dB SPL, but not 60 dB SPL. For the subjective evaluation, participants rated the sufficiency of the noise suppression in a range of quiet and low-level, real-life listening environments over a period of 2 weeks. Expansion increased satisfaction with background noise levels and was generally preferred to no expansion. The research described here extends the existing literature by comparing single-channel and multiple-channel noise suppression, and by investigating the perceptual effects of increasing and decreasing noise levels known as "pumping."

This paper presents the design and comparative evaluation of two algorithms that utilize expansion below a threshold level to suppress low-level noise in a modern DSP hearing aid. The aim of both algorithms was to suppress internal noise while preserving the intelligibility of speech presented at 55 dB SPL. The paper will focus on the multiple-channel noise suppression (MCNS) algorithm, which was designed to suppress sounds below long-term average speech levels independently in each channel. MCNS is compared with a previously developed single-channel noise suppression (SCNS) algorithm (Fiket *et al.*, 2005), which was designed to suppress sounds below a particular A-weighted input level, thereby relating the degree of noise suppression to the loudness of the noise as perceived by normal-hearing listeners. The instrumental testing of each algorithm consisted of acoustic measurements at the output of the test aid in static conditions, and calculations of the speech intelligibility index (SII) from the amplified signal. For the subject testing, normal-hearing listeners rated the loudness of the internal noise and the noticeability of pumping noises generated by each algorithm, and speech understanding with each algorithm was measured with hearing-impaired listeners fitted with aids using the ADRO® amplification strategy (Blamey, 2005).

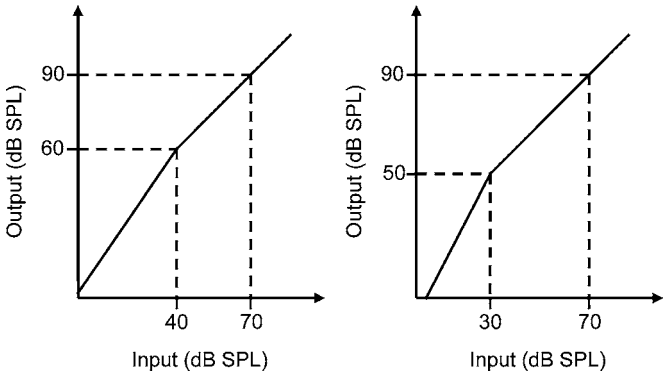


FIG. 1. Example of expansion below a threshold level in one frequency channel. For the left graph, the expansion ratio is 0.7:1 below the threshold of 40 dB SPL, resulting in a 0.43-dB gain reduction for every 1 dB the input level is below the threshold. For the right graph, the expansion ratio is 0.5:1 below a threshold of 30 dB SPL, resulting in a 1-dB gain reduction for every 1 dB the input level is below the threshold. In both cases, the operation of the amplifier is linear above the threshold with a maximum gain of 20 dB.

II. ALGORITHM DESIGN

The SCNS and MCNS algorithms both employed expansion below a threshold level independently in each frequency channel to suppress internal hearing-aid noise. Fig. 1 shows two examples of input-output functions that are expansive below an input threshold level, and linear (with 20 dB of gain) above the threshold. The left graph shows an expansion ratio (ER) of 0.7:1 below an expansion threshold (ET) of 40 dB SPL, and the right graph shows an ER of 0.5:1 below an ET of 30 dB SPL. The ER is the ratio between the change in the input level and the resultant change in the output level. Thus, with an ER of 0.7:1, every 0.7-dB reduction in the input level below the ET results in a 1-dB reduction in the output level. As a result, sounds that are just below the ET receive minimal suppression, while lower-level sounds, such as internal noise, receive a greater degree of suppression. Above the ET, no gain reduction is applied and the operation of both algorithms is linear.

Both algorithms were designed to run on an integrated circuit (IC) with a 16-bit DSP core, a 16-bit ADC, a configurable FFT coprocessor, and digital output to drive the

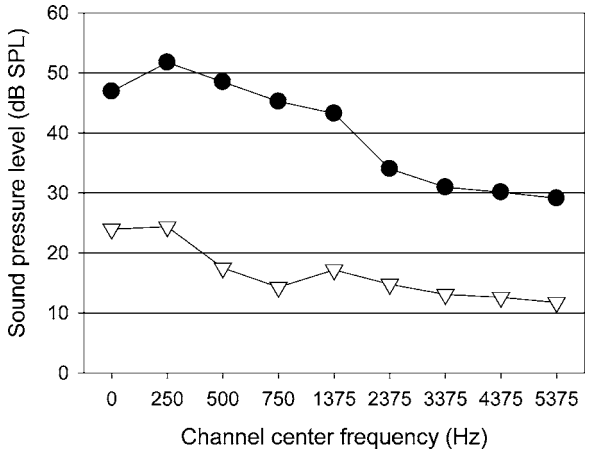


FIG. 2. Speech-shaped expansion thresholds (circles) for each MCNS channel for a wideband expansion threshold of 55 dB SPL, and typical noise levels specified for the Knowles low-noise EM-series of hearing-aid microphone (triangles) in each channel when using a 64-point FFT.

hearing-aid receiver. The IC was configured to sample the audio signal at a rate of 16 kHz, and process the samples with a 64-sample FFT every 32 samples. A spectral estimate of the audio signal was computed every 2 ms, with each estimate consisting of 32 FFT bins with center frequencies at multiples of 250 Hz from 0 to 7750 Hz.

A. Multiple-channel noise suppression (MCNS)

MCNS suppressed internal noise independently in each of nine MCNS channels. The recommended ET for each channel was based on the long-term average speech spectrum (LTASS) for "casual" speech, as measured by Pearsons *et al.* (1977) at a distance of 1 m for 100 talkers. The mean wideband level for "casual" speech was approximately 55 dB SPL for male and female adults. Thus, for convenience, a single, wideband, MCNS ET value of 55 dB SPL will be used to specify that the ET for each channel approximated the long-term average speech level at the channel's center frequency for 55-dB SPL speech. Figure 2 shows the ET for each channel (circles) for the recommended wideband ET of 55 dB SPL, and the typical noise level specified for Knowles, low-noise, EM-series, hearing-aid microphones (triangles) within each channel. The channel ET values were moved up or down together for different values of the wideband ET. For example, the channel ETs shown in Fig. 2 were reduced by 10 dB for a wideband ET of 45 dB SPL. The nine MCNS channels were formed from the FFT bins. Below 1 kHz, each channel was formed from a single FFT bin, which provided the maximum practical frequency resolution around the vowel formant frequencies. For bins from 1 to 5.75 kHz, each channel was formed by combining four consecutive FFT bins. Thus, the five highest-frequency channels consisted of the 1.00–1.75 kHz, 2.00–2.75 kHz, 3.00–3.75 kHz, 4.00–4.75 kHz, and 5.00–5.75 kHz bins. The input level for each channel was calculated from the sum of the power estimates for the FFT bins within the channel. The suppression gain for each channel was calculated from the channel's input level, and applied to the FFT bins within the channel. Wider channels were used above 1 kHz to smooth the random amplitude fluctuations of the internal noise in each channel (note that the magnitude of the amplitude fluctuations was expanded by the algorithm). The amplitude fluctuations were greater with narrower channels, and could be audible to normal-hearing listeners in very quiet environments. The formation of channels from a constant number of FFT bins above 1 kHz also allowed the DSP code length to be reduced and computational efficiency improved. The algorithm did not operate on bins with center frequencies of 6 kHz and greater, since the output of the aid reduced sharply after approximately 6 kHz. From approximately 6 kHz onward, the receiver's sensitivity fell sharply, which resulted in a sharp drop in the aid's output level for a given input level. The algorithm may be modified to operate at higher frequencies if used with a broadband receiver.

The ER was set to 0.7:1 to provide an effective gain reduction while maintaining sound quality. This ER results in a gain reduction of 0.43 dB for every 1-dB decrease in the input level below the ET. Note that the ER of 0.7:1 is equivalent to 1:1.43. Lower ERs result in greater noise suppression, but may also result in reduced sound quality due to the greater gain changes that are required after each FFT.

The time constants for rising and falling input levels below the ET will henceforth be referred to as the expansion offset and onset time constants, respectively. The time constants were defined as the time taken for the gain to settle to within 2 dB of the final value following a 25-dB change in the input level in the expansion region (this was based on IEC 60118-2, 1983). Experimentation with the algorithm showed that very fast time constants, such as 5 ms, resulted in large gain changes between successive FFTs that degraded speech quality and sound quality. Thus, the offset and onset time constants were respectively set to the slowest values that would maintain speech intelligibility and avoid pumping noises. Figure 2 shows that the difference between the channel ETs for a wideband ET of 55 dB SPL and typical microphone noise levels varies from approximately 31 dB in the low frequencies to 17 dB in the high frequencies. Thus, the offset time constant was set to 20 ms so that full release from noise suppression could be achieved in any channel during the first phoneme (approximately 30–40 ms) of speech following a period of silence. In conditions where the microphone noise dominates, release from expansion will be achieved in more than 20 ms in the low-frequency channels, and less than 20 ms in high-frequency channels that process consonants. An onset time constant of 200 ms was chosen to avoid an audible falling of the noise level after the cessation of speech in the same channel. This value was based on forward-masking data, which show that full release from forward masking is achieved in approximately 200 ms for normal-hearing listeners (Wilson and Carhart, 1971; Smiarowski and Carhart, 1975; Weber and Moore, 1981), and the rate of decay of forward masking is similar for normal- and impaired-hearing listeners at equal sensation levels (Glasberg *et al.*, 1987). Thus, the internal noise is masked by speech for approximately 200 ms after the cessation of the speech, so gain reductions applied during this period should not elicit a pumping percept. The exact masking duration depends on

the level difference between the speech (the masker) and the noise in the channel, with smaller differences resulting in smaller forward-masking durations.

In summary, the recommended settings for MCNS were a wideband ET of 55 dB SPL, an ER of 0.7:1, and expansion offset (rising input level) and onset (falling input level) time constants of 20 and 200 ms, respectively.

B. Single-channel noise suppression (SCNS)

SCNS operated in the time domain, and the recommended time constants and ER were the same as for MCNS (Fiket *et al.*, 2005). The recommended ET was different, since SCNS used an approximately A-weighted version of the input signal to calculate the suppression gain with the appropriate time constant. An ET of 45 dBA was previously recommended as a good compromise between maintaining the audibility of softly spoken speech and effective noise suppression (Fiket *et al.*, 2005). The A-weighting of the input signal ensured that relatively intense low-frequency noise (below approximately 50 Hz) that may be below normal hearing threshold levels did not prevent the algorithm from entering suppression. The A-weighting also related the degree of noise suppression to the loudness of the noise as perceived by normal-hearing listeners.

III. INSTRUMENTAL TESTING

The instrumental testing consisted of two experiments. For the first experiment, the noise levels at the output of a hearing aid were measured with and without each noise-suppression algorithm enabled. For the second experiment, speech was presented to the aid at 55 dB SPL, while the long-term average SPLs at the aid's output were measured and the speech intelligibility index calculated. This was performed without noise suppression, and with each noise-suppression algorithm enabled with a selection of different ET values and the recommended ER and time constants. The instrumental testing was performed with the algorithms running in a BTE hearing aid that was calibrated to have a flat gain-frequency response (as measured in a 2-cc coupler) for pure-tone inputs up to 6 kHz.

A. Effect on noise levels

1. Method

The aid was programed so that it operated in linear-gain mode with the gain set to provide either 15 or 35 dB of real-ear insertion gain (REIG) for an average adult ear, using average real-ear unaided gain and real-ear-to-2-cc-coupler difference values from Tables 4.1 and 4.6 of Dillon (2001). The aid was placed in a Brüel & Kjær type 4232 anechoic test box, which was placed on a table in a double-walled, sound-insulating test booth. The aid's front microphone port was blocked to minimize the entry of ambient noise (the rear microphone signal was not processed). The receiver was coupled to a Brüel & Kjær type 4946 2-cc coupler via the BTE's ear hook, a 25-mm length of plastic hearing-aid tubing (size #13), and a Brüel & Kjær DB3869 ear-mold simulator. The 2-cc coupler was connected to a Brüel & Kjær Type 2260 Investigator, which measured the L_{eq} with slow time weighting and linear frequency weighting. The L_{eq} data were measured at each gain setting without noise suppression, and with SCNS or MCNS enabled with the recommended settings. The measurement data for each condition were saved in the Investigator and uploaded onto a personal computer for analysis.

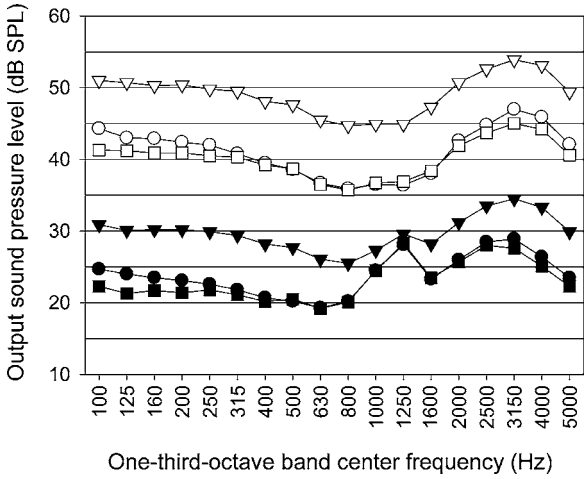


FIG. 3. Noise levels in one-third-octave bands at the output of a hearing aid measured in a 2-cc coupler. The upper curves (open symbols) are for 35 dB of real-ear insertion gain (REIG), while the lower curves (filled symbols) are for 15 dB of REIG. The output levels are shown without noise suppression (inverted triangles), and with SCNS (squares) and MCNS (circles). The expansion threshold was 45 dBA and 55 dB SPL for SCNS and MCNS, respectively. The expansion ratio was 0.7:1 for both algorithms.

2. Results

Figure 3 shows the output noise levels measured in one-third-octave bands from 100 to 5000 Hz without noise suppression, and with SCNS and MCNS with ETs of 45 dBA and 55 dB SPL, respectively, and an ER of 0.7:1. The data for 15 dB of REIG show a noise peak in the 1250-Hz band that could not be suppressed, since it was due to output-circuit noise that was amplified by acoustic resonances in the connecting tubing. This peak is absent in the data for 35 dB of REIG, which shows the effect of each algorithm more clearly. For 15 dB of REIG, the average noise-level reduction across the 100–5000-Hz bands was 6.7 and 5.9 dB with SCNS and MCNS, respectively. The noise-level differences between algorithms were less than 1.0 dB in the 250–2500-Hz bands, although this was partly due to the domination of the output-circuit noise in some bands. For 35 dB of REIG, the average noise-level reduction across the 100–5000-Hz bands was 9.0 and 8.1 dB for SCNS and MCNS, respectively. The output-level differences between algorithms were less than 1.0 dB in the 315–2000-Hz bands, with a maximum difference of 3.0 dB in the 100-Hz band. Across the bandwidth of the 100–5000-Hz one-third-octave bands, the equivalent input noise (the output noise minus the gain) with 35 dB of REIG was calculated as 27.8, 18.7, and 20.0 dB SPL without noise suppression, with SCNS, and with MCNS, respectively.

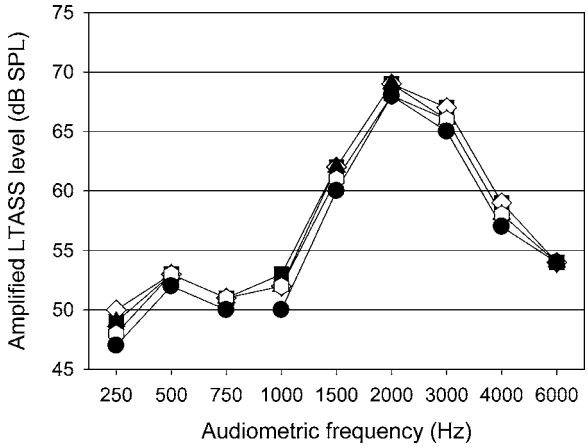


FIG. 4. The long-term average speech spectrum at the output of an aid programed with the ADRO amplification strategy for a sloping hearing loss, without MCNS (filled squares) and with MCNS using a wideband expansion threshold of 55 (open diamonds), 60 (filled triangles), 65 (open hexagons), and 70 (filled circles) dB SPL with an expansion ratio of 0.7:1.

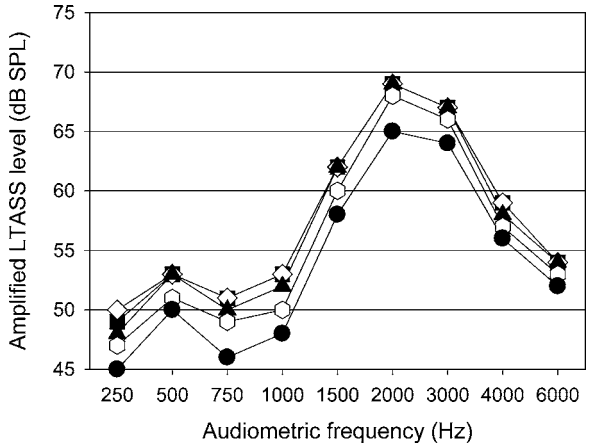


FIG. 5. The long-term average speech spectrum at the output of an aid programed with the ADRO amplification strategy for a sloping hearing loss, without SCNS (filled squares) and with SCNS using an expansion threshold of 45 (open diamonds), 50 (filled triangles), 55 (open hexagons), and 60 (filled circles) dBA with an expansion ratio of 0.7:1.

B. Effect on speech levels

1. Method

The aid was programed to provide amplification with the ADRO (Adaptive Dynamic Range Optimization) amplifier using the fitting predictions for a sloping audiogram (10, 10, 20, 30, 50, 70, 80, and 80 dB HL at 250, 500, 1000, 1500, 2000, 3000, 4000, and 6000 Hz, respectively). The ADRO amplifier is different from wide dynamic range compression amplifiers in that it uses a statistical analysis of the output level in each FFT bin and fuzzy-logic rules to calculate the bin gains that optimize the range of output levels in each bin (Blamey, 2005). The noise-suppression algorithms calculated the noise-suppression gain for each bin prior to the amplification of the audio signal, and applied the noise-suppression gains to the audio signal after amplification by ADRO. This avoided the possibility of ADRO counteracting the gain reductions applied by the noise-suppression algorithms.

The aid was placed in the chamber of an Audioscan Verifit VF-1 hearing aid analyzer and connected to the Verifit's 2-cc coupler via 25 mm of size #13 plastic tubing. The Verifit presented continuous speech at 55 dB SPL while it recorded the LTASS at the output of the aid, and calculated the speech intelligibility index with the one-third-octave-band method of ANSI S3.5 (1997) but without the 160-Hz band and without masking effects. Measurements were made with and without SCNS and MCNS activated. The noise-suppression ER was set to the recommended value of 0.7:1, while the ET was varied to assess the effect of this parameter on the amplified LTASS and SII.

2. Results

Figure 4 shows the LTASS at the output of the aid while running the ADRO amplifier without noise suppression (filled squares) and with MCNS using a wideband ET of 55 (open diamonds), 60 (filled triangles), 65 (open hexagons), and 70 dB SPL (filled circles) with an ER of 0.7:1. The average LTASS difference between processing with and without MCNS was 0 dB with an ET of 55 dB SPL. The LTASS was reduced by 1 dB at most frequencies when the ET was increased to 65 dB SPL. The calculated SII was 0.58 with and without MCNS until the ET was increased to 70 dB SPL (SII=0.57).

Figure 5 shows the LTASS at the output of the aid while running the ADRO amplifier without noise suppression (filled squares) and with SCNS using an ET of 45 (open diamonds), 50 (filled triangles), 55 (open hexagons), and 60 dBA (filled circles) with an ER of 0.7:1. The LTASS with SCNS and an ET of 45 dBA was not lower than the LTASS without noise suppression. The LTASS with SCNS and an ET of 55 dBA was at least 1 dB lower than without noise suppression at all frequencies. The calculated SII was 0.58 without noise suppression and with SCNS for ETs up to 50 dBA. The SII was 0.57 and 0.55 with ETs of 55 and 60 dBA, respectively.

IV. SUBJECT TESTING

The subject testing consisted of two experiments. The first experiment investigated the perceptual effects of the noise-suppression algorithms (i.e., noise loudness and pumping) with participants with normal hearing. The second experiment investigated the effects of the noise-suppression algorithms on the intelligibility of softly spoken speech for participants with impaired hearing fitted with the ADRO amplification strategy. The subject testing was performed with the same model of hearing aid that was used for the instrumental testing.

A. Perceptual effects

1. Methods

Twelve participants with normal hearing were recruited for the evaluation of the perceptual effects of the noise-suppression algorithms. Normal-hearing participants were recruited to minimize the effects of different HTLs and loudness-growth and masking characteristics on the evaluation results. A custom, hard-acrylic, occluding ear mold was made for each participant and inserted into the aided ear, and

a foam plug was inserted into the unaided ear to minimize entry of ambient noise. The same BTE hearing aid was fitted to each person monaurally, and the aid was programed to provide linear amplification with 15 dB of REIG for an average normal ear (see Sec. III A) in all four programs. This ensured that the internal noise would be audible, while speech would not be too loud. A flat REIG was the most appropriate fit for normal-hearing participants, as it did not unduly affect the loudness of the noise at some frequencies more than others. The aid's omnidirectional microphone was selected and the volume control was disabled.

The participants were initially trained to identify internal noise and pumping sounds by listening without noise suppression, and with SCNS configured with a relatively high ET (55 dBA). For the collection of the experimental data, the noise-suppression settings in the four programs were no noise suppression, SCNS with an ET of 45 dBA, and MCNS with ETs of 45 and 55 dB SPL. The allocation of settings to program numbers was randomized among participants who were "blind" with regard to the differences between the programs. For all noise-suppression conditions, the ER was set to 0.7:1 and the offset and onset time constants were 20 and 200 ms, respectively. All testing was performed with the participants seated in a double-walled, sound-insulating test booth with the audiologist situated outside of the booth.

In order to evaluate the noise loudness, the participants were instructed to listen to the loudness of the internal noise in all four programs and rank the programs from loudest to quietest. The participants could use the aid's pushbutton to cycle through all four programs as often as needed before deciding on the ranking (when changing programs, the aid emitted a number of short beeps that was equal to the program number).

In order to evaluate pumping effects, the participants were seated approximately 1 m from a loudspeaker and presented with the same CUNY sentence (Boothroyd *et al.*, 1985) at 55 dBA for each program (female speaker). The participants were asked to rate how noticeable pumping noises were using the following scale: 0-unnoticeable; 1-just noticeable; 2-noticeable; 3-very noticeable. This was repeated for the other three programs. A rating was given for each program as a practice run, and was repeated twice for two test runs. Three different CUNY sentences were used (sentence list 2, numbers 1–3), with the allocation of sentences to runs randomized among participants. The above was repeated with AB word lists (Boothroyd, 1968) instead of sentences, with the allocation of three word lists to three runs (one practice and two test runs) randomized among participants. Word lists were used to provide a stimulus with more opportunities to hear pumping noises since there were clear pauses between words.

2. Results

Table I shows the ranking responses of the noise-suppression algorithms according to the loudness of the internal noise for the 12 listeners with normal hearing. Eleven of the 12 participants ranked the condition without noise suppression as the loudest (one subject ranked the condition with MCNS and a wideband ET of 45 dB SPL as the loudest). All participants ranked either SCNS with an ET of 45 dBA or MCNS with an ET of 55 dB SPL as the quietest condition, with the other as the second quietest. The Kendall coefficient of concordance showed that the agreement on ranking among participants was significant ($W=0.894$, Table T critical value=0.222). The ranking of SCNS (ET =45 dBA) as quieter than MCNS (ET=55 dB SPL) was significant among participants (binomial one-tailed $p=0.0194$).

TABLE I. Ranking of single- and multiple-channel noise suppression algorithms (with different expansion thresholds) by 12 participants with normal hearing based on the perceived loudness of internal noise in quiet conditions. OFF=no noise suppression, M45=MCNS (ET=45 dB SPL), M55 =MCNS (ET=55 dB SPL), and S45=SCNS (ET=45 dBA). The expansion ratio was 0.7:1 for all noise-suppression conditions.

	Loudest	2nd loudest	2nd quietest	Quietest
OFF	11	1		
M45	1	11		
M55			9	3
S45			3	9

Table II shows the average pumping ratings for all participants for each processing condition and test stimulus. The average pumping ratings were no greater than 0.50, except for SCNS (1.88 for sentences, 2.71 for words). A Kruskal-Wallis one-way analysis of variance (ANOVA) by ranks for both sentences and words showed that overall the difference between the means for the conditions was significant ($p < 0.001$). Further analysis with a multiple-comparison test showed that the mean for SCNS (ET=45 dBA) was significantly different from the mean for any other processing condition at the $p=0.05$ level for both test stimuli. There was no significant difference between the mean without noise suppression and the mean for MCNS with either ET at the $p=0.05$ level for both test stimuli.

B. Effect on speech intelligibility

1. Methods

Ten participants with sensorineural hearing loss were recruited for the evaluation of the effect of the noise-suppression algorithms on the intelligibility of low-level speech. Nine participants were experienced hearing-aid users (two used ADRO aids, seven used compression aids), while one was a new aid user. The participants' pure-tone-average hearing loss ranged from 25 to 70 dB HL, with no more than a 5-dB difference between the ears of individual participants. All participants were fitted binaurally with the 32-channel

TABLE II. Average pumping-noise ratings for the single- and multiple-channel noise suppression algorithms (with different expansion thresholds) by normal-hearing participants for the following conditions: OFF=no noise suppression, M45=MCNS (ET=45 dB SPL), M55=MCNS (ET=55 dB SPL), and S45=SCNS (ET=45 dBA). The expansion ratio was 0.7:1 for all noise-suppression conditions. The rating scale was: 0-unnoticeable; 1-just noticeable; 2-noticeable; 3-very noticeable.

	OFF	M45	M55	S45
Sentences	0.13	0.21	0.08	1.88
Words	0.46	0.50	0.38	2.71

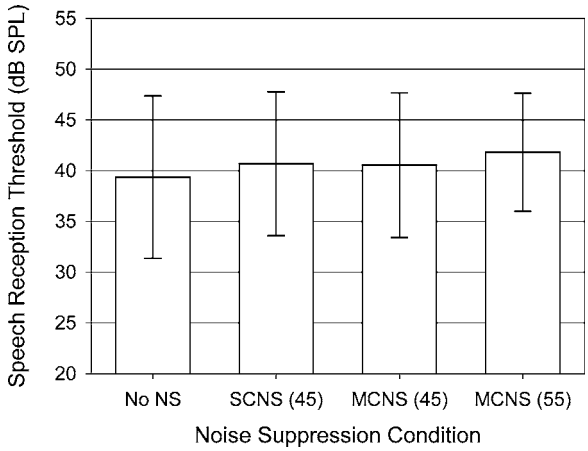


FIG. 6. Average sentence reception thresholds (SRTs) for HINT sentences in quiet for participants fitted with the ADRO amplification strategy. The bars show (from left to right) the average SRT without noise suppression, with SCNS (expansion threshold=45 dBA), and with MCNS with expansion thresholds of 45 and 55 dB SPL. The error bars show 1 standard deviation.

version of ADRO amplification with recommended fitting predictions derived from their audiograms. MCNS (recommended settings) and an adaptive-directional microphone were enabled. A second program used the telecoil input without noise suppression. The volume control was disabled (0-dB gain) for all programs. The participants used the aids in everyday environments for a period of 1 week to acclimatize to the aids.

After the acclimatization period, the participants returned for the speech tests. During testing, the participants wore aids that were configured the same way as for the acclimatization period, except that the microphone was set to omnidirectional mode. Four programs were used, and the noise-suppression conditions for the programs were no noise suppression, SCNS (ET=45 dBA), and MCNS with an ET of 45 or 55 dB SPL. The recommended ER (0.7:1) and offset and onset time constants (20 and 200 ms, respectively) were used by both noise-suppression algorithms. The allocation of the four noise-suppression conditions to the four programs was counterbalanced across participants. A Hearing-In-Noise-Test (HINT) system (Nilsson *et al.*, 1994) was used to measure the sentence reception threshold (SRT) in quiet twice for each of the four conditions after a practice run with the aid in the first program. The HINT system was also used to measure the percentage of words correctly identified in sentences presented at 55 dB SPL in quiet without noise suppression, and with SCNS and MCNS with the recommended ETs. One list of 20 sentences was presented without noise suppression, with SCNS (ET=45 dBA), and MCNS (ET =55 dB SPL). All 12 HINT sentence lists were presented to each subject without repetition during the above testing.

2. Results

Figure 6 shows the SRT in quiet averaged among participants for each noise-suppression condition. Without noise suppression, the average SRT was 39.4 dB SPL. The average SRT was 1.3 dB greater with SCNS (ET=45 dBA), and 1.2 and 2.4 dB greater for MCNS with the wideband ET set to 45 and 55 dB SPL, respectively. A repeated-measures ANOVA showed that the variations in SRT were statistically significant ($F=3.12$, df=3,27, $p<0.05$). *Posthoc* pairwise comparisons using the Bonferroni procedure showed a significant difference between MCNS with an ET of 55 dB SPL and the condition without noise suppression ($t=3.057$, df =10, $p<0.05$), and this finding was confirmed with a paired-comparison t-test ($t=4.04$, df=10, $p<0.01$). The other noise-suppression conditions had a nonsignificant effect on the SRT in quiet.

The mean percentage of words correct in sentences presented at 55 dB SPL was 99.6% without noise suppression, and 99.3% with either noise-suppression algorithm using its recommended settings (ET of 45 dBA or 55 dB SPL). The lowest individual score for any condition was 96.2%. A one-way ANOVA showed that the differences between the means without noise suppression and for either noise-suppression algorithm were not significant ($F=0.21$, df=2,27, $p=0.808$).

V. DISCUSSION

The results showed that with the recommended settings, SCNS reduced the level of internal noise at the output of the aid more than MCNS when measured in a 2-cc coupler. This finding was supported by the loudness-ranking data, which showed that most normal-hearing participants perceived the internal noise to be quietest with SCNS. A number of participants commented that it was difficult to hear a difference in loudness between the two algorithms, and some participants ranked the noise as quietest with MCNS. These subjective differences may have been due to individual variations in ear molds (sound bore shape, sealing, insertion depth) and residual ear-canal volume, as well as individual differences in hair cell survival (within the normal range) and neural processing in the brain. Figure 3 showed that with 15 dB of REIG, a peak in the output-circuit noise was dominant from approximately 630–2000 Hz, which restricted the perception of differences between algorithms to frequencies outside of this bandwidth. Whether people with impaired hearing would hear the internal noise as being louder with one algorithm or the other would depend on the above factors, as well as their audiogram, the amplification strategy used, and individual variations in psychoacoustic properties such as loudness recruitment (Fowler, 1936; Steinberg and Gardner, 1937) and loudness summation (Scharf and Hellman, 1966; Martin, 1974). The normal-hearing participants were fitted with 15 dB more gain than was required for normal loudness sensations to ensure the audibility of the microphone noise, while impaired-hearing people are not usually fit this way in clinical practice. For the aforementioned reasons, we do not expect the loudness or pumping perception of hearing-aid users to be similar to those of the normal-hearing participants.

Agnew (1997) showed that, for aid users with a moderate sensorineural hearing loss, simulated internal noise tends to become audible when it approaches the HTL at one or more audiometric frequencies, and becomes objectionable when its level is on average 8.8 dB higher than the audibility level. The noise-suppression algorithms presented in this pa-

per reduced the average noise level by up to 8–9 dB with the recommended settings, which suggests that they are capable of reducing the level of internal noise from an objectionable level to the threshold of audibility. The recommended settings are similar to the settings used in other hearing aids when "moderate" noise-suppression strength is chosen in the fitting software. A greater degree of noise suppression is possible with a lower ER and/or a higher ET, although the latter may result in decreased speech understanding in quiet. Dillon *et al.* (2003) used an Aurical hearing instrument test module (default settings) to measure the equivalent input noise of a selection of commercially available hearing aids with the noise-suppression strength set to moderate/medium. The range of reported measurement values was from 12.3 to 24.7 dB SPL. For comparative purposes, the equivalent input noise with and without SCNS and MCNS was measured with an Aurical hearing instrument test module using the default measurement settings. The equivalent input noise (average of ten measurements for each condition) without noise suppression and with SCNS and MCNS with the recommended settings were, respectively, 27.2, 19.4, and 18.9 dB SPL for international standard IEC 60118-7 (1983), and 26.1, 18.5, and 17.8 dB SPL for ANSI S3.22 (1996). These values are approximately in the middle of the range of values reported by Dillon *et al.* (2003). The slightly lower equivalent input noise value for MCNS than SCNS may seem to contradict the measurements reported in Sec. III A. This is because the above standards only specify measurements at a small number of frequencies around 1.6 kHz, where MCNS is most effective, while the data presented in Sec. III A were calculated across a 100–5000-Hz bandwidth.

An effect of the SCNS and MCNS algorithms that was not investigated was the audibility of soft environmental sounds. Sounds below the ET were assumed to be noises to suppress, with sounds just below the ET receiving less suppression than sounds that were further below the ET. A potential benefit was a reduction in the audibility of low-level environmental noises, while a potential disadvantage was a reduction in the audibility of desired soft sounds. Plyler *et al.* (2005; 2006) evaluated subjective preferences with other expansion algorithms in everyday listening environments. Hearing-aid users generally preferred single-channel expansion (ET=50 dB SPL) to no expansion (Plyler *et al.*, 2005), and gave significantly higher satisfaction ratings for four-channel expansion (ETs of 50, 50, 40, and 30 dB SPL) than no expansion (Plyler *et al.*, 2006). Thus, reduced noise levels appear to have outweighed any reduction in the audibility of desired soft sounds due to expansion. The SCNS and MCNS algorithms described in this paper used similar ETs to the algorithms evaluated by Plyler *et al.* (2005; 2006) and thus may result in a similar audibility trade-off.

The normal-hearing participants noticed pumping noises significantly more with SCNS than with MCNS. Since both algorithms reduced the microphone noise by a similar amount, the difference is likely to be primarily due to differences in the number of expansion channels. The MCNS channels were much closer to the critical bandwidth of the ear than the single 8-kHz-wide channel of SCNS. This allowed MCNS to adjust the gain only in channels where speech was present, thereby allowing speech to mask the noise as the gain was adjusted. In comparison, SCNS reduced the gain at all frequencies, reducing the forward masking effect of the speech on the noise. The pumping ratings for MCNS did not significantly differ from the ratings without noise suppression, which validates the selection of the onset time constant for MCNS. An interesting result was that some participants gave nonzero pumping ratings when noise suppression was not activated (one participant for sentences, five participants for word lists). A possible explanation is that masking effects may have been perceived as pumping effects. The audibility of the noise would have been reduced by the speech, and then fully restored approximately 200 ms after the cessation of the speech due to forward-masking effects (Wilson and Carhart, 1971; Smiarowski and Carhart, 1975; Weber and Moore, 1981). This is opposite to the effect of the noise-suppression algorithms, which increase the gain during speech and reduce the gain (and the audibility of the noise) after the cessation of speech. However, the participants were not trained to specifically listen for a decrease in the noise level after the cessation of speech, so masking effects could have been interpreted as algorithm effects. Another possible explanation for the nonzero pumping ratings without noise suppression activated is noise in the source recordings. Close examination of the recordings showed that the noise level increased shortly before the commencement of each utterance, and decreased to the original level shortly after the end of each utterance. The noise level during the utterances was fairly constant for the CUNY sentences, but the noise level audibly fluctuated during each AB word, which may partially explain the higher pumping ratings with AB words than CUNY sentences. With noise suppression activated, higher pumping ratings were expected with the AB word lists than CUNY sentences, since the former had long pauses between words for listening for pumping effects. It is possible that low levels of pumping that may have been introduced by MCNS were dominated by the pumping percepts caused by noise in the speech recordings and/or masking effects. However, the significant increase in pumping ratings with SCNS suggests that the effect of SCNS was greater than the other effects.

It was shown that with the recommended settings, the noise-suppression algorithms did not affect the SII or the long-term average level of speech (presented at 55 dB SPL) at the output of an aid running the ADRO amplification strategy for a sloping audiogram. The results for listeners with impaired hearing supported this finding, since neither algorithm significantly affected the understanding of words in sentences presented at 55 dB SPL in quiet. However, ceiling effects may have affected this finding, since six out of the ten participants had scores of 100.0% for all conditions, and no score was worse than 96.2%. This shows that each algorithm met the design objectives of suppressing internal noise while not affecting the intelligibility of speech presented at 55 dB SPL. Averaged among subjects, MCNS with an ET of 55 dB SPL resulted in a statistically significant increase in the SRT in quiet compared with the no-suppression condition (from 39.4 to 41.8 dB SPL). However, this increase in the SRT is not considered to be clinically significant, since these SRTs

were well below the mean conversational level of 55 dBA for suburban home environments (Pearsons *et al.*, 1977). MCNS with an ET of 45 dB SPL and SCNS with an ET of 45 dBA did not significantly affect the SRT. Although the SCNS and MCNS algorithms were evaluated in hearing aids, they may also be used in other audio applications where circuit noise and low-level environmental noise may be objectionable, such as wired or wireless headsets used in telephony systems. In these devices, the results with normal-hearing listeners are particularly applicable.

VI. CONCLUSIONS

This study showed that single- and multiple-channel expansion algorithms can be designed to provide similar degrees of suppression of hearing-aid circuit noise, both acoustically and perceptually, while preserving the intelligibility of speech presented at 55 dB SPL in quiet when used in conjunction with the ADRO amplification strategy. SCNS and MCNS did not significantly affect SRTs in quiet with an expansion threshold of 45 dBA or 45 dB SPL (LTASS shaped), respectively, when used with the ADRO amplification strategy. However, MCNS resulted in a statistically significant increase in the average SRT from 39.4 to 41.8 dB SPL with band ETs that followed the 55-dB SPL LTASS. The activation of noise suppression resulted in a significant increase in the noticeability of pumping sounds with SCNS, but not with MCNS, for normal-hearing listeners. Therefore, MCNS is expected to be preferred to SCNS by those listeners who can hear the pumping effect.

ACKNOWLEDGMENTS

Dr. Brenton Steele and Jutta Hau developed and tuned the SCNS algorithm. The authors would like to thank Professor Peter Blamey and Dr. Lois Martin for their many helpful suggestions and advice on the protocol and data analysis, and Anthony Shilton for his comments on a previous version of this paper. This research study was conducted with the approval of the Royal Victorian Eye and Ear Hospital Human Research and Ethics Committee (Project 03/529H).

Agnew, J. (**1988**). "Hearing instrument distortion: What does it mean for the listener?," Hear. Instr. **39**, 10–61.

Agnew, J. (**1997**). "Audible circuit noise in hearing aid amplifiers," J. Acoust. Soc. Am. **102**, 2793–2799.

ANSI (**1996**). ANSI S3.22-1996, "Specification of Hearing Aid Characteristics" (American National Standards Institute, New York).

ANSI (**1997**). ANSI S3.5-1997, "Methods for the Calculation of the Speech Intelligibility Index" (American National Standards Institute, New York).

Blamey, P. J. (**2005**). "Adaptive dynamic range optimization (ADRO): A digital amplification strategy for hearing aids and cochlear implants," Trends Amplif. **9**, 77–98.

Boothroyd, A. (**1968**). "Statistical theory of the speech discrimination score," J. Acoust. Soc. Am. **43**, 362–367.

Boothroyd, A., Hanin, L., and Hnath, T. (**1985**). "A sentence test of speech perception: Reliability, set equivalence, and short term learning," Speech & Hearing Report RCI, City University of New York, New York.

Dillon, H. (**1999**). "NAL-NL1: A new prescriptive fitting procedure for non-linear hearing aids," Hearing J. **52**, 10–16.

Dillon, H. (**2001**). *Hearing Aids* (Boomerang, Sydney).

Dillon, H., Keidser, G., O'Brien, A., and Silberstein, H. (**2003**). "Sound quality comparisons of advanced hearing aids," Hearing J. **56**(4), 1–6.

Fiket, H., Hau, J., and Steele, B. (**2005**). "Perceptual effects of noise reduction in hearing aids," Presented at the American Academy of Audiology Convention, Washington D.C., 30 March–2 April **2005**.

Fowler, E. P. (**1936**). "A method for the early detection of otosclerosis," Arch. Otolaryngol. **24**, 731–741.

Glasberg, B. R., Moore, B. C. J., and Bacon, S. P. (**1987**). "Gap detection and masking in hearing-impaired and normal-hearing subjects," J. Acoust. Soc. Am. **81**, 1546–1556.

IEC (**1983**). IEC-60118-2, "Hearing aids, Part 2: Hearing aids with automatic gain control circuits," International Electrotechnical Commission.

IEC (**1983**). IEC-60118-7, "Hearing aids, Part 7: Measurements of performance characteristics of hearing aids for quality inspection for delivery purposes," International Electrotechnical Commission.

Levitt, H., Bakke, M., Kates, J., Neuman, A., Schwander, T., and Weiss, M. (**1993**). "Signal processing for hearing impairment," Scand. Audiol. Suppl. **38**, 7–19.

Looney, C. G. (**1993**). "Noise," in *The Electrical Engineering Handbook*, edited by R. C. Dorf (CRC Press, Boca Raton).

Martin, E. S., and Pickett, J. M. (**1970**). "Sensorineural hearing loss and the upward spread of masking," J. Speech Hear. Res. **13**, 426–437.

Martin, M. C. (**1974**). "Critical bands in sensori-neural hearing loss," Scand. Audiol. **3**, 133–140.

Nelson, D. A., and Schroder, A. C. (**1997**). "Linearized response growth inferred from growth-of-masking slopes in ears with cochlear hearing loss," J. Acoust. Soc. Am. **101**, 2186–2201.

Nilsson, M., Soli, S. D., and Sullivan, J. A. (**1994**). "Development of the Hearing In Noise Test for the measurement of speech reception thresholds in quiet and in noise," J. Acoust. Soc. Am. **95**, 1085–1099.

Nilsson, M. J., Fang, X., Ghent, R. M., Murphy, P., and Bray, V. H. (**2000**). "Improved speech intelligibility in noise with a single-microphone noise reduction technique," Presented at the Joint 140th Meeting of the Acoustical Society of America/NOISE-CON 2000, California, 3–8 December **2000**, Poster 3pSC2.

Pearsons, K. S., Bennett, R. L., and Fidell, S. (**1977**). "Speech levels in various noise environments," Report No. EPA-600/1-77-025, Washington, DC, U.S. Environmental Protection Agency.

Plyler, P. N., Hill, A. B., and Trine, T. D. (**2005**). "The effects of expansion on the objective and subjective performance of hearing instrument users," J. Am. Acad. Audiol **16**, 101–113.

Plyler, P. N., Lowery, K. J., and Hamby, H. M. (**2006**). "The objective and subjective evaluation of multi-channel expansion in wide dynamic range compression hearing instruments," International Hearing Aid Research Conference 2006, poster C18, available from pplyler@utk.edu.

Scharf, B., and Hellman, R. (**1966**). "Model of loudness summation applied to impaired ears," J. Acoust. Soc. Am. **40**, 71–78.

Smiarowski, R. A., and Carhart, R. (**1975**). "Relations among temporal resolution, forward masking, and simultaneous masking," J. Acoust. Soc. Am. **57**, 1169–1174.

Steinberg, J. C., and Gardner, M. B. (**1937**). "The dependency of hearing impairment on sound intensity," J. Acoust. Soc. Am. **9**, 11–23.

Stelmachowicz, P. G., Lewis, D. E., Larson, L. L., and Jesteadt, W. (**1987**). "Growth of masking as a measure of response growth in hearing-impaired listeners," J. Acoust. Soc. Am. **81**, 1881–1887.

Thompson, S. C., LoPresti, J. L., Ring, E. M., Nepomuceno, H. G., Beard, J. J., Ballad, W. J., and Carlson, E. V. (**2002**). "Noise in miniature microphones," J. Acoust. Soc. Am. **111**, 861–866.

Weber, D. L., and Moore, B. C. J. (**1981**). "Forward masking by sinusoidal and noise maskers," J. Acoust. Soc. Am. **69**, 1402–1409.

Wilson, R. H., and Carhart, R. (**1971**). "Forward and backward masking: Interactions and additivity," J. Acoust. Soc. Am. **49**, 1254–1263.

Instantaneous orifice discharge coefficient of a physical, driven model of the human larynx

Jong Beom Park[a] and Luc Mongeau[b]
Ray W. Herrick Laboratories, School of Mechanical Engineering, Purdue University, 140 S. Intramural Drive, West Lafayette, Indiana 47907-2031

(Received 27 October 2005; revised 29 October 2006; accepted 2 November 2006)

The quasisteady approximation is often made in the study of phonatory aerodynamics to facilitate the modeling of time-varying air flows through the self-oscillating vocal folds. The unsteady, pulsating flow is approximated by a sequence of steady flows through representative configurations of the vocal folds at rest. Previous studies have discussed the accuracy of this approximation for a range of orifice geometries, and flow conditions. The purpose of the present study was to further evaluate the quasisteady approximation experimentally using an improved procedure, from a direct comparison between the discharge coefficients of steady jets through fixed orifices and unsteady jets through modulated orifices of identical shape, area, and transglottal pressures at a given time. Life-scale convergent and divergent glottis-shaped rubber orifices were used in a rigid-walled tube and a low Mach number flow representative of human phonation. It was found that the quasisteady approximation is valid during 70% of the duty cycle, when the Reynolds number was above 3000, for a frequency of oscillations of 100 Hz. The steady form of Bernoulli's equation along a streamline, and Bernoulli's flow obstruction theory were found to be reasonably accurate for the unsteady flows. These models break down at low Reynolds numbers, near the beginning and the end of the duty cycle, due to viscous effects and to the influence of flow displaced by the motion of the walls. [DOI: 10.1121/1.2401652]

PACS number(s): 43.70.Aj, 43.70.Bk, 43.28.Ra [AL] Pages: 442–455

I. INTRODUCTION

A. Review of previous work

Unsteady flow through the glottis constitutes the source of voiced sound, and determines the time-varying fluid pressure acting on the vocal-fold tissue as the vocal-folds change shape during phonation (Fant, 1960; Flanagan, 1965; Titze, 2000). The flow through the glottis is primarily governed by pressure gradients across the glottal orifice, quantified by convective acceleration terms in the equations of motion (Barney *et al.*, 1999). The motion of the vocal folds themselves, described by local acceleration contributions, may also be significant (Zhao *et al.*, 2002; Zhang *et al.*, 2002). Many previous laryngeal flow studies have been made using static physical models, which replicate time-frozen glottal shapes during motion (Gauffin *et al.*, 1983; Scherer *et al.*, 2001a, b). The fluid flow characteristics of these static replicas have been used to model the instantaneous state of the actual time-varying flows in applications such as articulatory voice synthesis. They also help the analysis and understanding of the flow-induced oscillations. Inherent in the use of quasistatic models lies the assumption that local flow acceleration effects related to wall motion are negligibly small relative to convective acceleration effects. This assumption has been referred to as the quasisteady approximation. It has been used for example in analytical models of the sound source (McGowan, 1988).

Various attempts have been made to verify the validity of the quasisteady approximation. Iguchi *et al.* (1990) studied the properties of steady and unsteady jets through an orifice with a constant area. Pelorson *et al.* (1994) evaluated the quasisteady approximation by comparing the flow separation point of steady and unsteady flows within a glottis-shaped channel. They discussed the effects of a moving flow separation point on the vocal fold's vibration pattern, and proposed numerical models with a more realistic description of vocal cord collision.

Deverge *et al.* (2003) calculated the wall pressures at the throat of moving wall channels of various shapes and made comparisons with experimental data. They considered the unsteady form of Bernoulli's equation, the steady form of Bernoulli's equation with additional terms for viscous losses, Reynolds' lubrication theory, and boundary layer equations for steady flows to establish the influence of local acceleration and viscous effects. They reported that local acceleration effects were generally negligible, except during glottal closure, for the case of a parallel wall orifice shape. Viscous effects were found to be significant, and predicted using boundary layer models for small glottal heights. Similarly, Hofmans *et al.* (2003) used boundary layer models for the wall (or "throat") pressures. They showed that predictions are more accurate when the pressure difference between separated flow and far field is used, which may indicate that boundary layer predictions are sensitive to outer flow assumptions. Vilain *et al.* (2004) studied boundary layer mod-

[a]Electronic mail: park44@purdue.edu
[b]Current address: Mechanical Engineering Department, McGill University, 3480 University Street, Montreal, Quebec, Canada H3A 2K6. Electronic mail: luc.mongeau@mcgill.ca

els for the prediction of the wall pressure for unsteady orifice flows with rigid and moving wall configurations. Their results showed that the predictions for steady and unsteady flows through straight channels were accurate, whereas they were less accurate for rounded orifice shapes.

Mongeau *et al.* (1997) conducted experiments using driven, rubber, glottis-shaped orifices to investigate the flow through the larynx. They found that the quasisteady approximation was accurate for cases where the orifice, which was converging, discharged in an open environment, with an anechoically terminated tube upstream. Zhang *et al.* (2002) used similar driven rubber physical models to extend the investigation of the range of validity of the quasisteady approximation to various orifice shapes with anechoic tube configurations on both sides of the orifice. They compared the sound pressure radiated downstream in a duct with predictions from a model that assumed plane-wave propagation with monopoles in the source region. Their evaluation relied on an iterative process from an initial guess of the downstream pressure, and on the steady form of Bernoulli's equation, ignoring the unsteady velocity potential (or local acceleration) term. The results showed that the predicted pressure was very similar to the measured one in cases where there was no strong acoustic loading upstream of the orifice. However, the orifice discharge coefficient (herein referred to as ODC) was assumed to be independent of the Reynolds number and the orifice area, which may degrade the accuracy of the prediction for the beginning and the end of the vibratory cycle, where the ODC varies rapidly with Reynolds number. The presence of acoustic loading also made their method somewhat complicated due to the need for decoupling the source and the frequency response of the reactive load through inverse filtering in the frequency domain.

From these and other recent investigations, it appears that the accuracy and range of validity of the quasisteady approximation still needs to be ascertained for unsteady glottal flows. In the present study, the axial velocity distribution of the unsteady flow through a driven, life-size rubber physical model of the glottis was measured at the orifice outlet. Other hydrodynamic and acoustic parameters such as flow rate, time-averaged, instantaneous upstream and downstream pressures were also measured. The same measurements were made independently for comparable steady flows through various orifice configurations (fixed orifice geometries). The ODC for steady flows and for comparable unsteady flows were computed and compared.

B. Theoretical background

The vocal folds act as a time-varying flow restriction that causes the static pressure to periodically rise in the subglottal space, as the area is reduced. Flow constrictions in engineering duct flow applications may be categorized into orifice plates, Venturi tubes, nozzles, and diffusers. Physical models of vocal folds generally idealize the glottis as an orifice plate, with a cross-section profile that is either converging or diverging based upon the phase of the motion. The orifice plate analogy may be preferable to the treatment of the converging orifice as a nozzle and the diverging one as a diffuser. Semiempirical models for such nozzles and diffusers are available for various types of industrial applications (Blevins, 1984). However, the relatively small dimensions of the glottal orifice relative to the vocal tract ($h/D \sim 0.05$) hampers the direct use of published standardized orifice data for the human glottis. Many characteristics of glottal flows are unique, and must be determined empirically or computationally for relevant geometries and dimensions.

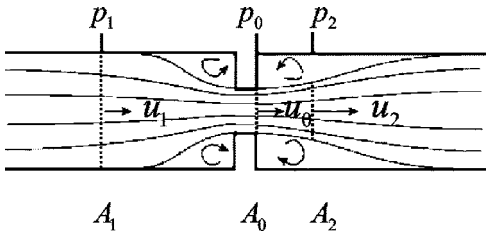


FIG. 1. Schematic diagram of flow streamlines (stream tube) through an orifice in a tube. P, u, and A denote pressure, flow velocity, and cross-sectional area of the stream tube upstream (1), at the orifice throat (o), and downstream (2).

The primary variable adopted here to characterize the flow constriction is the so-called orifice discharge coefficient, C_d, which is defined as

$$Q_{\text{real}} = C_d \cdot A_o \cdot u_2, \tag{1}$$

where Q_{real} is the actual flow rate through the orifice (as opposed to an ideal flow rate defined later), A_o is the orifice area at the throat (i.e., the minimum area), and u_2 is the flow velocity measured downstream of the orifice, within the jet core (Fig. 1). For orifices of complex shapes, for which the pressure head at the orifice throat may be different from that at the downstream measuring location, the discharge coefficient is expressed as the product of two coefficients,

$$C_d = C_L \cdot C_c, \tag{2}$$

where C_L is a ratio of the actual flow rate to the ideal flow rate,

$$C_L = Q_{\text{real}}/Q_{\text{ideal}}. \tag{3}$$

The ideal flow rate is that obtained for an incompressible, inviscid flow with a uniform velocity profile from the multiplication of the centerline flow velocity and the minimum orifice area. The contraction coefficient, C_c, is the ratio of the flow stream cross-sectional area downstream of the orifice, and the orifice area,

$$C_c^2 = \left(\frac{A_2}{A_o}\right)^2 = 1 + \frac{p_2 - p_o}{\rho u_2^2/2} + \frac{\Phi}{\rho u_2^2/2}, \tag{4}$$

where Φ is a viscous loss pressure head along a stream line. The detailed derivation of Eq. (4) is presented in Appendix A. The loss coefficient, C_L, measures flow rate losses due to the deviation of the actual velocity profile from the ideal velocity profile. Such deviation results from a pressure head loss at jet boundaries due to, for example, the kinetic energy of the entrained flow and viscous diffusion within the shear layers, as illustrated in Fig. 2. The contraction coefficient, C_c, quantifies the pressure variation (reversible process) and

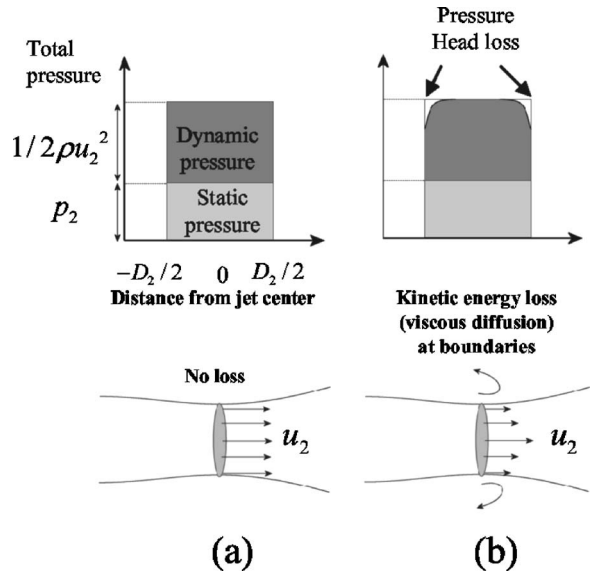


FIG. 2. An illustration of the total pressure head for (a) an idea flow with no viscous losses and (b) a flow with dynamic pressure head loss at boundaries.

viscous loss (irreversible process) along the stream line between the orifice throat and the measurement location. The contribution of each of these factors to C_c, however, may not be isolated unless the pressure distribution along a streamline is available. The orifice discharge coefficient for orifice plates is associated with the hydrodynamic changes of the jet through the orifice, and is characterized in terms of C_c and C_L.

The ODC definition, Eq. (1), is for a steady flow. The same definition can readily be extended to unsteady flows by making the orifice area, $A_o(t)$, time dependent, along with the ODC. The ODC for unsteady flows can be rewritten in terms of the time-varying flow rate, $Q_{\text{real}}(t)$, as

$$C_d(t) = \frac{Q_{\text{real}}(t)}{A_o(t) \cdot u_2(t)} = \frac{\int u(\mathbf{x},t)dA}{A_o(t) \cdot u_c(t)}, \tag{5}$$

where $u(\mathbf{x},t)$ is the velocity distribution over the jet plume cross section, and $u_c(t)$ is the velocity along a streamline through the center of the orifice. Note that the real flow rate is invariant due to the incompressible flow assumption ($Q_{\text{real}}=\int u_o(\mathbf{x})dA_o=\int u_2(\mathbf{x})dA_2$). All velocities are measured downstream, e.g., at point 2 in Fig. 1, and the jet centerline velocity, $u_c(t)$, hence replaces u_2 since they are equivalent by definition. Even when there is a pressure head loss at the jet boundary ($C_L<1$), it is shown in a subsequent section that the jet center velocity is indeed equivalent to the square root of the pressure differential since Bernoulli's equation holds for most of the oscillation period along the streamline through the orifice center.

The use of Eq. (5) in conjunction with the enforcement of measured values for velocity, area, and pressures implies that all unsteady effects are imbedded in the orifice coefficient. These unsteady effects include local flow acceleration, moving flow separation point on the wall, pressure feedback from downstream turbulence, viscous stresses, and changes in shear layer profiles. A comparison between the ODCs for unsteady and steady jets through comparable orifice configurations over one period of oscillation therefore indicates the influence of intrinsically unsteady effects. The ODC for steady jets, C_d, and for unsteady jets, $C_d(t)$, are referred to as stationary and instantaneous ODC, and noted as C_d^{st} and C_d^i, respectively, in subsequent sections.

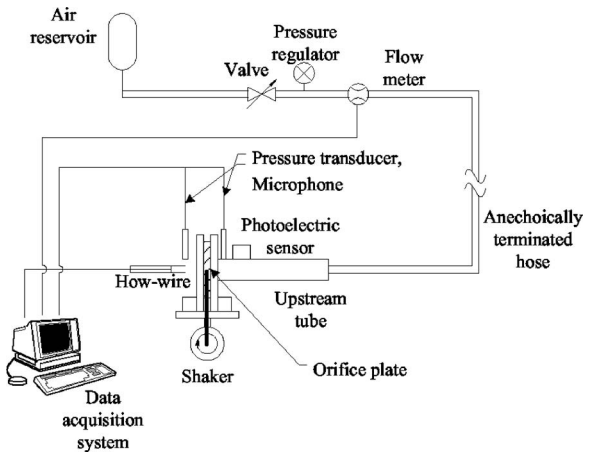


FIG. 3. Schematic diagram of the experimental setup.

II. EXPERIMENTAL METHODS

A. Experimental apparatus

A sketch of the experimental apparatus is shown in Fig. 3. Air supplied from a pressurized reservoir was used as the working gas. The time-averaged flow rate, Q_m, was measured far upstream of the test section using a precision mass flow meter, MKS 0558A. One anechoic termination was used to minimize sound wave reflections from the inlet and noise from the flow supply (Zhang *et al.*, 2002). The upstream (or "subglottal") test section consisted of a $2.2\times2.2\times20$ cm Plexiglas tube. The rubber orifice plate, sandwiched between two aluminum plates, was located at the end of the test section. The orifice plates were cast out of GE RTV11 silicone compound, and shaped after convergent and divergent lateral profiles of vocal folds with an included angle, α, as shown in Figs. 4(a) and 4(b). The angle was 30° for the convergent orifice and 20° for the divergent orifice. Details of the physical models can be found in previous related studies (Scherer *et al.*, 2001a; Zhang *et al.*, 2002). The coordinate axes, x, y, and z, are aligned along the lateral, spanwise, and streamwise directions, respectively, as shown in Fig. 4(c). A forced oscillation was employed to regulate the orifice wall motion. Two electrodynamic linear actuators, Labworks ET-126, drove the rubber orifice plate through actuating metal rods cast within the model. The transglottal, time-averaged pressure was measured using a differential pressure transducer, MKS Baratron model 100T. The upstream and downstream acoustic pressures were measured using B&K 4939 microphones: one mounted flush with the upstream tube and the other one placed downstream. Both were located 3 cm away from the orifice. A single hot-wire probe mounted on a three-dimensional linear traverse was used to measure the velocity[1] of the glottal jet in an x-y plane (referred to as *measurement plane*) located 1 mm downstream of the orifice

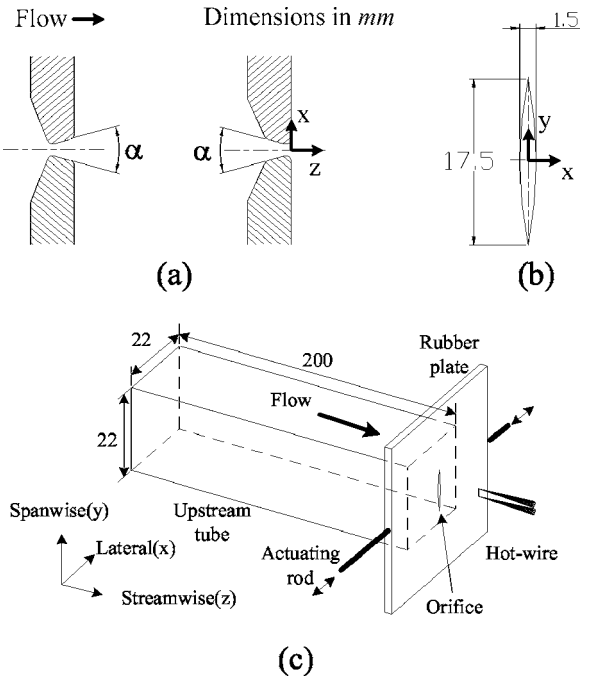


FIG. 4. Details of the orifice and test section: (a) profile of a divergent (left) and a convergent (right) orifice with an included angle, α; (b) frontal orifice dimensions. The origin of the coordinate system is located at the center of the orifice; (c) schematic of the test section and hot-wire anemometer.

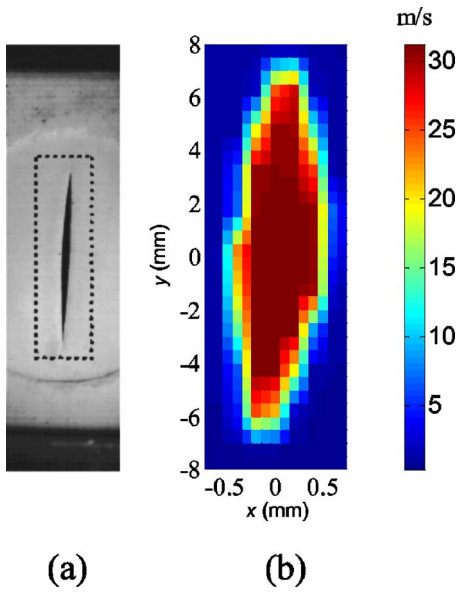


FIG. 5. (Color online) Typical velocity distribution through a convergent orifice at time of maximum opening, $t=T/2$ ($f=100$ Hz, $\Delta p=6$ cm H_2O, $Q_m=2.78\times10^{-4}$ m^3/s): (a) picture of the orifice with velocity measurement area (dotted); (b) velocity distribution of the jet, or *velocity map*, constructed from hot-wire measurements.

plate. The hot-wire anemometer yielded accurate instantaneous velocity estimates, typically within 1%, as ascertained from calibration using a TSI 1125 calibrator (Bruun, 1995). A TSI 1210 standard probe was used in an end-flow (parallel to the flow) arrangement for minimizing flow interference. Probe interference was verified to be negligible since the pressure differential across the orifice was not influenced by the presence of the probe. A photoelectric sensor, Thorlabs DET110, was placed in the upstream tube to measure the orifice area on-line, and synchronize the acquisition of other time-varying variables. A digital high-speed camera, NAC FX-K3, was used along the z direction for measuring the orifice area optically. Finally, an ion-argon laser, Lexel 85, was used for flow visualization, using vaporized oil droplets from a Rosco fog machine.

B. Measurement procedures

The area of the orifice was measured using two different methods. First, the high-speed camera (HSC) captured the orifice motion at a high frame rate (6000 fps). The pictures were converted to gray scale, and the orifice images were filtered by establishing a threshold value for the black color depth. The area was calculated from integration (pixel-by-pixel summation) of the resulting image after mapping the pixelated area with a known length reference. Synchronously with the HSC, the photoelectric sensor recorded the light intensity changes induced by the orifice motion. A correlation was established between the image-based area and the photoelectric sensor output, and used for on-line estimation of the area, $A_o(t)$. A calibration was performed at the beginning and the end of each experiment to ensure that the orifice area time history remained consistent.

The hot-wire probe was traversed along a measurement plane. The spatial resolution was 0.1 mm along x and 1.5 mm along y. One typical measured cross-sectional distribution of the flow velocity ($u_z(x,y)$) is shown in Fig. 5. The orifice was rotated slightly clockwise ($\theta\sim1.7°$), which was due to an error made during mold casting. Such a misalignment may induce a geometrical transform between nonaxial velocity components (u_x and u_y), which can be described by a rotational matrix $[\cos\theta\sin\theta,-\sin\theta\cos\theta]$. This influence is negligible because the misalignment angle is very small, and the flow velocity is predominantly axial. The acquisition of velocity signals at each location was gated, i.e., phase-locked, with respect to the shaker input signal (synchronously with the orifice motion). The velocity measurements were started after many cycles to ensure that the time-varying motion of the glottal jets reached a steady state. The velocity distribution, $u(\mathbf{x},t)(=u_z(x,y,t))$, or *velocity map*, at each instant of time was constructed by plotting all velocity data as shown in Fig. 5(b). The verification of the velocity map accuracy is discussed in the next section.

Although the glottal jet is primarily axial, the possible influence of lateral and spanwise velocity components, i.e., u_x and u_y, was considered. Spanwise flow motion may occur at the tips (or commissures) of the orifice major axis, and lateral motion may take place along the orifice walls. The spanwise velocity, u_y, was deemed negligible because the vertical deflection of the jet was found to be limited within a small region near the commissures from flow visualization. Its magnitude was less than 10% of that of the axial component since the deflection angle was small ($<10°$). In contrast, u_x was found to be significant in certain configurations. For a converging jet, u_x is negligible because the x-momentum components, gained from each convergent ori-

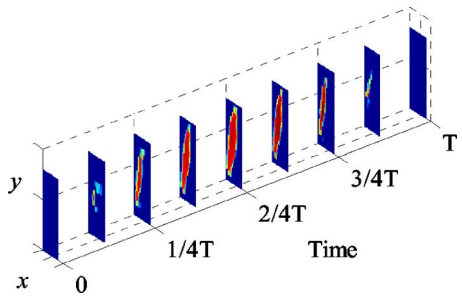


FIG. 6. (Color online) Time sequence of velocity maps. Integration of each velocity map yields the instantaneous flow rate at each sampling time.

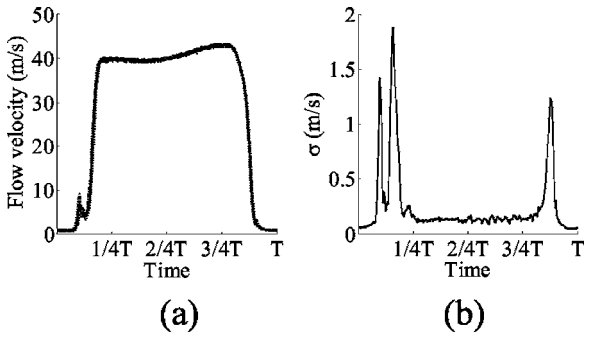


FIG. 7. Velocity vs time at $(x,y,z)=(0,0,1\ \text{mm})$: Converging orifice, $f=100$ Hz, $\Delta p=10$ cm H_2O, $Q_m=1.48\times10^{-4}$ m^3/s. (a) measured flow velocity: -.-.-.: maximum value; - - -: minimum value; ———-: ensemble average, (b) standard deviation of the ensembles.

fice wall upstream, are opposite and cancel each other. For a diverging jet, it is no longer negligible since the jet attaches to either orifice wall due to the *Coanda phenomenon*. Because the hot-wire is sensitive to both normal and binormal flow directions, but not to the tangential direction,[2] u_x was estimated from trigonometry assuming the flow direction was parallel to the diffuser wall angle, $\alpha/2$. This assumption was verified using flow visualization.

Two artifacts in the velocity data tend to bias the flow rate estimation from integration of the velocity map, $u(\mathbf{x},t)$. One is a nonzero velocity signal in absence of flow due to natural heat convection around the hot wire operated at a constant temperature around 205 °C. The other is flow induced from entrainment near the shear layers of the jet stream. To account for these factors, a cutoff velocity, u_{cut}, was introduced in the calculation of the flow rate, i.e., a lowest velocity limit below which the velocities were excluded from the integration. The value of the cutoff velocity was established such that the time-averaged value of the flow rate obtained from integration of the velocity map was equal to that measured upstream using the mass flow meter, i.e., $1/T\int u(\mathbf{x},t)dt=Q_m$. The cutoff velocity value was also based on the velocity map of steady jets for each configuration. The optimal cutoff value was determined such that the velocity integration of steady and unsteady glottal jets yielded similar estimation error values (~4%).

The frequency of the oscillations was 100 Hz for the unsteady cases. The maximum orifice width was similar to that of human vocal folds (~1 mm). The sampling frequency was 32.8 kHz, with a 30-s data record length. The pressure differentials across the orifice were selected to be 10 cm H_2O for the converging jet and 14 cm H_2O for the diverging jet, which are in the rage of the phonatory pressure for loud voice. Coverage of a wide range of Reynolds number was intended. A temporal sequence of velocity maps after filtering is shown in Fig. 6. Each frame in the figure represents the velocity map at one specific time. Area integration of each frame yields the instantaneous flow rate at that time. The time-averaged value of the instantaneous flow rate over one period, $\overline{Q(t)}$, was then compared to the time-averaged flow rate, Q_m, measured using the flow meter to ensure continuity of mass flow.

The instantaneous ODC of the jet was computed at every sampled time, $t=t_j$, from Eq. (5),

$$C_d^i(t_j)=\left.\frac{\sum u(\mathbf{x},t)\cdot\Delta A}{A_o(t)\cdot u_c(t)}\right|_{t=t_j}. \tag{6}$$

The ODCs for the steady jets were measured independently with fixed orifice geometries. The orifice area and the flow rate were adjusted to span the range of the time-varying experiments. The centerline velocity was measured in every experiment, and the stationary ODC was then calculated as

$$C_d^{st}(\text{Re})=\frac{Q_m}{A_o\cdot u_c}, \tag{7}$$

where Re is Reynolds number, $u_c\sqrt{A_o}/\nu$.

III. RESULTS

A. Converging orifice

The instantaneous centerline velocity over one cycle is shown in Fig. 7. At the very beginning of the duty cycle, for $t=T/8$, a small peak in the velocity waveform is apparent. This was due to a change in direction of the jet plume. The jet axis moved as the jet plume direction changes from slanted to straight (along z). The jet plume only temporarily wetted the hot-wire probe. The flow then accelerated rapidly due to the pressure gradient between the ambient pressure downstream and the pressurized tube upstream. After the initial velocity surge, the velocity was nearly constant until the end of the duty cycle around $t=7/8T$. The dotted lines in Fig. 7(a) were drawn by connecting minima and maxima recorded over successive velocity records to indicate the envelope of possible values. The standard deviation, σ, shown in Fig. 7(b), was less than 2 m/s over most of the duty cycle. The maximum deviation occurred early in the opening phase and was most likely due to the sweeping motion of the jet axis mentioned before. This variability, however, did not affect the flow rate significantly due to the small area at that time. This small scatter in the velocity history indicates that the pulsating jet is indeed a stationary periodic flow. A collection of phase-locked measurements thus allows the accurate reconstruction of the instantaneous velocity profile of the jet. The instantaneous flow rate is shown along with the orifice area in Fig. 8. It is smooth and symmetric throughout the cycle. The error in flow rate estimate, $|(\overline{Q(t)}-Q_m)/Q_m|$, was less than 4% over one period of oscillation.

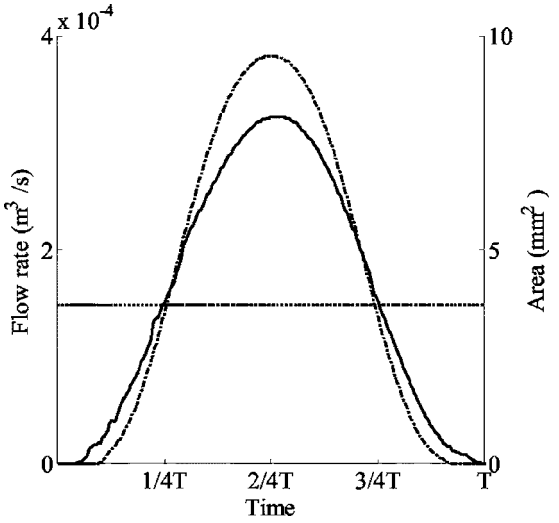


FIG. 8. Instantaneous flow rate and area function vs time (converging jet, $f=100$ Hz, $\Delta p=10$ cm H_2O, $Q_m=1.48\times10^{-4}$ m^3/s, $\overline{Q(t)}=1.54\times10^{-4}$ m^3/s): ———: $Q(t)$; –.–: $A_o(t)$. The estimation error from measured time-mean flow rate, Q_m, is about 4%.

Figure 9(a) shows the C_d^i curve over several cycles. Early in the opening phase, the coefficient is large. Its value decreases exponentially, tending asymptotically towards a value of 0.86 for large Reynolds numbers (large opening areas). The coefficient increases again along the same curve as the orifice closes, and the cycle repeats identically over the following periods. The C_d^{st} values for several orifice areas are shown versus Reynolds number in Fig. 9(b). The C_d^{st}'s follow similar trends as for the C_d^i curves. For large orifice areas, C_d^{st} asymptotically approaches a value of 0.8, almost

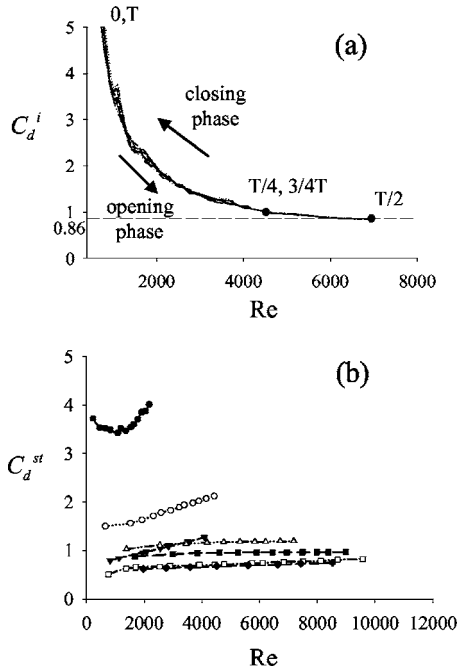


FIG. 9. Orifice discharge coefficient vs Reynolds number for (a) unsteady jets over one cycle for continuously varying area: converging orifice, $f=100$ Hz, $\Delta p=10$ cm H_2O, $Q_m=1.48\times10^{-4}$ m^3/s; and (b) steady jets for discrete areas: ●: 0.4 mm^2; ○: 1.5 mm^2; ▼: 2.1 mm^2; △: 3.5 mm^2; ■: 6.0 mm^2; □: 9.5 mm^2; ◆: 12.6 mm^2.

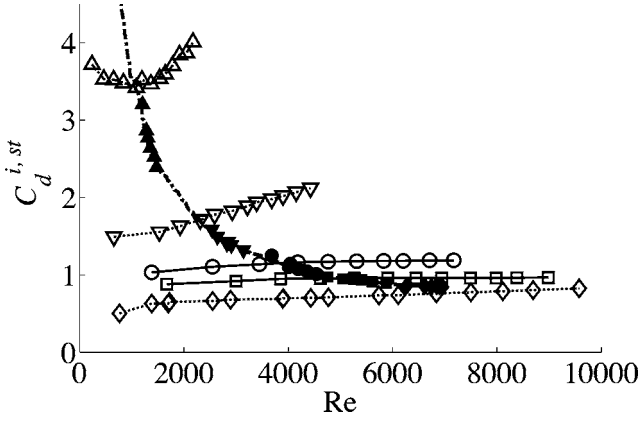


FIG. 10. Comparison between the orifice discharge coefficients of unsteady and steady jets for several opening areas. Coefficients values are the same as in Fig. 9. C_d^{st} and C_d^i are marked with open and solid symbols, respectively. Orifice area: △: 0.4 mm^2; ▽: 1.5 mm^2; ○: 3.5 mm^2; □: 6.0 mm^2; ◇: 9.5 mm^2.

identical to the value reported in the previous study by Zhang *et al.* (2002). For small orifice areas, C_d^{st} is larger than unity.

The glottal width, h, was larger than approximately 0.4 mm for large orifice areas ($A_o>3.5$ mm^2). This area range corresponds to Re>4400 (or $T/4<t<3/4T$) in Fig. 9. Note that the square root of the orifice area, $\sqrt{A_o}$, was used as length scale in the Reynolds number instead of the glottal width. For small orifice widths ($h<0.4$ mm), the rapid increase in C_d^i and C_d^{st} results from a large increase in the contraction coefficient, C_c, since C_L is always smaller than unity (Sec. IV, Appendix A). The large value of C_c can be attributed to the pressure rise and/or frictional loss along the stream from Eq. (4). For converging orifices with small areas, it is believed that frictional losses caused the large C_c values, because the orifice is nozzlelike in shape and the pressure recovery is probably very small. A similar observation was made by Deverge *et al.* (2003), where a viscous correction in Bernoulli's equation was needed for small $h(<0.3$ mm) in the case of a rounded vocal-fold replica. Viscous effects dominate for small areas in both steady and unsteady flows for converging and diverging orifices. Large C_d values were found in all cases (the diverging orifice is discussed in Sec. III B). As the orifice is further closed, the fixed spatial resolution and other factors degrade the accuracy of the ODC estimates for very small areas ($C_d\rightarrow\infty$).

The instantaneous coefficients, C_d^i, are compared to the stationary ones, C_d^{st}, in Fig. 10. The C_d^{st} values are denoted by open symbols, whereas the C_d^i values are denoted using solid symbols. The symbol shapes indicate the area values. The range of values for each symbol on the C_d^i curve indicates experimental accuracy, indicating a ±5% variability of the area values due to image processing errors. Both coefficients are in good agreement at high Reynolds number, specifically for Re>3000, which indicates that there is no significant differences between the steady and unsteady flows for the equivalent values of orifice area and dynamic pressure ($\rho(\mathrm{Re}\cdot\nu)^2/A_o$). The largest relative deviation, $|(\bar{C}_d^{st}-\bar{C}_d^i)/\bar{C}_d^{st}|$, was 12.2% at Re$\sim6800$. Hence, the quasisteady

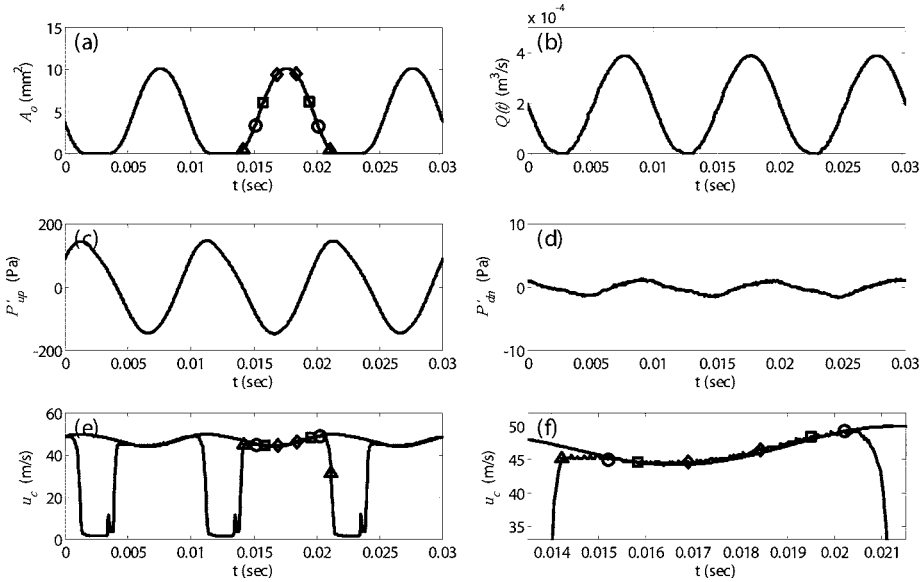


FIG. 11. Time history of flow and acoustic variables (converging jet, $f=100$ Hz, $\Delta p=10$ cm H_2O, $Q_m=1.48\times10^{-4}$ m^3/s). (a) orifice area; (b) flow rate; (c) upstream acoustic pressure; (d) downstream acoustic pressure; (e) centerline velocity at $(x,y,z)=(0,0,1$ mm); and (f) a close-up of centerline velocity: ———: measured; ·····: predicted from Bernoulli's equation. The areas are indicated by the following symbols: △: 0.4 mm^2; ○: 3.5 mm^2; □: 6.0 mm^2; ◇: 9.5 mm^2.

approximation is reasonable. Discrepancies were observed at low Reynolds numbers, for $\mathrm{Re}<3000$. The relative deviation, for example, was 20.5% at $\mathrm{Re}\sim1100$. Small measurement and image processing errors may be in part responsible for the differences between C_d^i and C_d^{st}. This explanation is however not entirely satisfactory because the same procedures were followed for both the steady and the unsteady cases. The difference suggests that the unsteady effects are influential for those conditions. In other words, the viscous effects are still important to both jets for small areas $(C_d^i, C_d^{st}>1)$, while the unsteady effects make a significant contribution to the aerodynamic properties of the unsteady jet $(C_d^i<C_d^{st})$. There are two possible explanations for the observation that $C_d^i<C_d^{st}$ in terms of the unsteady motion of the orifice. If C_c is assumed constant for both jets, a smaller C_L for the unsteady jet suggests stronger viscous diffusion developed in shear layers; the motion of the wall boundary for very small orifice width may induce fluid rotation. The squeezing action of the walls on the fluid may cause an increase in local vorticity, which increases viscous dispersion. On the other hand, if C_L is assumed constant for both jets, a smaller C_c for the unsteady jet implies that stream area at the measurement plane is smaller. This conjecture may be justified from a kinetic point of view since it is plausible that the size of the stream area is affected by the acceleration of the wall. In terms of Eq. (4), the pressure drop may occur due to unsteady motion. Most likely, those two effects may take place simultaneously. These considerations are discussed further in Sec. IV.

The time histories of key hydrodynamic and acoustic variables over several cycles are shown in Fig. 11. The flow rate reached a maximum when the orifice area was a maximum. The upstream acoustic pressure was out of phase with the orifice area (or flow rate) because the pressure was released as the orifice opened, and built up as the orifice closed. The downstream acoustic pressure was relatively small due to the rapid decay of the sound pressure with distance outside the orifice in this case, for which there was no tube or pseudo vocal tract; the sound was radiated in a free field.

Neglecting the unsteady velocity potential term, $\partial\phi/\partial t$, using measured data for the mean transglottal pressure, Δp and the acoustic pressure, $p'(t)$, and assuming that the orifice area is small with respect to the tube area $(A_o/A_t\ll1)$, the centerline velocity was computed from Bernoulli's equation as

$$u_c(t)=\sqrt{\frac{2\Delta p(t)}{\rho}}=\sqrt{\frac{2(\overline{\Delta p}+p'_{\mathrm{up}}(t)-p'_{\mathrm{dn}}(t))}{\rho}}. \tag{8}$$

As shown in Fig. 11(e), there is a good agreement between the measured and the computed center flow velocity for large orifice areas during $T/4<t<3/4T$ (0.015 s$<t<$0.02 s; $A_o>3.5$ mm^2); otherwise, the prediction differs from the measured data. The reason is that Bernoulli's equation cannot be applied because the streamline is no longer inviscid (0.4 mm$^2<A_o<3.5$ mm^2), or no longer existent ($A_o<0.4$ mm^2). These results are consistent with the relationship between the C_d^i and C_d^{st} in Fig. 10. When Bernoulli's equation does not hold (in the presence of unsteady and friction effect), the C_d^i and C_d^{st} values do not agree well. Since the open quotient is approximately 0.8, the quasisteady approximation can thus be considered to hold for about 70 percent of the "duty cycle," or time during which the orifice is open.

Pictures of the steady and unsteady flows are shown in Fig. 12. The unsteady jet [Figs. 12(b), 12(d), and 12(f)] was captured at the instant when the area of the orifice reached that of the stationary orifice for the steady jet [Figs. 12(a), 12(c), and 12(e)]. Despite a difference in flow motion far downstream, Figs. 12(g) and 12(h), both steady and unsteady jets in the vicinity of the orifice exit were nearly identical for large orifice areas. Again, the discrepancies are larger for small areas. Quantitative comparisons using for example particle image velocimetry (PIV) would be very useful to quantify the changes in the velocity field. Some results were obtained using PIV for cases with displacement flow alone $(\Delta p=0)$ and no jet flow. But, the equipment available did not allow the procedure to be used for large velocities due to the insufficient frame rate of the camera.

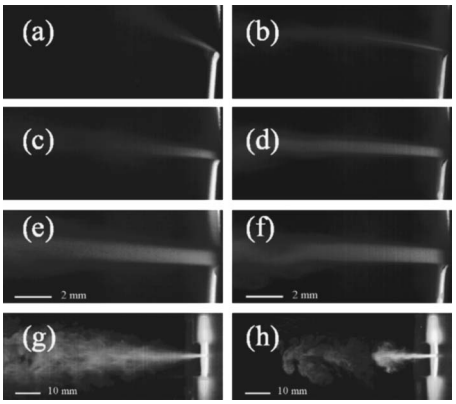


FIG. 12. Pictures of steady and unsteady jets through the convergent orifice for the same orifice area over view plane of $y=0$: $f=100$ Hz, $\Delta p=10$ cm H_2O. Steady jets: (a) $A_o=0.5$ mm^2, (c) $A_o=6$ mm^2, and (e) $A_o=10$ mm^2; unsteady jets: (b) $A_o=0.5$ mm^2, $t\sim 2/10T$, (d) $A_o=6$ mm^2, $t\sim 3/10T$, and (f) $A_o=10$ mm^2, $t\sim 5/10T$. The jet plume is shown, for the cases of (g) steady and (h) unsteady jet ($t\sim 1/4T$).

B. Diverging orifice

A diverging orifice flow has distinct features. Due to a fast transition into the turbulent mixing region downstream, the jet undergoes relatively large fluctuations in velocity. Moreover, the flow is bistable and very sensitive to boundary conditions. Any small upstream or downstream perturbation makes the flow deflect and attach to either of the orifice walls. Figure 13 shows a kymographic image of the flow velocity along the center line. The x location was measured from the center of the orifice. The jet swept around the orifice center early in the opening phase, then attached to the right wall (positive x values) during most of the cycle. Line A-B in Fig. 13 shows the approximate deviation of the jet axis from the orifice centerline. Both steady and unsteady jets through the divergent orifice did tend to attach to the same wall. Asymmetry in the model due to fabrication inaccuracies may have caused the preferential deflection of the jet.

The velocity waveform of the unsteady jet along the center line A-B is shown in Fig. 14. The first peak in the waveform is due to the change in the jet axis orientation. The time of this event corresponds to the leftmost section of line

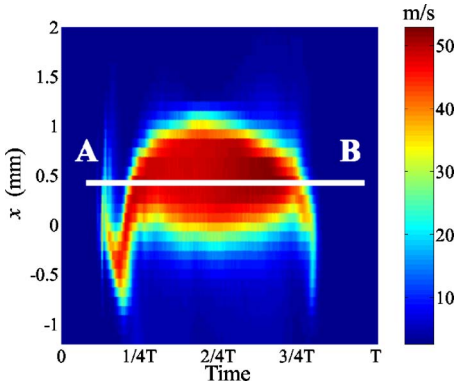


FIG. 13. (Color online) Time history of the unsteady diverging jet at center: $\{u_z(x,y,z,t):-1\leq x\leq 2,\ y=0,\ z=1\ (\text{mm}),\ 0\leq t\leq T\}$. The jet deflects to the right (upward) side due to the Coanda effect: $f=100$ Hz, $\Delta p=14$ cm H_2O, $Q_m=3.06\times 10^{-4}$ m^3/s. Line AB denotes the jet core during wall attachment.

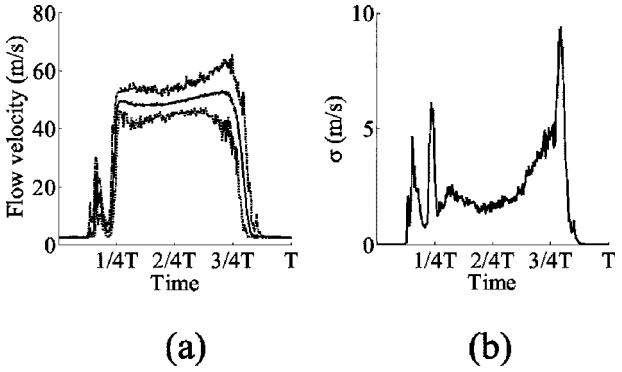


FIG. 14. Velocity vs time (along line AB in Fig. 13): diverging orifice, $f=100$ Hz, $\Delta p=14$ cm H_2O, $Q_m=3.06\times 10^{-4}$ m^3/s. (a) measured flow velocity: –.–.–.: maximum value; - - -: minimum value; ———: ensemble average. (b) standard deviation of the ensembles.

A-B in Fig. 13. The peak velocity envelope shows that the largest magnitude in velocity fluctuations is about 20% of the maximum velocity. The variance of the fluctuations remained small during most of the cycle. The largest variations occurred at both ends of the cycle where the contribution to flow rate is less significant due to the small area. The velocity waveform was stationary, as for the converging jet. Ensemble-averaged velocity waveforms were thus used for flow rate calculations.

The jet was observed to start generally straight with a slight sweeping motion. It deflected from the axial direction after approximately one quarter of one period. A similar behavior was also observed in previous studies (Hofmans, 2003; Erath and Plesniak, 2006). Since only the axial component of velocity contributes to the flow rate (note that the orifice area is normal to the axial direction as well), the measured flow velocities were corrected to obtain the axial component by assuming that the jet deflection angle, $\alpha/2$, was the same as the wall angle, as described in Sec. II. The change in flow direction caused a slightly asymmetric flow rate waveform, as shown in Fig. 15. The waveform of the

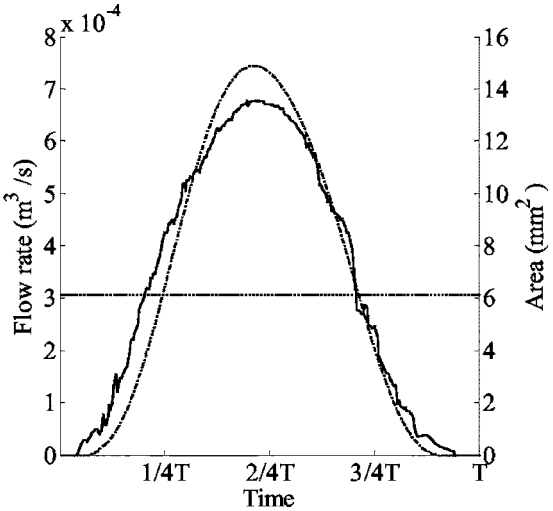


FIG. 15. Instantaneous flow rate and area function vs time (diverging jet, $f=100$ Hz, $\Delta p=14$ cm H_2O, $Q_m=3.06\times 10^{-4}$ m^3/s, $\overline{Q(t)}=3.20\times 10^{-4}$ m^3/s): ———: $Q(t)$; –.–: $A_o(t)$. The estimation error from the measured time-mean flow rate, Q_m, is about 4%.

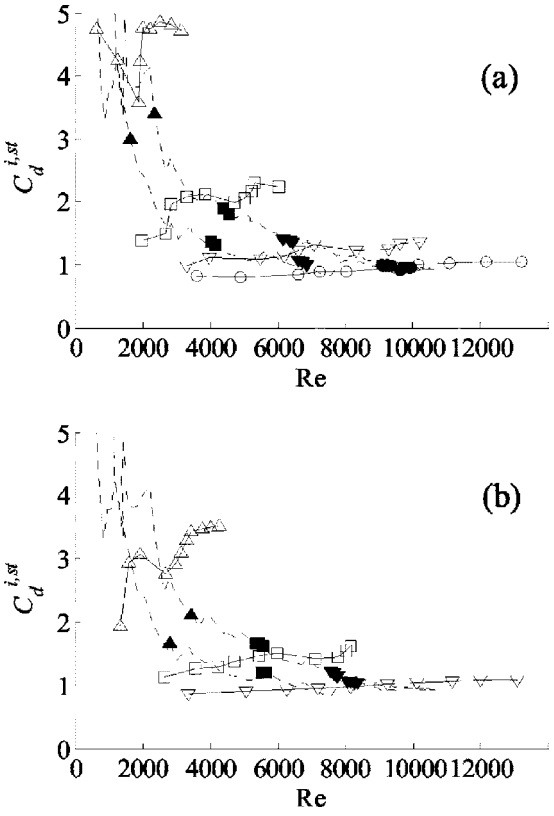


FIG. 16. Comparison between orifice discharge coefficients for unsteady diverging jet and those for steady diverging jets at several opening areas. C_d^{st} and C_d^i are marked with open and solid symbols, respectively ($f=100$ Hz, $\Delta p=14$ cm H_2O, $Q_m=3.06\times10^{-4}$ m^3/s). Orifice area: (a) △: 0.9 mm^2; □: 2.5 mm^2; ▽: 5.0 mm^2; ○: 11.5 mm^2; and (b) △: 1.5 mm^2; □: 3.7 mm^2; ▽: 7.6 mm^2.

instantaneous flow rate is noisy, but the time-averaged value agreed with that from the inlet flow meter within around 4% estimation error, as for the case of the converging orifice.

The C_d^i and C_d^{st} values are shown in Fig. 16. There is a difference between the C_d^i vs Reynolds number curves for the opening and the closing phases. Such differences are generally referred to as hysteretic. Interestingly, such hysteresis probably means that net power is transferred from the wall motion to the flow (or vice versa), although the shape of the orifice during the opening and closing phases is nominally identical. For $\mathrm{Re}<6000$, the coefficient during the opening phase is larger than that during the closing phase, whereas the coefficient is almost identical for $\mathrm{Re}>6000$. The period for $\mathrm{Re}>6000$ corresponds approximately to $T/4<t<3/4T$ or $A_o>4.0$ mm^2 (Fig. 15). The difference in C_d^i may be associated with the Coanda effect since the jet attached to the wall for $t>T/4$. One plausible reason for this difference is the influence of the wall motion. The wall motion may cause a suction, which initiates the Coanda effect (Prandtl and Tietjens, 1957). It is also hypothesized that the asymmetric jet development in the orifice channel due to the wall attachment (wall boundary layer on the attached side and free shear layer on the other) may cause the differences in C_d^i. In other words, the jet plume is not perfectly identical during the opening and closing half-cycles.

The C_d^i values for the diverging orifice are generally greater than those for the converging orifice. This is due primarily to an increase in C_c related to a pressure recovery within the diffuserlike orifice channel, with the C_L value being almost the same regardless of jet type. The C_c and C_L values are further discussed in Sec. IV.

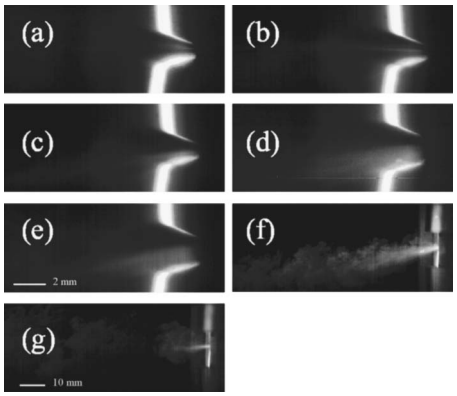


FIG. 17. Pictures of steady and unsteady jets through the divergent orifice for the same orifice area over view plane of $y=0$: $f=100$ Hz, $\Delta p=14$ cm H_2O. Steady jets: (a) $A_o=2$ mm^2 and (d) $A_o=11$ mm^2; unsteady jets: (b) $A_o=2$ mm^2, $t\sim1/4T$, (c) $A_o=2$ mm^2, $t\sim3/4T$, (e) $A_o=11$ mm^2, $t\sim2/4T$. The jet plume is shown, for the cases of (f) steady and (g) unsteady jet ($t\sim3/8T$).

The C_d^i values asymptotically converge for large orifice areas, and they are relatively large for small orifice areas. The C_d^i and C_d^{st} values are in good agreement for $\mathrm{Re}>6000$, where both steady and unsteady jets were attached to the orifice wall. In an intermediate range, $4000<\mathrm{Re}<6000$, the steady jet was bistable and attached to the orifice wall intermittently. The C_d^{st} values were computed in that case by averaging the ODCs for each flow pattern. They were found to be between the C_d^i values for the opening and the closing phase. Although such bistable behavior precludes a conclusive analysis, the coefficients seemed to be in agreement in the intermediate Reynolds number region. Significant discrepancies were again observed at low Reynolds numbers. As in the case of the converging jet, C_d^{st} values are greater than C_d^i, which suggests influences of the unsteady effects due to wall motion associated with small orifice widths. Pictures of the flow through the diverging orifice are shown in Fig. 17. The unsteady jets for small orifice areas, shown in Figs. 17(b) and 17(c), were initially straight and then attached to one wall. Pictures of steady jets for the corresponding area are shown in Fig. 17(a). For a large orifice area, the discharge patterns of the steady [Fig. 17(d)] and the unsteady jets [Fig. 17(e)] are almost identical near the orifice despite the different jet plume patterns downstream [Figs. 17(f) and 17(g)].

IV. DISCUSSION

The improvement in accuracy related to the inclusion of a time-dependent ODC in the pressure-flow relation of the orifice was illustrated by considering a pulsating orifice of area, $A_o(t)$, in an infinite tube. The radiated sound pressures estimated using a constant ODC value, C_d, and that using a time-varying ODC, $C_d(t)$, are the variables to compare. A monopole source representation was adopted from the previous work of Zhang *et al.* (2002) for the calculations, where

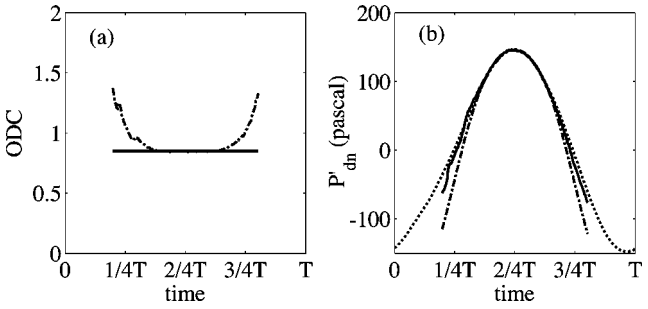


FIG. 18. Illustration of the influence of employing time-varying orifice discharge coefficients: (a) ———: stationary coefficient, C_d, –.–.–.: instantaneous coefficient, $C_d(t)$; (b) - - -: measured radiated pressure, –.–.–.: radiate pressure computed with C_d, ——: radiated pressure computed with $C_d(t)$. Converging jet confined in a downstream tube, $f=100$ Hz, $\Delta p=10$ cm H_2O, $Q_m=1.48\times10^{-4}$ m^3/s.

the source region was idealized as two identical but oppositely phased one-dimensional monopoles. In such an idealized case, the radiated acoustic pressure, $p'_{dn}(t)$, can be expressed in terms of the ODC, the orifice area, the mean pressure differential, $\Delta\bar{p}$, and the flow rate, Q_m, as

$$p'_{dn}(t)=\frac{\rho c}{A_t}\left\{-\left(\frac{2c}{A_t}\right)(C_d(t)A_o(t))^2+C_d(t)A_o(t)\times\sqrt{\frac{4c}{A_t}Q_m+\frac{2}{\rho}\Delta\bar{p}+\left(\frac{2c}{A_t}C_d(t)A_o(t)\right)^2}-Q_m\right\}$$

$$=f(C_d(t),A_o(t),\Delta\bar{p},Q_m), \tag{9}$$

where ρ is the medium density, and A_t is the tube cross-sectional area (see Appendix B).

Figure 18 shows a comparison between predicted and measured downstream acoustic pressures for the converging orifice with an anechoically terminated downstream duct. Instantaneous ODC data shown in Fig. 18(a) were used in Eq. (9). The resulting downstream acoustic pressures show that the evaluation using $C_d(t)$ improves the radiated pressure estimation by as much as 20% over the transient phase of the duty cycle. Although such a correction does not affect the peak pressure value, it does affect the rate of change of the pressure during glottal opening and closing, altering the harmonic content of the radiated pressure at higher frequency.

The values of C_c and C_L can be calculated numerically from the integration of the flow area in the measurement plane, and from the instantaneous ODC data, respectively. The instantaneous C_L time histories for converging and diverging jets are shown in Fig. 19. The C_L curves for both jets are symmetric, and the loss pattern seems to be little influenced by the presence of Coanda effect in the case of the diverging jet [the difference is of the order $O(10^{-2})$ between $t<T/4$ and $t>3/4T$]. The C_L pattern is similar between the converging and the diverging orifice, but the loss undergoes a larger change in the case of the converging jet. Turbulence developed downstream of the diverging jet may lead to more uniform viscous diffusion, and hence may induce a larger head loss (overall lower C_L) with a smaller change or excursion over one cycle.

The instantaneous C_c curves for the converging and diverging jets are shown in Fig. 20. For the converging jet at a

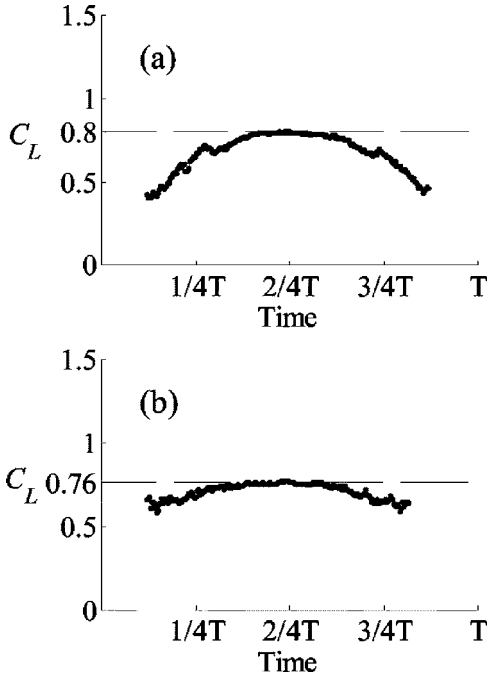


FIG. 19. Instantaneous loss coefficient vs time for (a) the unsteady converging and (b) the unsteady diverging jet over one cycle.

large orifice area ($1/4T<t<3/4T$), the C_c value is slightly greater than unity, i.e., $C_c\geq1.07$. This may be because the flow separation point is not exactly located at the throat, but slightly downstream due to the rounded edge. Hence, there is a small increase in the pressure (pressure recovery) along the streamline. For the diverging jet case, the pressure recovery is much higher ($C_c\geq1.24$). For small areas in both cases ($t<1/4T, t>3/4T$), the curve increases dramatically, and little difference is observed between converging and diverging jets, which suggests that the frictional loss along the stream overwhelms the pressure rise in the same way for both orifice profiles.

The effect of orifice wall motion is significant when the orifice area is small, as shown by the discrepancies between C_d^i and C_d^{st}. The volume displaced by the pumping action of the orifice wall oscillating motion in absence of flow is referred to as *displacement flow*. The axial volume velocity induced by the displacement flow together with the wall normal velocity may induce bound vorticity, presumably around

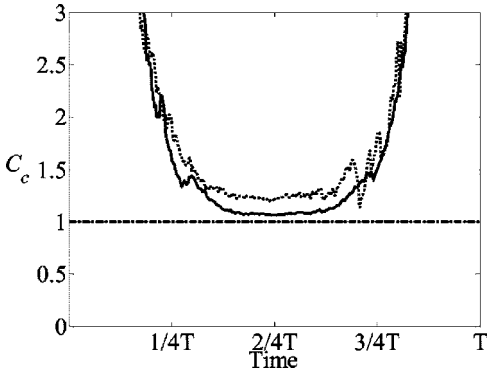


FIG. 20. Instantaneous contraction coefficient vs time: ———: converging jet; - - -: diverging jet; –.–.–.: reference line of value 1.

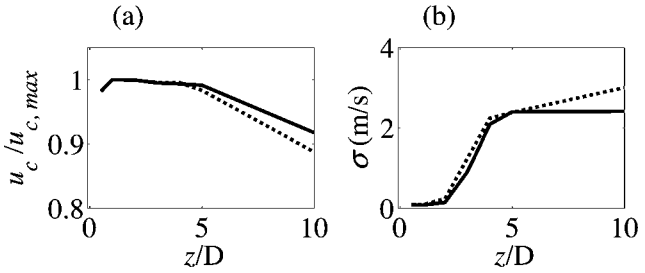


FIG. 21. Jet center velocity as a function of streamwise distance from orifice exit: (a) normalized center velocity and (b) velocity standard deviation: converging orifice, $f=0$ Hz (steady jet), $\Delta p=10$ cm H_2O, ———: no confinement downstream; - - -: confined by a rigid tube downstream.

one boundary layer thickness from the wall. This local unsteady fluid behavior may in part cause the orifice coefficient difference between steady and unsteady flows. When the orifice area is large, this effect may not be significant due to the relative importance of the pressure-driven velocity fields. Wall pressure and transglottal flow field measurements are needed to confirm this hypothesis.

Although the orifices have a rounded edge at their exit, *vena contracta* was observed. The centerline velocity was measured as a function of the streamwise distance from the converging orifice as shown in Fig. 21. The results include two discharge configurations of an open jet and a confined jet. A rigid uniform tube having the same dimensions of the upstream tube was used for the downstream confinement. The centerline velocity for both cases increased and reached a maximum at $z\sim 1D$, which should correspond to the minimum stream area based on the assumption of an incompressible stream tube. The presence of the downstream confinement had no influence on the location of the *vena contracta*. Beyond this point, the velocity decreased faster with z in the case of the confined jet due to increased turbulent diffusion. This is further substantiated by the observed increase in fluctuation amplitudes, which are shown in the form of a standard deviation in Fig. 21(b). This *vena contracta* may be due to the elliptical shape of the orifice. Tip vortices at the commissures may cause local flow acceleration downstream. This phenomenon was observed in previous studies, along with an axis switching of the elliptical jet plume (Zhang *et al.*, 2004).

In previous studies (Zhang *et al.*, 2002), the instantaneous flow rate was calculated from the instantaneous radiated sound-pressure data because no flow map was available. Because the ducts were not perfectly anechoic, an acoustic model was needed to describe the sound field, and inverse filtering of the sound-pressure signals was required in order to calculate the glottal flow waveform. Due to difficulties in the robust estimation of the flow rate, a low-pass filter was used to stabilize the numerical scheme. This process may have resulted in numerical errors in the flow rate estimates. One benefit of the direct mapping method utilized in the present study is that such inverse filtering is no longer needed, and accurate flow rate estimates can be obtained more easily for arbitrary acoustic loading, including cases with highly reflecting tube configurations.

The use of a single wire probe for the velocity measurements yields sufficient accuracy in the present study because of the relatively symmetric orifice geometry. For more physiologically realistic configurations, e.g., self-oscillated models (Thomson *et al.*, 2005) and the models with false vocal folds (Triep *et al.*, 2005), direct measurement of the glottal flow close to the orifice may be impeded by a strong out-of-plane axial (inferior/superior or "bulging") folds motion or by geometric restrictions. In such cases, the measurement plane may need to be relocated with the use of cross-wire or slanted wire probes. A signal filtering of the measured velocity may also be required in order to remove highly turbulent components in the data.

V. CONCLUSIONS

The accuracy of the quasisteady approximation often utilized in voice production models was evaluated. The orifice discharge coefficients of comparable unsteady and steady jets through convergent and divergent glottis-shaped orifices were directly measured using hot-wire anemometry. The experiments confirmed that the approximation is valid within reasonable accuracy over 70% of the glottal duty cycle. Bernoulli's steady streamline theory was found to be accurate for the prediction of the unsteady flow over the same range. Unsteady and viscous effects were found to be significant only during a short time interval near opening and closure. The flows through the diverging orifices were complex, and featured significant hysteretic effects. More work is still needed to clarify the relative contributions of viscous effects and displacement flow over the very low Reynolds number range.

ACKNOWLEDGMENTS

This study was supported by Research Grants No. R01 DC03577 and R01 DC 05788 from the National Institute of Deafness and Other Communication Disorders, National Institutes of Health. Thanks are expressed to the Herrick Laboratories technical staff for their technical support, and to Dr. Ronald Scherer and Dr. Fariborz Alipour for their pertinent comments and suggestions. Finally, the continuous help and support of our colleagues Byron Erath, Jungsoo Suh, Michael Plesniak, Steven Frankel, Bogdan Kucinschi, Scott Thomson, and many others are gratefully acknowledged.

List of Symbols

A = cross sectional area of flow stream tube (mm^2)
A_t = tube area (mm^2)
A_o = orifice area (mm^2)
C_c = area contraction coefficient
C_L = loss coefficient
C_d = orifice discharge coefficient
C_d^{st} = stationary orifice discharge coefficient for steady jets
C_d^i = instantaneous orifice discharge coefficient for unsteady jets
$Q(t)$ = time-varying flow rate (m^3/s)

Q_{ideal} = flow rate without loss (m^3/s)
Q_m = flow rate measured far upstream using flow meter (m^3/s)
p = static pressure (Pa)
p' = acoustic pressure (Pa)
Δp = pressure differential; transglottal pressure (Pa)
Re = Reynolds number, $u_c\sqrt{A_o}/\nu$
t = time (s)
T = period of one cycle (s)
$u(\mathbf{x},t)$ = flow velocity profile measured on the measurement plane (m/s)
u_{cut} = cutoff flow velocity (m/s)
α = included angle of orifice (°)
ρ = air density (kg/m^3)
σ = standard deviation of flow velocity (m/s)

Subscripts

1 = measuring point 1
2 = measuring point 2
c = jet centerline
dn = downstream
up = upstream
x = lateral direction
y = spanwise direction
z = streamwise direction

Acronym

ODC = orifice discharge coefficient

APPENDIX A: ORIFICE DISCHARGE COEFFICIENT DEFINITION

Consider an infinitely long tube interrupted by an orifice of area, A_o, as shown in Fig. 1, where p, u, and A denote the static pressure, the flow velocity, and the cross-sectional area of stream tube at point 1 and 2. For incompressible, frictionless flow in the duct, the center flow velocity at the orifice throat (at minimum area) can be computed as

$$u_o = \frac{1}{\sqrt{1-(A_o/A_1)^2}}\sqrt{\frac{2(p_1-p_o)}{\rho}}. \tag{A1}$$

The orifice discharge coefficient is defined as the ratio of the actual (real) flow rate to the ideal flow rate (center velocity times area),

$$Q_{real} = C_d \cdot Q_{ideal} = C_d \cdot A_o \cdot u_o \tag{A2}$$

$$\therefore Q_{real} = \frac{C_d \cdot A_o}{\sqrt{1-(A_o/A_1)^2}}\sqrt{\frac{2(p_1-p_o)}{\rho}}. \tag{A3}$$

The downstream pressure, p_2, is usually measured instead of the orifice throat pressure, and Eq. (A3) becomes, in terms of p_2,

$$Q_{real} = \frac{C_d^* \cdot A_o}{\sqrt{1-(A_o/A_1)^2}}\sqrt{\frac{2(p_1-p_2)}{\rho}}. \tag{A4}$$

Note that $C_d = C_d^*$, where it is assumed $p_o \sim p_2$ for orifice plates. The spatially averaged flow velocity through the orifice is then computed as

$$u_{o,real} = \frac{C_d^*}{\sqrt{1-(A_o/A_1)^2}}\sqrt{\frac{2(p_1-p_2)}{\rho}} \tag{A5}$$

from $Q_{real} = A_o \cdot u_{o,real}$ (Blevins, 1984; Munson *et al.*, 1993).

To avoid any confusion, the orifice discharge coefficient is defined as in Eq. (A4) rather than in Eq. (A3) because it is practically based on the variables measured downstream. The ratio of the real flow rate to the ideal one will be denoted by C_L instead of C_d from Eq. (A2) as

$$Q_{real} = C_L \cdot Q_{ideal}, \tag{A6}$$

and C_d^* will be replaced with C_d for the orifice coefficient herein after. The real flow rate can now be expressed from Eq. (A4) in terms of flow velocity measured downstream as

$$\begin{aligned} Q_{real} &= \frac{C_d \cdot A_o}{\sqrt{1-(A_o/A_1)^2}}\sqrt{\frac{2(p_1-p_2)}{\rho}} \\ &= C_d \cdot A_o \frac{\sqrt{1-(A_2/A_1)^2}}{\sqrt{1-(A_o/A_1)^2}} \cdot \frac{1}{\sqrt{1-(A_2/A_1)^2}} \\ &\quad \times \sqrt{\frac{2(p_1-p_2)}{\rho}} \\ &= C_d \cdot A_o \frac{\sqrt{1-(A_2/A_1)^2}}{\sqrt{1-(A_o/A_1)^2}} \cdot u_2. \end{aligned} \tag{A7}$$

Because the jet stream area as well as the orifice area is small with respect to the tube area ($A_o, A_2 \ll A_1$), the real flow rate further reduces to

$$Q_{real} = C_d \cdot A_o \cdot u_2. \tag{A8}$$

The orifice discharge coefficient can be decomposed through the introduction of a contraction coefficient, C_c. Because the flow is assumed incompressible, the ideal flow rate is always constant,

$$Q_{ideal} = A_o \cdot u_o = A_2 \cdot u_2, \tag{A9}$$

and, therefore,

$$u_o = A_2/A_o \cdot u_2 = C_C \cdot u_2, \tag{A10}$$

where C_C is defined as the ratio of flow stream area to the orifice area. Therefore, from Eqs. (A6) and (A8)–(A10),

$$C_d = C_C \cdot C_L \quad (C_L \leq 1). \tag{A11}$$

Note that the loss coefficient, C_L, is always smaller than unity because the real flow rate cannot exceed the ideal flow rate. The contraction coefficient can be shown from Bernoulli's equation to indicate the pressure head variation along the streamline. If one takes the Bernoulli's equation with frictional loss pressure head, Φ, along the stream line from the orifice throat to the measurement location,

$$\frac{1}{2}u_0^2 + \frac{p_o}{\rho} = \frac{1}{2}u_2^2 + \frac{p_2}{\rho} + \frac{\Phi}{\rho}, \tag{A12}$$

the contraction coefficient can be rewritten from Eqs. (A10) and (A12) as

$$C_c^2 = 1 + \frac{p_2 - p_o}{\frac{\rho}{2}u_2^2} + \frac{\Phi}{\frac{\rho}{2}u_2^2}. \tag{A13}$$

It should be pointed out that Eq. (A13) is almost identical to the relationship for diffuser performance that can be found in other references (Blevins, 1984),

$$K = 1 - (A_o/A_2)^2 - C_P, \tag{A14}$$

where K is loss coefficient and C_P is static pressure recovery, except that (A14) is normalized by the dynamic pressure head at the diffuser throat.

Therefore, from Eqs. (A11) and (A13), the orifice coefficient is bounded by

$$C_d^2 \leq 1 + \frac{p_2 - p_o}{\frac{\rho}{2}u_2^2} + \frac{\Phi}{\frac{\rho}{2}u_2^2}. \tag{A15}$$

For flows in converging orifices, where the pressure rise (except for a possible pressure drop at the *vena contracta*) and viscous effects are insignificant, C_d is always smaller than unity. For flow through a diverging orifice (with significant pressure recovery) or through a very small opening (with strong viscous frictional losses), C_d can be larger than unity. The relative contribution of the reversible (pressure recovery) and irreversible (viscous loss) processes may not be discriminated from the orifice coefficient value alone, unless either p_o or Φ is available.

APPENDIX B: RADIATED SOUND PRESSURE IN TERMS OF C_D

This section presents the details of a simple model for the sound pressure radiated by flow through a time-varying orifice in an infinitely long uniform tube. First, rewrite the definition of the orifice discharge coefficient with $u_2 = u_c$ from Eq. (A8) as

$$Q(t) = C_d(t)A_o(t)u_c(t), \tag{B1}$$

and express the centerline velocity from Eq. (8) as

$$\frac{1}{2}\rho u_c(t)^2 = \Delta\bar{p} + p'_{\mathrm{up}}(t) - p'_{\mathrm{dn}}(t). \tag{B2}$$

The fluctuating velocity is related to flow rate through a simple conservation of mass argument,

$$u'_{\mathrm{dn}}(t) = -u'_{\mathrm{up}}(t) = \frac{1}{A_t}\{Q(t) - Q_m\}. \tag{B3}$$

Note that $\overline{Q(t)} = Q_m$ within a tolerable error (4%), as discussed in Sec. III. Assuming infinitely long tubes, the acoustic impedance of the tube is

$$\frac{p'_{\mathrm{up}}(t)}{u'_{\mathrm{up}}(t)} = \frac{p'_{\mathrm{dn}}(t)}{u'_{\mathrm{dn}}(t)} = \rho c, \tag{B4}$$

where ρ is the density, p'_{up} and p'_{dn} are the acoustic pressures upstream and downstream, respectively. By substitution of (B4) into (B3), the acoustic pressure can be expressed as

$$p'_{\mathrm{dn}}(t) = -p'_{\mathrm{up}}(t) = \frac{\rho c}{A_t}\{Q(t) - Q_m\}. \tag{B5}$$

Then, a quadratic equation for the jet centerline velocity can be obtained from Eq. (B2) by using (B1) and (B5),

$$\frac{1}{2}\rho u_c(t)^2 + \frac{2\rho c}{A_t}C_d(t)A_o(t)u_c(t) = \Delta\bar{p} + \frac{2\rho c}{A_t}Q_m, \tag{B6}$$

which yields

$$u_c(t) = -\frac{2c}{A_t}C_d(t)A_o(t) + \sqrt{\frac{4c}{A_t}Q_m + \frac{2}{\rho}\Delta\bar{p} + \left(\frac{2c}{A_t}C_d(t)A_o(t)\right)^2}. \tag{B7}$$

Note that $u_c(t)$ is chosen positive since the mean flow is driven by a positive mean pressure differential across the orifice. Finally, the propagating acoustic pressure downstream, $p'_{\mathrm{dn}}(t)$, is obtained from (B1), (B5), and (B7),

$$\begin{aligned} p'_{\mathrm{dn}}(t) &= \frac{\rho c}{A_t}\{C_d(t)A_o(t)u_c(t) - Q_m\} \\ &= \frac{\rho c}{A_t}\left\{-\left(\frac{2c}{A_t}\right)(C_d(t)A_o(t))^2 + C_d(t)A_o(t) \right. \\ &\quad \left. \times \sqrt{\frac{4c}{A_t}Q_m + \frac{2}{\rho}\Delta\bar{p} + \left(\frac{2c}{A_t}C_d(t)A_o(t)\right)^2} - Q_m\right\} \\ &= f(C_d(t), A_o(t), \Delta\bar{p}, Q_m). \end{aligned} \tag{B8}$$

[1]The single hot-wire probe primarily measures the axial flow velocity.

[2]The effective velocity, U_e, measured by a hot wire is related to each fluid velocity component by $U_e^2 \sim U_n^2 + \alpha^2 \cdot U_b^2 + \beta^2 \cdot U_t^2$, where n, b, and t denote normal, binormal, and tangential directions to the hot wire, α and β are directional sensitivity constants, generally about 1.05 and 0.2, respectively.

Barney, A., Shadle, C. H., and Davies, P. O. A. L. (**1999**). "Fluid flow in a dynamic mechanical model of the vocal folds and tract. I. Measurements and theory," J. Acoust. Soc. Am. **105**(1), 444–455.

Blevins, R. D. (**1984**). *Applied Fluid Dynamics Handbook* (Van Nostrand Reinhold, New York).

Bruun, H. H. (**1995**). *Hot-wire Anemometry: Principles and Signal Analysis* (Oxford University Press, New York).

Deverge, M., Pelorson, X., Villain, C., Lagrée, P., Chentouf, F., Willems, J., and Hirschberg, A. (**2003**). "Influence of collision on the flow through *in vitro* rigid models of the vocal folds," J. Acoust. Soc. Am. **114**(6), 3354–3362.

Erath, B. D., and Plesniak, M. W. (**2006**). "An investigation of bimodal jet trajectory in flow through scaled models of the human vocal tract," Exp. Fluids **40**(5), 683–696.

Fant, G. (**1960**). *Acoustic Theory of Speech Production* (Mouton, the Hague).

Flanagan, J. L. (**1965**). *Speech Analysis, Synthesis, and Perception* (Springer, Berlin).

Gauffin, J., Binh, N., Ananthapadmanabha, T. V., and Fant, G. (**1983**). "Glottal geometry and volume velocity waveform," in *Vocal Fold Physiology: Contemporary Research and Clinical Issues*, edited by D. M. Bless and J. H. Abbs (College-Hill, San Diego), pp. 194–201.

Hofmans, G. C., Groot, G., Ranucci, M., Graziani, G., and Hirschberg, A. (**2003**). "Unsteady flow through *in vitro* models of the glottis," J. Acoust. Soc. Am. **113**(3), 1658–1675.

Iguchi, M., Yamazaki, H., Yamada, E., and Morita, Z. (**1990**). "Velocity and turbulence intensity in a pulsating jet through a sudden expansion," Trans. Jpn. Soc. Mech. Eng., Ser. A **56**, 1659–1664.

McGowan, R. S. (**1988**). "An aeroacoustic approach to phonation," J.

Acoust. Soc. Am. **83**(2), 696–704.
Mongeau, L., Franchek, N., Coker, C. H., and Kubli, R. A. (**1997**). "Characteristics of a pulsating jet through a small modulated orifice, with application to voice production," J. Acoust. Soc. Am. **102**(2), 1121–1133.
Munson, B. R., Young, D. F., and Okiishi, T. H. (**1993**). *Fundamentals of Fluid Mechanics*, 2nd ed. (Wiley, New York).
Pelorson, X., Hirschberg, A., van Hassel, R. R., Wijnands, A. P. J., and Auregan, Y. (**1994**). "Theoretical and experimental study of quasi-steady flow separation within the glottis during phonation. Application to a modified two-mass model," J. Acoust. Soc. Am. **96**, 3416–3431.
Prandtl, L., and Tietjens, O. G. (**1957**). *Applied Hydro- and Aeromechanics* (Dover, New York), p. 294.
Scherer, R. C., Witt, D., and Kucinschi, B. R. (**2001a**). "The effect of exit radii on intraglottal pressure distributions in the convergent glottis," J. Acoust. Soc. Am. **110**, 2267–2269.
Scherer, R. C., Shinwari, D., De Witt, K. J., Zhang, C., Kucinschi, B. R., and Afjeh, A. A. (**2001b**). "Intraglottal profiles for a symmetric and oblique glottis with a divergence angle of 10 degrees," J. Acoust. Soc. Am. **109**(4), 1616–1630.
Thomson, S. L., Mongeau, L., and Frankel, S. H. (**2005**). "Aerodynamic transfer of energy to the vocal folds," J. Acoust. Soc. Am. **118**, 1689–1700.
Titze, I. R. (**2000**). *Principles of Voice Production* (National Center for Voice and Speech, Denver).
Triep, M., Brücker, Ch., and Schröder, W. (**2005**). "High-speed PIV measurements of the flow downstream of a dynamic mechanical model of the human vocal folds," Exp. Fluids **39**, 232–245.
Vilain, C. E., Pelorson, X., Fraysse, C., Deverge, M., Hirschberg, A., and Willems, A. (**2004**). "Experimental validation of a quasi-steady theory for the flow through the glottis," J. Sound Vib. **276**, 475–490.
Zhang, Z., Mongeau, L., and Frankel, S. (**2002**). "Experimental verification of the quasisteady approximation for aerodynamic sound generation by jets in tubes," J. Acoust. Soc. Am. **112**(4), 1652–1663.
Zhang, Z., Mongeau, L., Frankel, S., Thomson, S. L., and Park, J. B. (**2004**). "Sound generation by steady flow through glottis-shaped orifices," J. Acoust. Soc. Am. **116**(3), 1720–1728.
Zhao, W., Zhang, C., Frankel, S., and Mongeau, L. (**2002**). "Computational aeroacoustics of phonation. I. Computational methods and sound generation mechanisms," J. Acoust. Soc. Am. **112**(5), 2134–2146.

Generation of the vocal tract spectrum from the underlying articulatory mechanism

Tokihiko Kaburagi
Department of Acoustic Design, Faculty of Design, Kyushu University, 4-9-1 Shiobaru, Minami-ku, Fukuoka, 815-8540 Japan

Jiji Kim
NTT Communications Corporation, 26-1 Sakuragaoka-cho, Shibuya-ku, Tokyo, 150-8512 Japan

(Received 2 April 2006; revised 8 October 2006; accepted 9 October 2006)

A method for synthesizing vocal-tract spectra from phoneme sequences by mimicking the speech production process of humans is presented. The model consists of four main processes and is particularly characterized by an adaptive formation of articulatory movements. First, our model determines the time when each phoneme is articulated. Next, it generates articulatory constraints that must be met for the production of each phoneme, and then it generates trajectories of the articulatory movements that satisfy the constraints. Finally, the time sequence of spectra is estimated from the produced articulatory trajectories. The articulatory constraint of each phoneme does not change with the phonemic context, but the contextual variability of speech is reproduced because of the dynamic articulatory model. The accuracy of the synthesis model was evaluated using data collected by the simultaneous measurement of speech and articulatory movements. The accuracy of the phonemic timing estimates were measured and compared the synthesized results to the measured results. Experimental results showed that the model captured the contextual variability of both the articulatory movements and speech acoustics.
[DOI: 10.1121/1.2384847]

PACS number(s): 43.70.Bk, 43.72.Ja [BHS] Pages: 456–468

I. INTRODUCTION

Speech synthesis methods that mimic processes involved in the production of speech have been widely studied. For example, Coker and Fujimura (1966), Mermelstein (1972), and Engwall (1999) have proposed articulatory models that control the shape of the vocal tract by varying the values of a small number of parameters specifying the positions of articulators. The transfer function of the vocal tract is then determined using an acoustic tube model (Flangan, 1972; Maeda, 1982; Sondhi and Schroeter, 1987), and in combination with the glottal and noise source models, the synthesis model produces speech signals that can sound natural. Since the articulatory positions can be measured by various observation techniques (Rokkaku *et al.*, 1986; Perkell *et al.*, 1992; Stone, 1990), it is possible to determine the accuracy of the model position estimates. Such articulatory measurements are also useful for the dynamic control of articulatory parameters and speech synthesis of continuous utterances.

To accurately synthesize continuous speech, the effects of coarticulation needs to be modeled (Sharf and Ohde, 1981). Coarticulation is a source of articulatory variability because it causes the interaction of neighboring phonemes and temporal spreading of phonemic articulatory features. Therefore, speech scientists have performed physiological (MacNeilage and DeClerk, 1969), kinematic (Gay, 1977; Farnetani, 1989), and acoustic (Stevens and House, 1963; Öhman, 1966; Blumstein, 1986) investigations to clarify the influence of the phonemic context on the attainment of phonemic articulatory targets. Contextual variability is also one of the central problems in the field of speech synthesis and recognition technologies. To solve this problem, speech units are usually constructed on the basis of context-dependent frameworks such as the digram and trigram models (Jelinek, 1997).

To cope with the problem of contextual variability, dynamic articulatory models have been proposed (Saltzman and Munhall, 1989; Kaburagi and Honda, 1996) based on the idea that articulatory movements are generated so that phonemic articulatory targets are satisfied. Although the targets are universal and context independent, these models are capable of reproducing articulatory variability by blending target gestures (Saltzman and Munhall, 1989) or by a reduction in the dimension of the task space (Kaburagi and Honda, 1996). Generally, the dimension of the task space becomes smaller than that of the articulatory variables because of the coordinated structure of the articulators (Saltzman 1979) and the existence of articulator priority (Coker, 1976), and then the kinematic states of the articulators are only partly constrained by each phonemic task. In the latter model, it was assumed that the unconstrained degrees of freedom of the articulatory variables are resolved by minimizing an energy criterion so that the tasks of the adjacent phonemes are smoothly interpolated. Then, the values of redundant degrees of freedom are determined depending on the phonemic context, and the model can reproduce the contextual effects.

This paper presents a speech synthesis model which generates the vocal-tract spectrum on the basis of articulatory dynamics. Given a sequence of phoneme symbols, our model first specifies articulatory timings and target vocal-tract ges-

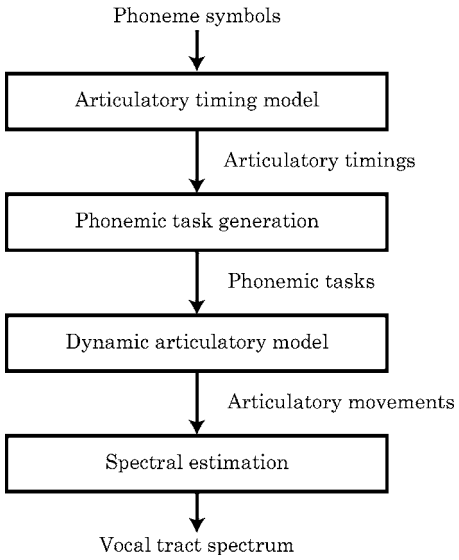


FIG. 1. The framework of the spectrum generation model. The phonemic information is first used to specify the articulatory timings and a sequence of phonemic tasks. Continuous movements of the articulatory organs are then formed by a dynamic articulatory model. Finally, the vocal-tract spectrum is estimated from the articulatory movements on a frame-by-frame basis using a data search method.

tures using invariant features (Kaburagi and Honda, 2001). Then articulatory movements are generated by the dynamic model (Kaburagi and Honda, 1996). Finally, the vocal-tract spectrum is estimated by searching a data set composed of articulatory and acoustic data pairs. Unlike typical speech synthesis methods, the temporal pattern of speech is formed as continuous articulatory movements instead of a concatenation of inflexible segmental units. Also, contextual variability is automatically reproduced by using a compact set of phonemic tasks. As such, our synthesis model is constructed by mimicking the actual process of human speech production. In addition, an electromagnetic articulography (EMA) is used to measure the positions of the articulators (Kaburagi and Honda, 1994, 2002). These articulatory data are used to construct the synthesis model and quantitatively evaluate the accuracy of the generated articulatory movements and vocal-tract spectra.

This paper is organized as follows. Section II explains the mathematical framework for specifying the articulatory timing and phonemic tasks, forming articulatory movements, and estimating the vocal-tract spectrum. Section III presents the results of the experiments, and finally Section IV provides a summary of this work and our conclusions.

II. THE GENERATIVE MODEL

A. Outline of the model

This section explains the generative model and shows how the vocal-tract spectrum is generated from phonemic information. The model is composed of processes of articulatory timing determination, phonemic task specification, articulatory movement generation, and spectrum estimation (Fig. 1). Given a sequence of phoneme symbols, the timing

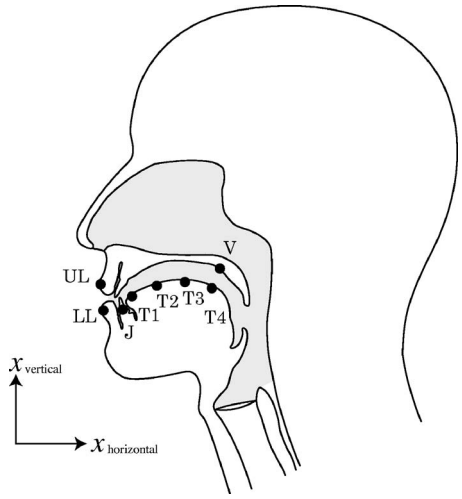


FIG. 2. Articulatory variables represent the positions of eight marker points on the midsagittal plane of a speaker. The axes of the coordinate system, $x_{\text{horizontal}}$ and x_{vertical}, are fixed to the head position.

model determines the moment at which each phoneme should be articulated, and the obtained temporal information is used to constrain the dynamic behavior of the articulators. Phonemic tasks represent articulatory targets for forcing the vocal tract into phoneme-specific configurations. Then a dynamic articulatory model generates continuous articulator movements so that every task is satisfied. Finally, the vocal-tract spectrum is estimated from the positions of the articulators on a frame-by-frame basis.

B. Articulatory and acoustic variables

Articulatory variables of the model represent the absolute positions of eight points fixed on the articulatory organs. As illustrated in Fig. 2, the number of points is one each for the jaw (J), upper lip (UL), lower lip (LL), and velum (V), and four for the tongue (T1, T2, T3, T4). Each marker position is expressed in a two-dimensional space corresponding to the midsagittal plane of a speaker, where the coordinate system is fixed to the head position. Therefore, the dimension of the articulatory variables becomes 16. This point representation of articulatory variables has the advantage that their actual values are directly measured by using an electromagnetic observation technique (Kaburagi and Honda, 1994, 2002). As will be explained later, measurement data are usable for determining the values of model variables and phonemic articulatory targets.

On the other hand, the vocal tract transfer function is represented by LSP (line spectrum pair) parameters (Furui, 2000). In this study, audio signals are recorded at a sampling frequency of 8 kHz and the order of the LSP analysis, i.e., the number of poles of the all-pole spectral model, is set at 14. The articulatory and acoustic variables are respectively represented as

$$\mathbf{x} = (x_1, x_2, x_3, \ldots, x_{Lx})^T \tag{1}$$

and

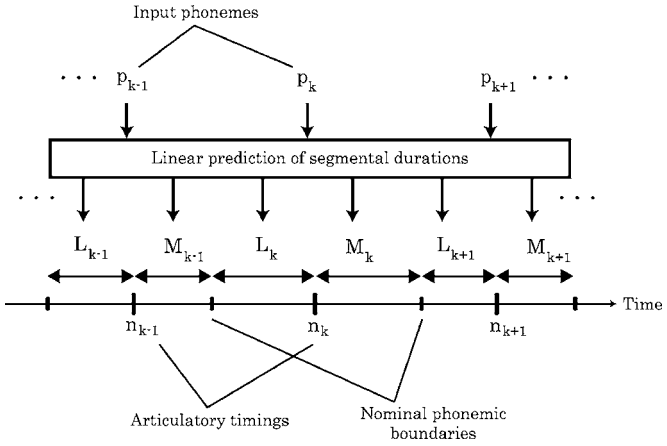


FIG. 3. Estimation of phonemic lengths to determine articulatory timings. Phoneme onset, L_k, and phoneme offset, M_k, are estimated using linear models from given phonemic information. The initiation time of the kth phoneme is illustrated as the phonemic boundary between the $(k-1)$th and kth phonemes.

$$\mathbf{y} = (y_1, y_2, y_3, \ldots, y_{Ly})^T, \tag{2}$$

where T denotes transposition. $Lx(=16)$ and $Ly(=14)$ are the dimension numbers of vectors.

C. Determination of articulatory timings

When a sequence of phoneme symbols is input to the model as

$$p_1 p_2 p_3 \cdots p_k \cdots p_K, \tag{3}$$

the articulatory timing of each phoneme is first determined as n_k $(k=1,2,\ldots,K)$, where K represents the number of phonemes. Suppose that the timing of the first phoneme is set at $n_1=1$. The timing of the second phoneme is then determined as $n_2=n_1+M_1+L_2$, where M_1 and L_2 represent segmental durations determined from the types of the phonemes p_1 and p_2. As illustrated in Fig. 3, the articulatory timing of the kth phoneme is generally predicted as

$$n_k = n_{k-1} + M_{k-1} + L_k \quad (k = 2,3,\ldots,K). \tag{4}$$

L_k is the time duration, referred to as the phoneme onset, between the initiation of the phoneme production and the reaching of the articulatory goal. Similarly, M_k is the duration, referred to as the phoneme offset, between the articulation time of the kth phoneme and the initiation time of the succeeding phoneme. In the figure, each initiation time is indicated as a nominal phonemic boundary between adjacent phonemes.

The lengths of the phoneme onset and offset are predicted as a function of phonemic types using the type-I quantification method (Hayashi, 1954)

$$L_k = \bar{L} + L_1(p_k) + \sum_{i=1}^{N_c} \delta_i(p_{k-1}) L_{2i} + \sum_{i=1}^{N_c} \delta_i(p_{k+1}) L_{3i} \tag{5}$$

and

$$M_k = \bar{M} + M_1(p_k) + \sum_{i=1}^{N_c} \delta_i(p_{k-1}) M_{2i} + \sum_{i=1}^{N_c} \delta_i(p_{k+1}) M_{3i}, \tag{6}$$

where $\bar{L}$ and $\bar{M}$ are the mean durations and L_1 and M_1 are additional terms specific to each phoneme. Also, contextual effects are considered to improve the prediction accuracy. Phonemes are first classified into N_c groups on the basis of their articulatory features. The terms L_{2i} and M_{2i} then represent influences of the preceding phoneme which belongs to the ith group. Similarly, L_{3i} and M_{3i} represent influences of the following phoneme. The value of the function, $\delta_i(\,)$, is one when the argument of the function, p_{k-1} or p_{k+1}, belongs to the ith group; if not, it takes zero.

As will be explained in the third section, the prediction parameters in Eqs. (5) and (6) are learned using empirical articulatory data taken by electromagnetic articulography. By a visual inspection of measured articulatory movements, the articulatory timing of each phoneme can be determined. Then, the instants of the nominal phonemic boundaries are determined as

$$n_{k,k+1} = \arg\min_n \{e_k(n) - e_{k+1}(n)\}, \tag{7}$$

where $n_{k,k+1}$ is the boundary between the kth and $(k+1)$th phonemes. $e_k(n)$ is the articulatory distance, defined as

$$e_k(n) = (\mathbf{x}(n) - \mathbf{x}_k)^T W (\mathbf{x}(n) - \mathbf{x}_k), \tag{8}$$

where $\mathbf{x}(n)$ is a sequence of measured articulatory movements, and $\mathbf{x}_k[=\mathbf{x}(n_k)]$ is the articulatory position of the kth phoneme at time n_k. The components of the weighting matrix $W=\mathrm{diag}(v_1, v_2, \ldots, v_{L_x})$ are given as $v_l \propto c_l^{-0.5}$ $(\Sigma v_l = 1)$ for c_l, the variance of the articulatory variables in the data set. These expressions indicate that the phoneme boundary is determined as the time at which the difference between $e_k(n)$ and $e_{k+1}(n)$ becomes the minimum. In other words, we suppose that the articulatory distance between $\mathbf{x}_k$ and $\mathbf{x}(n_{k,k+1})$ is the same as that between $\mathbf{x}_{k+1}$ and $\mathbf{x}(n_{k,k+1})$.

When the articulatory timings and phonemic boundaries are determined from the articulatory data, the values of the prediction parameters are calculated following the ordinal type-I quantification method so that the mean squared error of the segmental durations, i.e., phoneme onset and offset, is minimized. In the present study, the phonemes of Japanese are classified into 13 groups $(N_c=13)$, as will be discussed in the third section, and the linear expressions in Eqs. (5) and (6) are obtained for each of these groups. In accord with the structure of the expressions, the *items* of the type-I quantification method are set as the types of the preceding, central, and following phonemes. For the item of the central phoneme, the *categories* of the method are formed by the members of the phoneme group. Therefore, the values of L_1 and M_1 are determined for each phoneme separately. On the other hand, for the items of the preceding and following phonemes, the categories are formed by the 13 phoneme groups to determine the values of L_{2i}, L_{3i}, M_{2i}, and M_{3i} for $i=1,2,\ldots,N_c$. Because the linear models are trained for each group independently, these parameters should be expressed, for example, as $L_{2i}(c_k)$ as a function of c_k, the group number

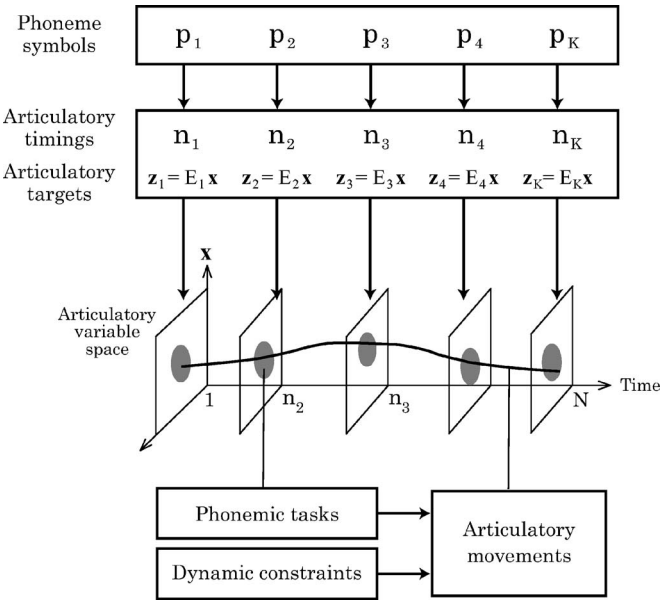


FIG. 4. Illustration of the task generation model and the dynamic articulatory model constructed by combining the results of the foregoing studies (Kaburagi and Honda, 1996, 2001). Phonemic tasks represent when and how the articulatory organs should be organized to form the vocal tract into phoneme-specific configurations. They are used to constrain the values of the articulatory variables at the specified articulation timings. Articulatory movements are then determined so that they jointly satisfy phonemic tasks and dynamic constraints about the smoothness of movements.

in which the phoneme p_k is included. Similarly, $\bar{L}$ and $\bar{M}$ are the mean values of phoneme onset and offset within a group, and they should be written as $\bar{L}(c_k)$ and $\bar{M}(c_k)$ in the strict sense. Therefore, the number of parameters is N_c for $\bar{L}$ and $\bar{M}$, N_p for L_1 and M_1, and N_c^2 for L_{2i}, L_{3i}, M_{2i}, and M_{3i}, where N_p is the number of phoneme types. Because N_p is 23 and N_c is 13, the total number of parameters is 748.

D. Task generation and articulatory movement formation

After the articulatory timings are determined, phonemic tasks, i.e., articulatory constraints that must be met for the production of each phoneme, are specified as (Kaburagi and Honda, 1996)

$$\mathbf{z}_k = E_k \mathbf{x}(n_k) \quad (1 \leqslant k \leqslant K), \tag{9}$$

where $\mathbf{z}_k$ is a vector representing the target values of the articulatory variables and E_k is the transformation matrix. As shown in Fig. 4, conditions $n_1=1$ and $n_K=N$ are supposed at both ends of the utterance. If the components of $\mathbf{z}_k$ and E_k are specified, then it is clear that the expression in Eq. (9) forms simultaneous linear equations with respect to the unknown values of $\mathbf{x}$.

The values of $\mathbf{z}_k$ and E_k are determined using the invariant-feature representation (Kaburagi and Honda, 2001) by transforming $\mathbf{x}$ to a feature subspace and setting the projected values at zero

$$0 = \mathbf{f}_{pl}^T(\mathbf{x} - \bar{\mathbf{x}}_p) \quad (l = 1,2, \ldots, L_p), \tag{10}$$

where p represents the type of phoneme, $\bar{\mathbf{x}}_p$ is the mean articulatory position, $\mathbf{f}_{pl}$ is the transformation vector, and $L_p(\leqslant L_x)$ is the task dimension specific to each phoneme. The phonemic task in Eq. (9) is obtained by rewriting Eq. (10) as $\mathbf{f}_{pl}^T\bar{\mathbf{x}}_p = \mathbf{f}_{pl}^T\mathbf{x}$ and setting $\mathbf{z}_k = (\mathbf{f}_{p1}^T\bar{\mathbf{x}}_p, \mathbf{f}_{p2}^T\bar{\mathbf{x}}_p, \ldots, \mathbf{f}_{pLp}^T\bar{\mathbf{x}}_p)^T$ and $E_k = (\mathbf{f}_{p1}, \mathbf{f}_{p2}, \ldots, \mathbf{f}_{pLp})^T$. The mathematical derivation of $\mathbf{f}_{pl}$ and the meaning of the invariant-feature representation have been precisely described in the literature, but it is noteworthy that $\mathbf{f}_{pl}$ is designed so that it projects articulatory variables onto a subspace where the articulatory variance becomes small. From a small articulatory variability, we can expect that the transformed variables will be related to characteristic articulatory gestures and less dependent on the phonemic context. In Eq. (10), a multiple number of transformation vectors are used to constrain the articulatory variables simultaneously, but the remaining $L_x - L_p$ degrees of freedom are set free from the constraints.

Next, movements of the articulatory organs are determined as the values of articulatory variables $\mathbf{x}(n)$ $(1 \leqslant n \leqslant N)$ so that every task is satisfied. However, the task dimension (L_p) in Eq. (10) is usually smaller than the dimension of articulatory variables (L_x). Also, the articulators can take an arbitrary position in the interval between adjacent tasks. Therefore, dynamic constraints are imposed to resolve these redundancies and determine articulatory movements uniquely (Kaburagi and Honda, 1996). First, each articulatory variable is represented as the output of a dynamical system

$$x_l(n) - 2\tau_l x_l(n-1) + \tau_l^2 x_l(n-2) = (1-\tau_l)^2 s_l(n) \quad (1 \leqslant l \leqslant L_x), \tag{11}$$

where s_l is the input force and τ_l is a parameter specifying the time constant of the system. Also, an energy criterion is introduced as

$$C = \sum_{n=1}^{N-1} (C_x(n) + C_s(n)) + C_x(N), \tag{12}$$

where C_x is related to the velocity of movements as

$$C_x(n) = (\mathbf{x}(n) - \mathbf{x}(n-1))^T W_x (\mathbf{x}(n) - \mathbf{x}(n-1)), \tag{13}$$

and C_s to the change of system inputs as

$$C_s(n) = (\mathbf{s}(n) - \mathbf{s}(n-1))^T W_s (\mathbf{s}(n) - \mathbf{s}(n-1)), \tag{14}$$

and $\mathbf{s}(n)$ is a vector storing input forces. The weighting matrices are expressed as $W_x = \mathrm{diag}\{w_1, w_2, \ldots, w_{L_x}\}$ and $W_s = \mathrm{diag}\{d_1 w_1, d_2 w_2, \ldots, d_L w_{L_x}\}$, where the parameter d_l adjusts the overall shape of the movement pattern and w_l controls the relative amplitude among the articulatory variables.

Finally, the articulatory movements are determined as the outputs of linear systems [Eq. (11)] so that they satisfy every phonemic task [Eq. (9)] and minimize the energy criterion [Eq. (12)]. The formation of articulatory movements in this mathematical framework generates an optimal control problem with linear dynamics and quadratic criteria. It can be explicitly solved by applying the technique of dynamic programming (Kaburagi and Honda, 1996).

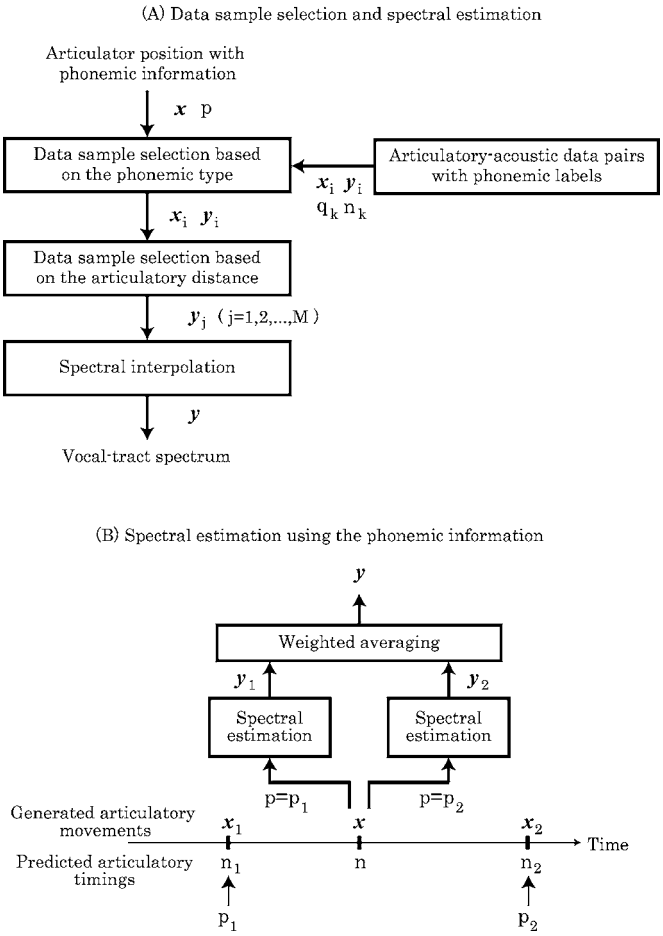


FIG. 5. The procedure for estimating the vocal-tract spectrum from the positions of the articulatory organs. The upper part of the figure shows the procedure for selecting neighboring samples of the articulatory-acoustic data pairs and interpolating the selected spectral samples. Here, the sample selection is first performed based on the matching of the phonemic type. The neighboring samples are then selected using the distance of the articulator positions. In the lower part of the figure, the whole estimation procedure is illustrated where the phonemic context is taken into account to achieve a smooth temporal change of the resultant spectral sequence. To estimate the spectrum ($\mathbf{y}$) for an articulatory vector ($\mathbf{x}$) located between the articulatory timings of phonemes p_1 and p_2, the estimation procedure in (a) is performed for each phonemic condition separately, and the output spectra ($\mathbf{y}_1$ and $\mathbf{y}_2$) are interpolated to produce the estimation result.

E. Estimation of the vocal-tract spectrum

Finally, the vocal-tract spectrum is estimated from the articulatory movements. Because the articulatory variables used in this study are not capable of providing enough information about the whole configuration of the vocal tract, the estimation of the area function and the adoption of an acoustic tube model are difficult. Therefore, we use a statistical method based on a search of articulatory-acoustic data pairs. Because these data pairs directly represent the relationships between the articulator positions and the vocal-tract spectra, the estimation can be performed by finding data samples that agree with the articulator positions generated by the articulatory model.

In the upper part of Fig. 5, the block diagram shows the procedure for determining the vocal tract spectrum ($\mathbf{y}$) from the articulator position ($\mathbf{x}$) and the associated phonemic symbol (p). As will be described in the following section, the articulatory and acoustic data are obtained by a simultaneous measurement of the articulatory movements and audio signals performed for various utterances. After the measurement, the vocal tract transfer function is estimated by applying an LPC-based speech analysis method. Phonemic labels are also assigned at the articulatory timing of every phoneme. Here, we denote samples of the articulatory-acoustic data as $\mathbf{x}_i$ and $\mathbf{y}_i$ $(i=1,2,3,\ldots N_1)$ where N_1 is the number of data pairs. The phonemic labels are expressed as q_k and the corresponding timing information as n_k $(k=1,2,3,\ldots,N_2)$, where N_2 is the total number of labels.

Selection of the articulatory-acoustic data pairs is first performed based on phonemic matching to ensure the acoustical reliability of the resulting spectrum. When the phonemic type of $\mathbf{x}$ coincides with a label as $p=q_k$, the data samples $\mathbf{x}_i$ and $\mathbf{y}_i$ located in the time interval $n_{k-1} \leqslant n \leqslant n_{k+1}$ are selected as the candidates. Here, n_{k-1} and n_{k+1} are the articulatory timings of the preceding and following phonemes of q_k in an utterance. This sample selection is also effective in reducing the cost of computing articulatory distances. The articulatory distance given in Eq. (8) is rewritten as

$$e_i = (\mathbf{x}-\mathbf{x}_i)^T W(\mathbf{x}-\mathbf{x}_i), \tag{15}$$

and used to select data samples, where $\mathbf{x}$ is the input articulatory position and $\mathbf{x}_i$ is a data sample with the same phonemic type. The articulatory distances calculated for $\mathbf{x}_i$ $(i=1,2,3,\ldots)$ are sorted in ascending order with the first M samples selected as the neighboring samples, as $\mathbf{x}_j$ and $\mathbf{y}_j$ for $j=1,2,\ldots,M$, so that their articulatory distances are smaller than the remaining samples. Here, the parameter M specifies the number of neighboring samples. The vocal-tract spectrum is then calculated using a weighted interpolation of selected samples as

$$\mathbf{y} = \sum_{j=1}^{M} \alpha_j \mathbf{y}_j. \tag{16}$$

The weighting coefficient α_j is given as $\alpha_j \propto e_j^{-1}(\Sigma\alpha_j=1)$ so that a data sample closer to the input position is weighted more heavily.

The lower part of Fig. 5 shows the entire estimation procedure in which the phonemic context is taken into account. Suppose p_1 and p_2 denote phoneme symbols input to the generative model, n_1 and n_2 their calculated articulatory timings, and $\mathbf{x}(n)$ the generated articulatory movements. The spectral estimation is performed on a frame-by-frame basis to produce a sequence of spectral patterns as $\mathbf{y}(n)$. In the interval $n_1 \leqslant n < n_2$, the values of the articulatory variables change smoothly from $\mathbf{x}_1=\mathbf{x}(n_1)$ to $\mathbf{x}_2=\mathbf{x}(n_2)$. Both phonemes ($p_1$ and p_2) can be associated with $\mathbf{x}=\mathbf{x}(n)$, and then the output spectrum is determined as

$$\mathbf{y} = \beta_1 \mathbf{y}_1 + \beta_2 \mathbf{y}_2. \tag{17}$$

$\mathbf{y}_1$ and $\mathbf{y}_2$ are the spectra estimated by the procedure depicted in Fig. 5(a), where the phonemic information is set as $p=p_1$ and $p=p_2$, respectively. The weights are calculated as

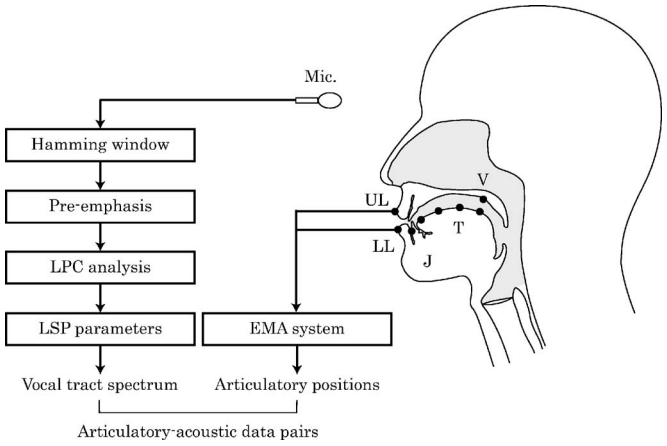


FIG. 6. Construction of the articulatory and acoustic data pairs using the simultaneous measurement of articulatory movements and audio signals. Articulatory data were taken using an electromagnetic sensor that tracked the positions of the jaw (J), upper lip (UL), lower lip (LL), velum (V), and four fixed points on the tongue (T). The acoustic data were obtained as the vocal-tract spectra by applying a linear predictive analysis. Both types of data were synchronized by positioning the window function of the signal processing at each sampling point of the magnetic data.

$$\beta_1 = \frac{e_1^{-1}}{e_1^{-1} + e_2^{-1}}, \quad \beta_2 = \frac{e_2^{-1}}{e_1^{-1} + e_2^{-1}}, \tag{18}$$

where e_1 is the articulatory distance [Eq. (15)] between $\mathbf{x}$ and $\mathbf{x}_1$, and e_2 is that between $\mathbf{x}$ and $\mathbf{x}_2$. When n is n_1, e_1 becomes zero and the weights are not defined and, therefore, the output spectrum is determined as $\mathbf{y}=\mathbf{y}_1$. When n approaches n_2 from n_1, β_1 approximately decreases from one to zero, and β_2 increases from zero to one. Our method can thus produce a smooth change of the resulting spectrum.

III. EXPERIMENTS

This section presents results of experiments for measuring the articulatory and acoustic data, evaluating each part of the generative model, and generating the vocal-tract spectrum. In particular, the third experiment demonstrates the ability of the model for reproducing context-dependent variability of speech using context-independent tasks and the dynamic articulatory mechanism.

A. Measurement of articulatory and acoustic data

A set of articulatory and acoustic data was assembled based on the simultaneous measurement of articulatory movements and audio signals as illustrated in Fig. 6. To monitor movements of the articulatory organs that can influence the vocal tract transfer function, receiver coils of an EMA system (Carstens Articulograph AG100, Germany) were attached to the jaw, upper lip, lower lip, tongue, and velum on the midsagittal plane, and their movements were recorded at a sampling rate of 250 Hz. Four receiver coils on the tongue were placed from the tip to the dorsum at almost equal intervals. The positions of two coils on the nose bridge and upper incisors were also measured to calibrate the movement of the head. By using nonparametric magnetic field representation (Kaburagi and Honda, 2002), it was proved that the accuracy of the measurement system is about 0.1 mm for a 14×14 cm^2 region when the receiver coil is exactly positioned on the midsagittal plane. The error is about 1 mm when the off-center misalignment is 2 mm and the tilt angles with respect to the x and y axes are less than 20 deg.

Speech signals were recorded in synchronization with the EMA system at a sampling frequency of 8 kHz and processed using a 30 ms hamming window and second-order pre-emphasis. The center position of the hamming window was set at each sampling point of the EMA measurement to synchronize the articulatory and acoustic data in time. The pre-emphasis was designed to cancel the spectral tilt of glottal source signals and the radiation characteristics of the lips. Finally, a 14th order LPC analysis was performed and the values of the LSP parameters were determined as the acoustic data representing the vocal-tract spectrum. Articulatory and acoustic measurements were performed with a male Japanese subject while he uttered 50 sentences twice. In addition, the instant at which each phoneme was articulated was determined manually by an operator, and a label representing the corresponding phonemic type was assigned to the data set.

B. Evaluation of the estimation accuracies

The obtained data were first used to quantitatively evaluate the generative model. Because the accuracy of the dynamic articulatory model was investigated in the foregoing article (Kaburagi and Honda, 2001), examinations were performed for the temporal and spectral estimation methods. For estimating the articulatory timings, phonemic groups were made as shown in Table I, and the models were trained using the articulatory data. Note that even though the total number of groups was 13, the last group was excluded from the table because it was used only to represent the neutral state of the articulators.

All of the articulatory data were used for the parameter learning, and the number of data samples was 1542 for the vowel group. For the consonant groups two to nine, the average sample number was 201. The nasal group had the maximum number of samples at 344, while the bilabial stops had the minimum at 46. For groups 10–13, the sample number was 18 for the geminate obstruent but more than 66 for the others. When the number of groups was increased, we could not obtain a proper solution of the parameter estimation problem due to the lack of training data. Thus, the present phoneme classification is reasonable from the viewpoint of parameter learning.

In the table, S.D. and error, respectively, indicate the standard deviation and the mean absolute difference between the predicted and actual values. Also, the difference between them was calculated as $100\times$[(S.D.)-(prediction error)/(S.D.)](%) and shown as I.R., the improvement rate. I.R. is a measure of error reduction achieved by applying the linear estimation models [Eqs. (5) and (6)], since the standard deviation represents the mean error when the phoneme onset and offset are predicted by their mean durations. As a result, the prediction error averaged for every group was

TABLE I. Classification of Japanese phonemes and results of the experiment for predicting the durations of the phoneme onset and offset. Syllabic nasal, geminate obstruent, and syllable elongation are specific in Japanese. They usually occupy a certain period and constitute a mora. The 13th group represents a "neutral" state of the articulators, which can be found at the beginning or the terminal end of an utterance, but this state is not shown in the table. Unvoiced vowels are classified to the first group. In the experiment, the values of L_1 and M_1 were determined for each phoneme, and those for L_{2i}, L_{3i}, M_{2i}, and M_{3i} were determined using the 13 phonemic groups. For the explanations of S.D., error, and I.R. (improvement rate), see the text.

	Phoneme onset (L_k)			Phoneme offset (M_k)		
Category (phonemes)	S.D. (ms)	Error (ms)	I.R. (%)	S.D. (ms)	Error (ms)	I.R. (%)
1. Vowels (/a,i,u,e,o/)	14.0	11.2	20.0	20.3	15.6	23.2
2. Semivowels (/j,w/)	10.5	8.0	23.8	12.4	9.2	25.8
3. Bilabial stops (/p,b/)	5.6	3.1	44.6	8.4	5.6	33.3
4. Alveolar stops and affricates (/t,d/)	17.0	12.6	25.9	9.7	8.8	9.3
5. Velar stops (/k,g/)	17.7	11.6	34.5	12.9	10.1	21.7
6. Alveloar fricatives (/s,z/)	15.3	13.7	10.5	13.8	11.2	18.8
7. Glottal and labial fricatives (/h,f/)	22.1	14.8	33.0	15.1	8.3	45.0
8. Nasals (/m,n/)	11.5	10.0	13.0	10.8	7.1	34.3
9. Liquid (/r/)	9.5	7.9	16.8	7.8	7.0	10.3
10. Syllabic nasal (/N/)	14.2	7.8	45.1	26.8	13.5	49.6
11. Geminate obstruent (/Q/)	17.6	9.1	48.3	44.1	16.0	63.7
12. Syllable elongation (/H/)	19.7	13.5	31.5	26.4	12.8	51.5
(mean)	14.6	10.3	28.9	17.4	10.4	32.2

about 10.3 (ms) for the onset and 10.4 (ms) for the offset. The mean improvement rate was about 28.9% for the onset and 32.2% for the offset, respectively. The improvement rate of each group was more than 10% except for the offset of the fourth group. Because the mean duration is the first term of both Eqs. (5) and (6), we can conclude that improvements were achieved by considering the phoneme-specific and context-dependent variations of segmental durations.

Next, the accuracy of the spectrum estimation procedure was examined using pairs of the articulatory and acoustic data. Of all of the utterances of the data set, five were randomly selected as the test set, and the data for the remaining utterances were used as the training set, i.e., the articulatory-acoustic data pairs appearing in the block diagram of Fig. 5(a). The vocal-tract spectrum was estimated for each sample of test utterances, and the error between the estimated and actual spectra was calculated using a 30th order cepstrum distance. This evaluation procedure was repeated three times by changing the combination of test and training utterances.

Experimental results are shown in Fig. 7 as a function of the number of neighboring samples M of Eq. (16). First, the articulatory distance between the test and training samples was calculated as the Euclidean distance of the articulatory variables for each sample of test utterances, and then the mean distance was plotted in Fig. 7(a). When M was one (zero bits), the articulatory distance was the minimum because the nearest sample was selected for the estimation. As expected, the articulatory distance monotonically increased as the sample number increased. On the other hand, the mean spectral error plotted in Fig. 7(b) shows that a minimum error of about 3.08 dB was obtained at sample number 64 or 128. This indicates that the sample interpolation [Eq. (16)] is effective for achieving continuous mapping and reducing the estimation error.

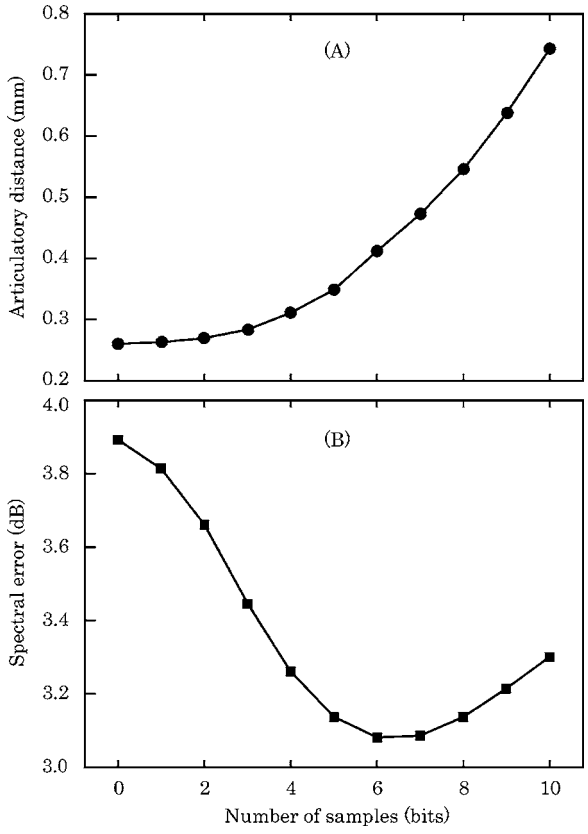


FIG. 7. Results of the experiment for determining the accuracy of the spectral estimation method. The plots in (a) show the mean articulatory distance between each test sample and selected samples of data pairs. The horizontal axis is the number of data pair samples used for the spectral interpolation. In (b), the mean spectral error is shown as a function of the sample number.

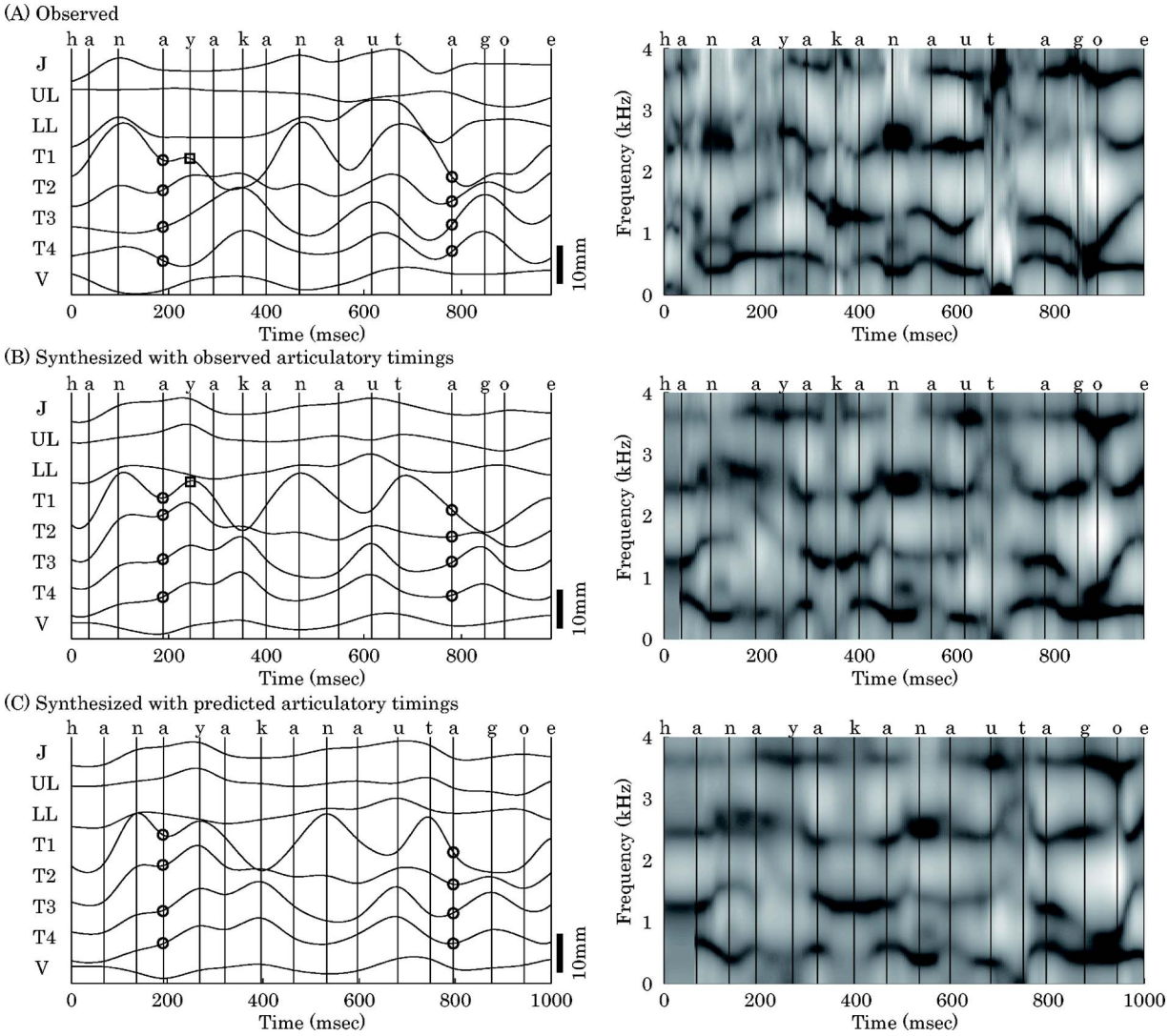


FIG. 8. Results of the spectrum generation experiment for the utterance /hanayakanautagoe/ (*happy singing voice*). In (a), the articulatory movements and spectrogram are shown for the observed data. In (b), the articulatory movements and spectral sequence were generated by the model, but the articulatory timings were obtained from the measured articulatory data. In (c), the synthesis was performed with the estimated articulatory timings. Note that the traces on the left side of (a)–(c) show vertical movements of the articulators. Squares in the articulatory plots indicate the positions of the tongue tip marker when /y/ was articulated. Circles denote the positions of four tongue markers for /a/.

C. Spectrum generation

Next, Fig. 8 shows the experimental results of the phonemic sequence /hanayakanautagoe/, a Japanese phrase meaning *happy singing voice*, taken from a sentence used in collecting the measurement data. In the figure, the articulatory movements and the time sequences of the vocal-tract spectrum are shown on the left and right, respectively. Measurement data are plotted in Fig. 8(a) and the synthesized results are plotted in Figs. 8(b) and 8(c). Here, the articulatory timings were given by the measurement in Fig. 8(b) and predicted by the model in Fig. 8(c). Therefore, the articulatory timings shown by the vertical lines in Fig. 8(b) are the same as those in Fig. 8(a). The symbol at the top of each vertical line represents the input linguistic information. The articulatory plots indicate the vertical movements of the articulators. In the experiment, the dimension of the phonemic task (L_p) and the values of the parameters included in the dynamic articulatory model, i.e., τ_l, w_l, and d_l, were determined so that the mean difference between the predicted and actual movements was minimized for all of the articulatory data. The number of the neighboring samples in the spectrum estimation was set at $M=64$.

In Fig. 8(a), we can observe upward movements of the tongue tip (T1) for the consonants /n/ and /t/ and those of T3 and T4 for /k/ and /g/ closing the vocal tract. These articulatory features are inherent in the production of each phoneme and used to represent the phonemic tasks of the dynamic articulatory model (Kaburagi and Honda, 2001). In general, the generated articulatory patterns in Fig. 8(b) agreed well with the actual movements in Fig. 8(a), but a discrepancy was found for /y/ where the movement of T1 marked by the square was more prominent in Fig. 8(b) than in Fig. 8(a). In the case of Fig. 8(a), this result suggests that the preceding and following vowels affected the articulation of /y/ and that the tongue remained in the lower portion of the tract, but this contextual effect was less pronounced in Fig. 8(b). On the other hand, when the tongue position for /a/ in the /nay/ context is compared with that in the /tag/ context in Fig. 8(a)

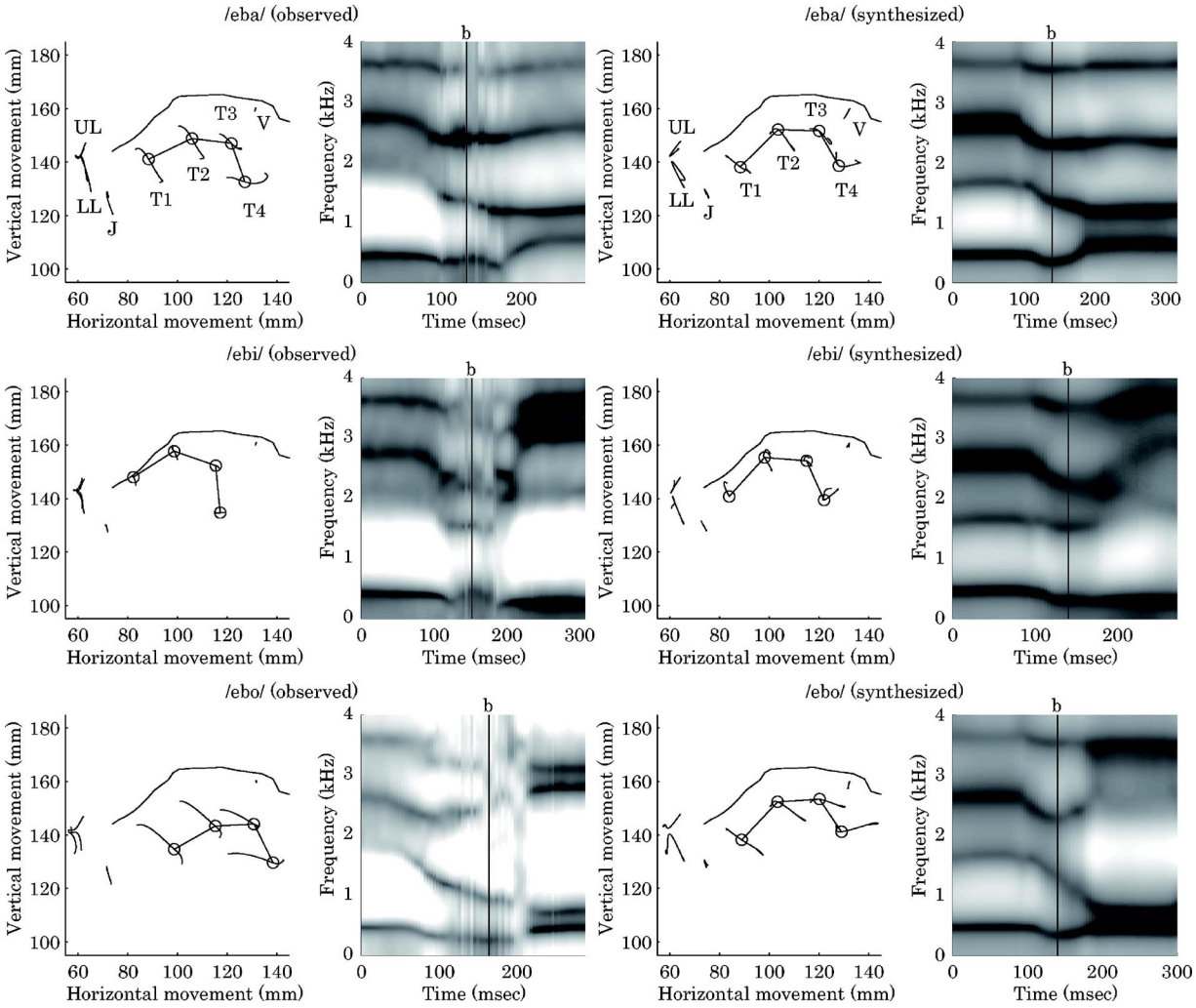


FIG. 9. Comparison of the observed and synthesized patterns of the articulatory and acoustic data for three VCV syllables. The preceding vowel and the central consonant were, respectively, fixed at /e/ and /b/, but the following vowel was changed to /a/, /i/, or /o/. The articulatory pattern is shown as the trajectory of eight marker points. The anterior side of the head is to the left. The shape of the palate was extracted from an x-ray of the subject. It was superimposed on the figure by using a magnetically measured palate shape as a reference. On the other hand, the spectrogram was calculated from the sequence of the vocal-tract spectrum. The vertical line in each spectrogram shows the articulatory timing of the central consonant. The preceding and following vowels were articulated at the left and right sides of the spectrogram, respectively.

(circles in the figure), the tongue tip is clearly more raised in /nay/ than in /tag/. This observation also holds for Figs. 8(b) and 8(c), indicating the ability of the model for predicting the context-dependent variability of articulatory movements.

In the right part of the figure, the respective spectrograms made from time sequences of the LSP parameters are shown. Note that speech signals were pre-emphasized before the LPC analysis so that the spectral tilt was flattened. Formants in Figs. 8(b) and 8(c) were less prominent than those in Fig. 8(a), indicating that the bandwidths were widened. It is possible that this was caused by the interpolation of multiple spectral patterns in converting the articulatory variables to the acoustic ones.

D. Representation of the contextual variability

Finally, experiments were performed for a number of VCV utterances to see the contextual effects in more detail. Figure 9 shows the observed and synthesized patterns of the bilabial consonant /b/ for the three sequences /eba/, /ebi/, and /ebo/. On the left, the articulatory trajectories and spectrograms are plotted for the measurement data. The synthesized results are shown on the right. The vertical and horizontal axes of the articulatory plots correspond to the coordinate system of the electromagnetic device. The circles indicate the positions of the tongue markers (T1, T2, T3, and T4) when the consonant was articulated. They are also connected by line segments to show the tongue configuration. The vertical line in each spectrogram indicates the articulatory timing of the consonant.

The figure shows that the tongue position of /b/ is affected by the neighboring vowels because it is not strongly constrained by the task of the bilabial consonant (Kaburagi and Honda, 2001). The generated tongue configuration of each utterance is in agreement with the actual configuration, but the model tends to predict a reduced or neutralized tongue posture. This inconsistency can in part be explained by the utterance condition of the measurement data from which the phonemic tasks were calculated, i.e., these data were measured for sentences, but the observed data plotted in the figure were taken for VCV utterances. In addition, the

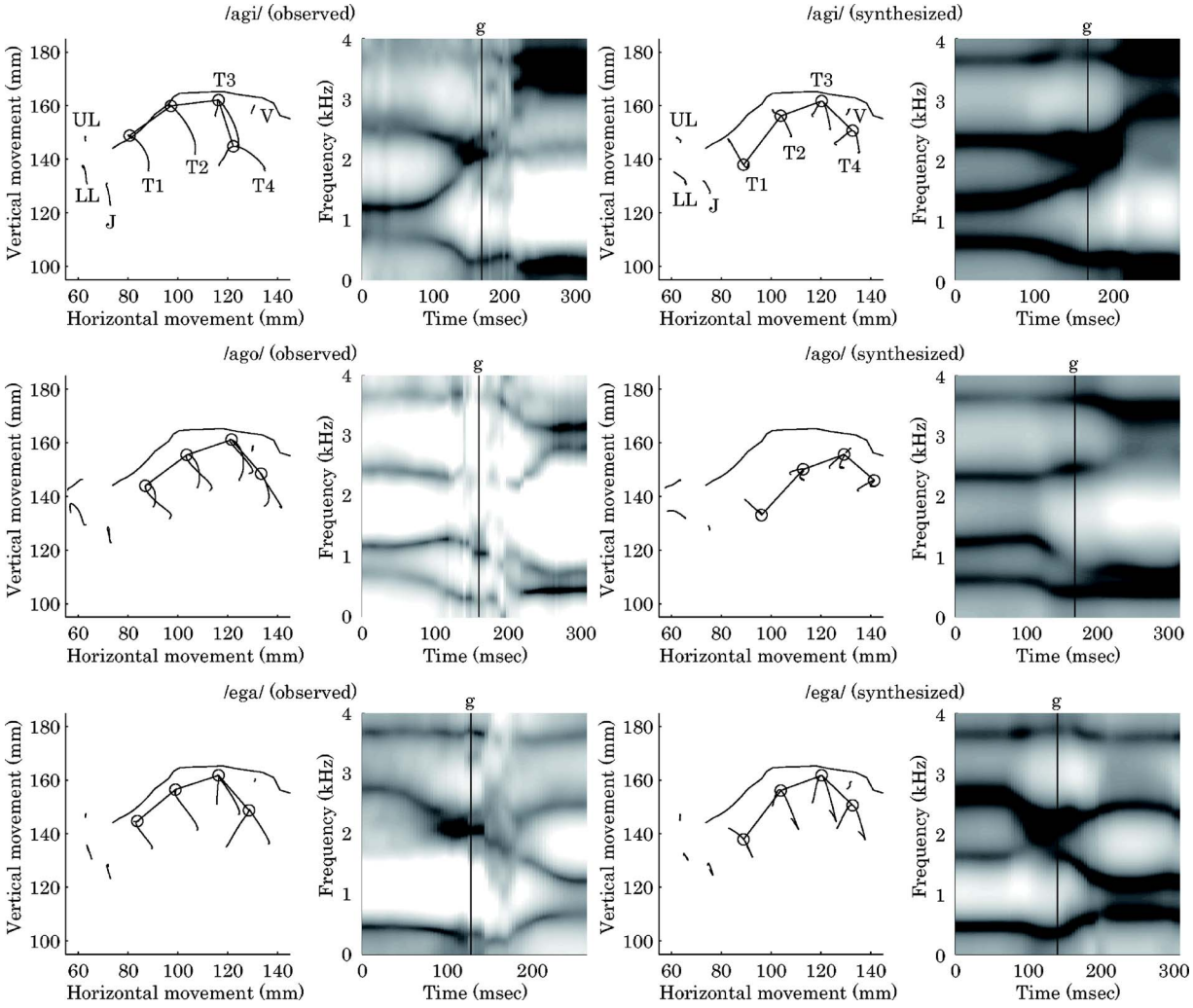


FIG. 10. Comparison of the observed and synthesized patterns of the articulatory and acoustic data for syllables /agi/, /ago/, and /ega/.

phonemic tasks were statistically calculated by blending a number of articulatory configurations uttered with various phonemic contexts, and this was another reason for the neutralization. On the other hand, the spectrograms show that the model can generate smooth movements of the formants. Also, the influence of the second vowel is clearly reproduced. For example, the second formant of the sequences /eba/ and /ebo/ falls down smoothly in the vicinity of the consonant. For the sequence /ebi/, however, it falls and raises quickly, just as it did in the observed speech.

Experimental results are shown in Figs. 10–12 for /g/, /m/, and /r/. On the left side of Fig. 10, the tongue shapes are shown to be influenced by the vowel context, but the position of the third marker (T3) is relatively fixed in order to close the vocal tract. On the right side, the predicted tongue shapes are shown to also be affected by the vowel context. For the sequence /ago/, the tongue is located posteriorly since the consonant is surrounded by back vowels. By careful inspection, we can see that the velum is in a lower position in order to touch the tongue. In the spectrograms, the dynamic changes of the formants are well generated. For example, the predicted frequency of the second formant is about 1 kHz for the sequence /ago/, which is lower than the corresponding frequencies of the sequences /agi/ and /ega/, as was the case in the observed speech.

The articulatory movements shown in Fig. 11 are quite similar to those in Fig. 9 except for the noticeable movements of the velum. The synthesized spectral patterns agreed well with the actual ones, but the formants moved quickly in the vicinity of the consonant. As explained before, this can partly be attributed to the difference in utterance conditions; the temporal prediction model was learned using sentence data, but the observed data shown in the figure were taken for VCV utterances. Moreover, the articulators were constrained by the tasks only at given points in time. Therefore, our modeling scheme makes it difficult to keep the articulators in an articulatory position. Lastly, Fig. 12 shows that the characteristic movements of the tongue were well simulated for /r/. While the tip position was constrained in order to touch the hard plate, the shape of the back portion was determined depending on the vowel context. Also, the generated movements of the formants were quite similar to the actual patterns.

The accuracy of the dynamic articulatory model and spectral estimation method was determined for the utterances shown in Figs. 8–12, and the results are summarized in Table II. The articulatory error in Table II(A) indicates the mean error of the articulatory variables. The error for these 13 utterances was about 1.27 mm on average, indicating that the

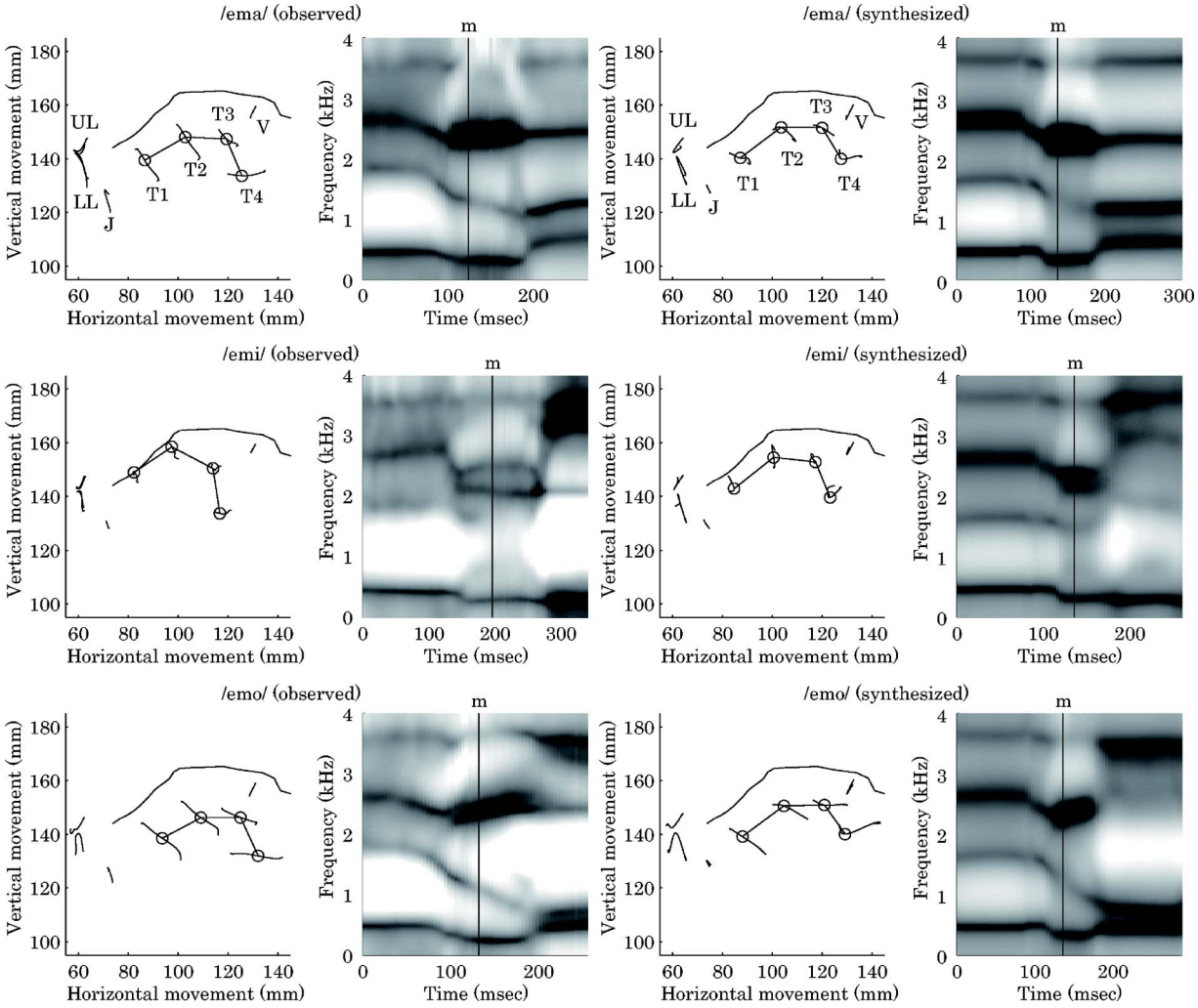


FIG. 11. Comparison of the observed and synthesized patterns of the articulatory and acoustic data for syllables /ema/, /emi/, and /emo/.

simulated movements agreed well with the observed data. On the other hand, the spectral error was about 2.53 dB on average when the vocal-tract spectra were estimated from the observed articulatory movements, as shown in Table II(B). This error increased to 3.44 dB, as shown in Table II(C), when the spectra were estimated from the generated articulatory movements.

IV. SUMMARY AND CONCLUSIONS

Recent speech technologies rely greatly on stochastic methods such as hidden Markov models (HMM) and huge amounts of speech data by which stochastic models are trained (Rabiner and Juang, 1993). In this modern approach, the contextual variability of phonemic features, such as formants or spectral patterns, is largely expressed by the variation of speech units. Such a stochastic approach has also been adopted to recover movements of the articulators from speech (Hiroya and Honda, 2004).

On the other hand, articulatory-based models were developed to explore the physiologic (Kakita *et al.*, 1985), kinematic (Mermelstein, 1972), and acoustic (Flanagan, 1972) processes of human speech production. In the articulatory-based approach, contextual variability are explained as the change of articulatory behavior adapted to the phonemic context, and the variability of acoustic features are then explained using an acoustic model of the vocal tract as a deterministic process. Such variability has been explained based on the superposition of phonemic patterns (Broad and Clermont, 1987; Saltzman and Munhall, 1989). Also, the dimensionality of the articulatory targets and smoothness of the articulatory movements have been explicitly formulated as a dynamic articulatory model (Kaburagi and Honda, 1996, 2001).

By combining the dynamic articulatory model with statistical methods for estimating the temporal structure of articulatory movements and determining the vocal-tract spectrum from the articulatory positions, this paper presented a model for generating the vocal-tract spectra based on the underlying articulatory mechanism. In the model, the linguistic information of an utterance was first converted to a sequence of phonemic tasks. Articulatory movements were then formed so that these tasks and other dynamical constraints were jointly satisfied. The spectral property of the vocal tract was finally determined from the configuration of the tract represented by the articulatory variables. In addition, experiments were conducted to determine the accuracy

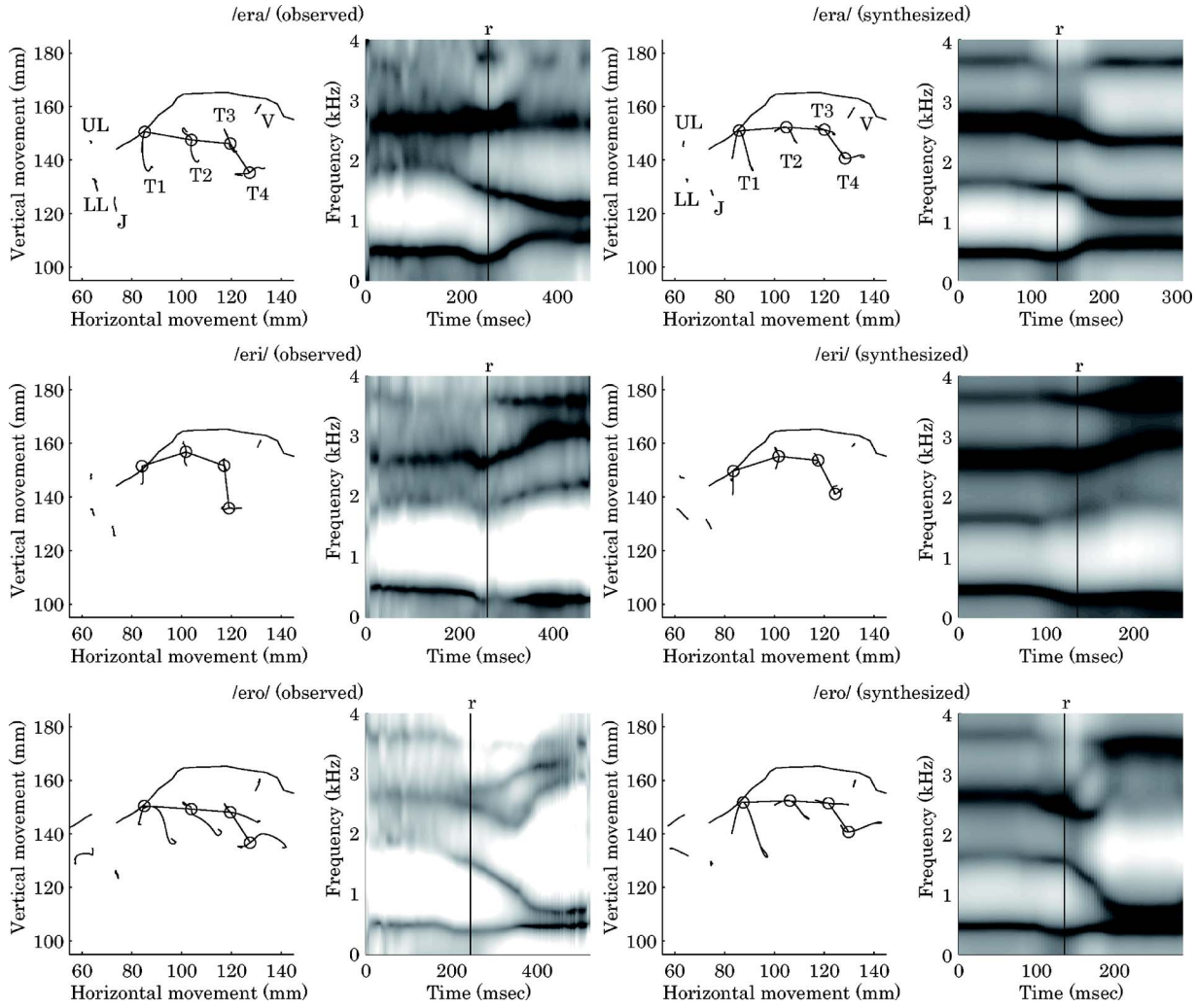


FIG. 12. Comparison of the observed and synthesized patterns of the articulatory and acoustic data for syllables /era/, /eri/, and /ero/.

and effectiveness of the proposed model. The main experimental results and related considerations are as follows:

(1) A set of articulatory and acoustic data was obtained using an electromagnetic position sensor. The sensor measured positions of eight marker points fixed on the articulators with high spatial and temporal resolutions. The vocal-tract spectrum was analyzed from speech and represented by the LSP parameters.
(2) The articulatory timing of each phoneme was estimated by the type-I quantification method trained using the articulatory data. The accuracy of this method was experimentally examined, and it was shown that the estimation error can be reduced by considering the type of the central, preceding, and following phonemes. The articulatory targets were represented using the invariant features (Kaburagi and Honda, 2001) and also trained by the articulatory data.
(3) After the articulatory movements were generated, the vocal-tract spectrum was estimated by finding appropriate samples of the articulatory-acoustic data pairs in terms of the phonemic condition and articulatory distance. Experiments showed that the spectral error of the estimation method was about 3 dB on average.
(4) Spectral generation experiments were performed for a number of utterances and the results were compared with the measured data. The model was capable of reproducing the variability of the phonemic features of both the articulatory movements and speech acoustics.

In the model, articulatory variables were defined as the positions of fixed points on the articulators. When compared to the articulatory parameters studied thus far (Mermelstein, 1972), they are not so instinctive and cannot provide enough information to represent the configuration of the entire vocal tract. However, their direct link with the electromagnetic observation method was effectively utilized for learning the phonemic tasks and estimating the vocal-tract spectra. Construction of a huge articulatory database is still difficult because of the instability of marker coils and the burden to the subject. However, the basic units of our model are phoneme specific, and therefore they were obtained from a relatively small amount of articulatory data. Our model can simultaneously generate articulatory movements and a sequence of vocal-tract spectra, thus providing a framework for constructing an auditory-visual speech communication system. Further study will be carried out to extend the model using

TABLE II. Simulation errors of the utterances shown in Figs. 8–12. Articulatory error was calculated as the Euclidean distance of the articulatory variables between the simulated and observed movements at each moment of an utterance. Then the error was averaged for the entire utterance. Spectral error was calculated using a 30th order cepstrum distance and averaged for each utterance. In (B), each numerical shows the error for when the vocal-tract spectrum was estimated from the observed articulatory movements. On the other hand, spectral estimation was performed for the generated articulatory movements in (C). Note that the articulatory timings were extracted from the observed articulatory movements in this experiment.

	(A) Articulatory error (mm)	(B) Spectral error (dB)	(C) Spectral error (dB)
/hanayaka…/	130	2.76	3.64
/eba/	1.11	2.11	2.87
/ebi/	0.99	2.66	4.15
/ebo/	1.64	2.85	3.66
/agi/	1.54	2.53	4.39
/ago/	1.70	2.66	3.67
/ega/	1.20	2.71	3.93
/ema/	1.17	2.47	3.19
/emi/	1.29	2.69	3.93
/emo/	1.35	2.22	2.80
/era/	0.93	2.03	2.53
/eri/	0.89	2.71	3.09
/ero/	1.36	2.61	2.97
Mean	1.27	2.53	3.44

three-dimensional articulatory variables. To this end, development of a three-dimensional observation system (Kaburagi *et al.*, 2005) will also be performed.

ACKNOWLEDGMENTS

We thank Mr. Takemi Mochida for his support given in recording the acoustical and articulatory data at NTT Communication Science Laboratories. This research was partly supported by the Grant-in-Aid for Scientific Research from the JSPS (Grant Nos. 14101001 and 18500134).

Blumstein, S. E. (**1986**). "On acoustic invariance in speech," in *Invariance and Variability in Speech Processes*, edited by J. S. Perkell and D. H. Klatt (Lawrence Erlbaum, N.J.).

Coker, C. H., and Fujimura, O. (**1966**). "Model for specification of the vocaltract area function," J. Acoust. Soc. Am. **40**, 1271.

Coker, C. H. (**1976**). "A model of articulatory dynamics and control," Proc. IEEE **64**, 452–460.

Broad, D. J., and Clermont, F. (**1987**). "A methodology for modeling vowel formant contours in CVC context," J. Acoust. Soc. Am. **81**, 155–165.

Engwall, O. (**1999**). "Modeling of the vocal tract in three dimensions," Proc. Eurospeech **1999**, 113–116.

Farnetani, E. (**1989**). "V-C-V lingual coarticulation and its spatiotemporal domain," in *Speech Production and Modeling*, edited by W. J. Hardcastle and A. Marchal (Kluwer, Dordrecht).

Flanagan, J. L. (**1972**). *Speech Analysis, Synthesis, and Perception*, 2nd ed. (Springer-Verlag, Berlin).

Furui, S. (**2000**). *Digital Speech Processing, Synthesis, and Recognition*, 2nd ed. (Marcel Dekker,).

Gay, T. (**1977**). "Articulatory movements in VCV sequences," J. Acoust. Soc. Am. **62**, 185–193.

Hayashi, C. (**1954**). "Multidimensional quantification—With the applications to analysis of social phenomena—," Ann. Inst. Stat. Math. **5**, 121–143.

Hiroya, S., and Honda, M. (**2004**). "Estimation of articulatory movements from speech acoustics using an HMM-based speech production model," IEEE Trans. Acoust., Speech, Signal Process. **12**, 175–185.

Jelinek, F. (**1997**). *Statistical Methods for Speech Recognition* (MIT Press, Cambridge, Mass.).

Kaburagi, T., and Honda, M. (**1994**). "Determination of sagittal tongue shape from the positions of points on the tongue surface," J. Acoust. Soc. Am. **96**, 1356–1366.

Kaburagi, T., and Honda, M. (**1996**). "A model of articulator trajectory formation based on the motor tasks of vocal-tract shapes," J. Acoust. Soc. Am. **99**, 3154–3170.

Kaburagi, T., and Honda, M. (**2001**). "Dynamic articulatory model based on multidimensional invariant-feature task representation," J. Acoust. Soc. Am. **110**, 441–452.

Kaburagi, T., and Honda, M. (**2002**). "Electromagnetic articulograph based on a nonparametric representation of the magnetic field," J. Acoust. Soc. Am. **111**, 1414–1421.

Kaburagi, T., Wakamiya, K., and Honda, M. (**2005**). "Three-dimensional electromagnetic articulography: A measurement principle," J. Acoust. Soc. Am. **118**, 428–443.

Kakita, Y., Fujimura, O., and Honda, K. (**1985**). "Computation of mapping from muscular contraction patterns to formant patterns in vowel space," in *Phonetic Linguistics*, edited by V. A. Fromkin (Academic, New York).

MacNeilage, P. F., and DeClerk, J. L. (**1969**). "On the motor control of coarticulation in CVC monosyllables," J. Acoust. Soc. Am. **45**, 1217–1233.

Maeda, S. (**1982**). "A digital simulation method of vocal-tract system," Speech Commun. **1**, 199–229.

Mermelstein, P. (**1972**). "Articulatory model for the study of speech production," J. Acoust. Soc. Am. **53**, 1070–1082.

Öhman, S. E. G. (**1966**). "Coarticulation in VCV utterances: Spectrographic measurements," J. Acoust. Soc. Am. **39**, 151–168.

Perkell, J. S., Cohen, M. H., Svirsky, M. A., Matthies, M. L., Garabieta, I., and Jackson, M. T. T. (**1992**). "Electromagnetic midsagittal articulometer (EMMA) systems for transducing speech articulatory movements," J. Acoust. Soc. Am. **92**, 3078–3096.

Rabiner, L., and Juang, B. H. (**1993**). *Fundamentals of Speech Recognition* (Prentice-Hall International, Englewood Cliffs, N.J.).

Rokkaku, M., Hashimoto, K., Imaizumi, S., Niimi, S., and Kiritani, S. (**1986**). "Measurements of the three-dimensional shape of the vocal tract based on the magnetic resonance imaging technique," Ann. Bull. RILP **20**, 47–54.

Saltzman, E. L. (**1979**). "Levels of sensorimotor representation," J. Math. Psychol. **20**, 91–163.

Saltzman, E., and Munhall, K. G. (**1989**). "A dynamical approach to gestural pattering in speech production," Ecological Psychol. 1, 333–382.

Sharf, D. J., and Ohde, R. N. (**1981**). "Physiological, acoustic, and perceptual aspects of coarticulation: Implications for the remediation of articulatory disorders," in *Speech and Language: Advances in Basic Research and Practice, Vol. 5*, edited by N. J. Lass (Academic, New York).

Sondhi, M. M., and Schroeter, J. (**1987**). "A hybrid time-frequency domain articulatory speech synthesizer," IEEE Trans. Acoust., Speech, Signal Process. **35**, 955–967.

Stevens, K. N., and House, A. S. (**1963**). "Perturbations of vowel articulations by consonantal context: An acoustical study," J. Speech Hear. Res. **6**, 111–128.

Stone, M. (**1990**). "A three-dimensional model of tongue movement based on ultrasound and x-ray microbeam data," J. Acoust. Soc. Am. **87**, 2207–2217.

Voicing and silence periods in daily and weekly vocalizations of teachers

Ingo R. Titze
National Center for Voice and Speech, The Denver Center for the Performing Arts, Denver, Colorado 80204 and Department of Speech Pathology and Audiology, University of Iowa, Iowa City, Iowa 52242-1012

Eric J. Hunter
National Center for Voice and Speech, The Denver Center for the Performing Arts, Denver, Colorado 80204

Jan G. Švec
National Center for Voice and Speech, The Denver Center for the Performing Arts, Denver, Colorado 80204 and the Groningen Voice Research Lab, Department of Biomedical Engineering, University of Groningen Medical Center, Groningen, the Netherlands

(Received 9 August 2005; revised 13 October 2006; accepted 17 October 2006)

The National Center for Voice and Speech (NCVS) data bank on voice dosimetry was used to study the distributions of continuous voicing periods and silence periods in 31 teachers over the duration of two weeks. Recordings were made during all awake hours of the day. Voicing periods were grouped into half decades, ranging from 0.0316 to 0.10 s for the shortest periods of phonation to 31.6–100 s for the longest periods of phonation. Silence periods were grouped into similar half decades, but ranged up to periods of several hours. On average, the teachers had 1800 occurrences of voicing (onset followed by offset) per hour at work and 1200 occurrences per hour while not at work. Voicing occurred 23% of the total time at work, dropping to 13% during off-work hours and 12% on weekends. The greatest accumulation of voicing occurred in the 0.316–1.0 s voicing periods, whereas the greatest accumulation of silence occurred in the 3–10 s silence periods. The study begins to lay the groundwork for understanding vocal fatigue in terms of repetitive motion and collision of tissue, as well as recovery from such mechanical stress. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2390676]

PACS number(s): 43.70.Fq, 43.70.Bk, 43.70.Jt [BHS] Pages: 469–478

I. INTRODUCTION

Teaching is one of the most vocally demanding occupations. The vocal load of teachers is beginning to be quantified thoroughly and systematically. Voice accumulation and voice dosimetry devices have been developed for monitoring vocal use at work (Airo *et al.*, 2000; Buekers *et al.*, 1995; Cheyne *et al.*, 2003; Granqvist, 2003; Ohlsson *et al.*, 1989; Popolo *et al.*, in press; Švec *et al.*, 2003; Szabo *et al.*, 2001). Analyses of teachers' voices, recorded over the course of a working day, have been carried out in order to detect changes in voice quality and parameters such as the fundamental frequency and sound pressure level (Rantala *et al.*, 1998; Södersten *et al.*, 2002; Szabo *et al.*, 2003). Voice accumulation times and the voicing percentages relative to total time at work have been found to be higher in teachers than in other professions. Masuda *et al.* (1993) measured a voicing percentage of 21% for teachers in an 8 h workday, compared to 7% for office workers. Sala *et al.* (2002) found that the average *speaking* time of day care center teachers was 40% of the time at work, compared to 28% for nurses. But this speaking time is not equal to voicing time, given that speech carries many unvoiced segments. Comparing the Masuda *et al.* results to the Sala *et al.* results suggests that voicing time is about half of speaking time.

Ten healthy female preschool teachers in daycare centers studied by Södersten *et al.* (2002) had a voicing percentage of 17% as determined by a voice accumulator. This result is basically in agreement with Masuda *et al.*, considering a somewhat different teacher population. Watanabe *et al.* (1987) found the voicing percentage of 20 normal adults (not teachers) to be 11%, while that of 30 children was 16%. While the measures were collected over several hours and several days, an exact number of recording hours was not given. Although data sets across the globe are beginning to converge, it appears that there still exists some uncertainty about voicing percentages among teachers and other voice user populations.

If a fundamental frequency of 150 Hz is assumed for a mean between male and female teachers, 1 h of voicing time (17% of 6 h of teaching) translates to about half a million vocal fold collisions per day. These collisions may be a contributor to one aspect of vocal fatigue, the continual deformation of lamina propria material in the vocal folds (Gotaas and Starr, 1993; Sapir *et al.*, 1993; Smith *et al.*, 1997; Morton and Watson, 1998; Russell *et al.*, 1998; Yiu, 2002; Roy *et al.*, 2004; Vilkman, 2004; Popolo *et al.*, 2004).

In the abovementioned studies, the primary focus has been on the accumulated phonation time, also referred to as the vocal load. What is missing is an equal focus on recovery

time, or the off load. One might argue that, on a daily basis, recovery time is simply 24 h minus the accumulated phonation time. Although such a calculation is correct, it does not consider short-term recovery from moments of silence between speech segments. Also, the largest number of recording days thus far has not been adequate to distinguish weekday loads from weekend loads, nor have the at-work versus the not-at-work differences in the same corpus of teachers been documented. This study is an attempt to fill some of these gaps.

We suspect that vocal fatigue in teachers is attributable to the fact that teachers instruct primarily in a monologue style, which allows little vocal rest in a typical class period. We further suspect that total voice accumulation over a typical workday may not tell the whole story. The *distribution* of rest periods may play an important role. For example, 4 h of near-monologue speech may be more injury prone to vocal fold tissues than 8 h of dialogue with frequent turn taking. At this point in time, however, it is not known what constitutes a minimum rest period for tissues to experience any degree of recovery. On the one hand, it could be as little as a few seconds to a few minutes if increased blood circulation occurs during voice rest (Švec and Sram, 2001) or if internal tissue fluid redistribution is a critical factor (Fisher *et al.*, 2001). On the other hand, it could be as much as several days if epithelial cells or extracellular matrix products of the vocal fold lamina propria need to be regenerated (Gray *et al.*, 1987).

Given this current state of uncertainty in the duration of the recovery process, it seems prudent to begin with the simple question of how voicing periods and rest periods are distributed in a typical teacher's workday. We expect that short rest periods are more frequent than long rest periods, but is the relation simply inverse? In continuous speech, is there a rest period of a specific duration that accounts for more accumulated rest time than any other? How are voicing and silence periods distributed during work and after work? How are voicing and silence periods distributed on weekends as compared to workdays?

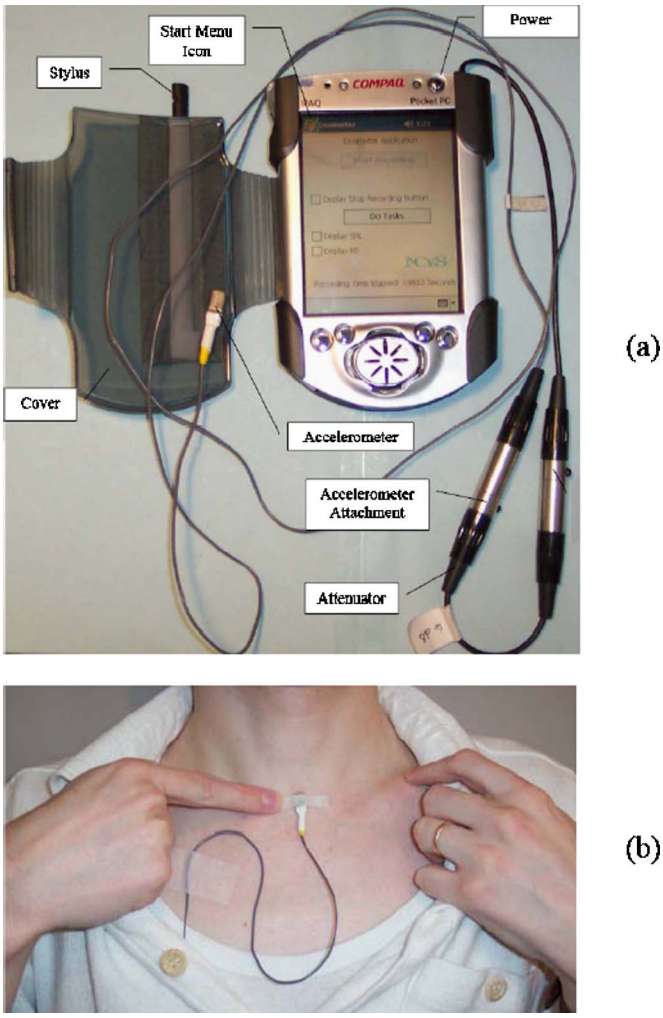


FIG. 1. (a) (Color online) The NCVS Voice Dosimeter, consisting of a modified Pocket PC with an accelerometer as the transducer. (b) The accelerometer is attached using hypo-allergenic medical adhesive and medical tape, to the sternal notch. The cabling runs underneath the clothing and the dosimeter is worn in a fanny pack.

II. METHODS

A. Field recordings

The National Center for Voice and Speech data bank on teachers was used as the primary resource. For this data bank, which is still growing, voice recordings are captured daily with the NCVS Voice Dosimeter [Fig. 1(a)] as described previously (Švec *et al.*, 2003; Švec *et al.*, 2004; Popolo *et al.*, 2004). The long-range goal is to recruit 80 teachers and engage each of them in 14 consecutive days of recording. The dosimeter's transducer (an accelerometer) is attached to the teacher's neck, at the sternal notch, using a special medical adhesive (Mastisol®, Ferndale Laboratories); additionally, adhesive tape is used to secure the transducer for an entire day, as shown in Fig. 1(b). The dosimeter can record up to 24 h of real-time processed data (with the help of an external battery) before needing to be recharged, enough time for a normal day of speaking. The dosimeter captures raw acceleration data at a sampling frequency of 11,025 Hz and processes the data in 30 ms intervals. Calculations of voicing decisions, skin acceleration level (which is converted to an estimate of sound pressure level, Švec *et al.*, 2005), fundamental frequency, and voice duration are then stored every 30 ms. Figure 2 shows an example of a dosimeter recording for test utterances we ask the teacher to give several times a day. The test utterances are performed softly and with a high fundamental frequency to test the equipment and to elicit statements about vocal fatigue, although vocal fatigue is not discussed in detail in this paper. In Fig. 2, silence is followed by sustained phonation at 550 Hz, followed by an upward pitch glide, followed by five repeated syllables /hi/ at 600 Hz, followed by a portion of the "Happy Birthday" song, followed by silence, and finally by counting "one, two, three." These recordings offer a rich data stream to study short-term, intermediate-term, and long-term voicing periods and recovery periods from vibrational exposure. Specifically, the lowest trace in Fig. 2 is a unit step function for voicing periods (value $k_v=1.0$) and silence periods (value $k_v=0$). For any teacher who has completed the two-week

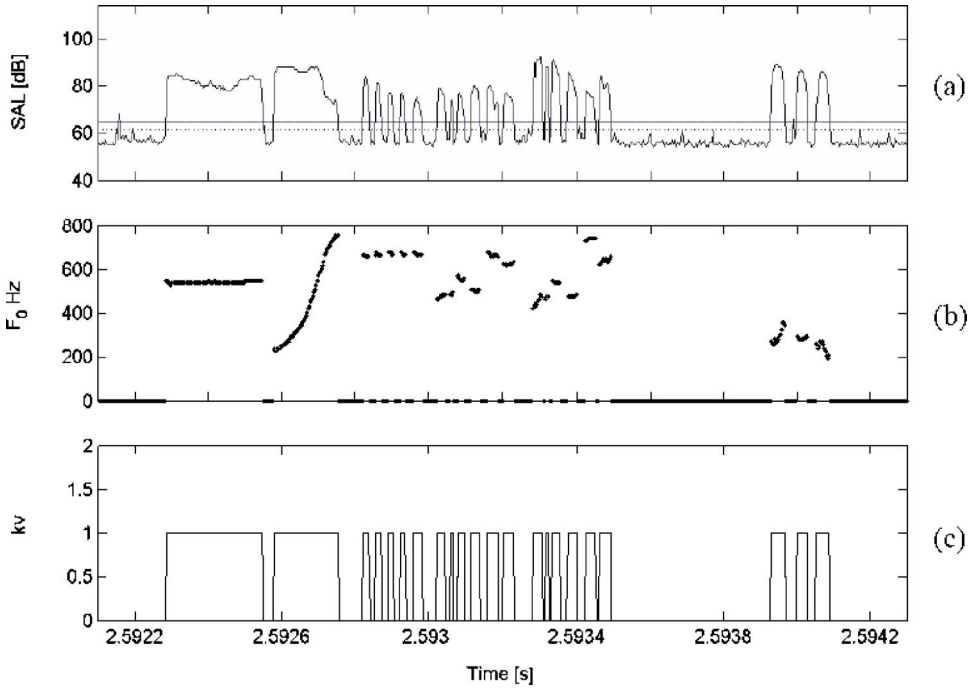


FIG. 2. A dosimeter recording: (a) skin acceleration level (SAL) in dB, (b) voicing fundamental frequency, (c) and voice/no-voice switch k_v (1 for voice, 0 for no voice).

dosimetry study, there are on the order of 100,000 data records/h, 1.4 million records per day, or 20 million records over an entire 14-day period.

B. Subjects used for the current study

To date, a complete data set has been gathered on 31 teachers. The teachers (all from public schools in six Denver metropolitan area school districts) consisted of 26 females and 5 males (the U.S. teacher population is about 20% male). Of these subjects, 23% reported having some sort of voice or speech training. Two of the subjects (both physical education instructors) use an amplification system sometime during the workday; a third teacher (not a physical education instructor) uses an amplification system every day. In a study on vocal intensity, these subjects would have been excluded, but since we are reporting only voice duration results here, the amplification is of little or no consequence. The subject breakdown by teaching grade was 68% K–6th grade, 23% Junior High (7–9th grade), and 14% High School (10–12th grade). Subject breakdown by topic was 45% general education instruction, 23% physical education instruction, 18% music/theater instruction, and 14% other (library instruction, special education, etc.). It will be shown that the results from this moderate-size corpus are statistically significant to draw preliminary conclusions about vocalization of teachers in general, but not specific teacher subclasses.

III. RESULTS

Consider k_v to be the unit step function for voicing described earlier. Mathematically, it has a value 1.0 when there is voicing and 0.0 when there is no voicing. Let voicing duration (also referred to as voicing period) or silence duration (silence period) be the independent variable, constructed in logarithmic bins ranging from 0.0316 to 0.10 s for the shortest bin duration to 3160–10,000 s for the longest bin duration. Let *occurrence* of a given duration be the dependent variable, and *accumulation* within the duration be a derived variable. Figure 3(a) shows an occurrence histogram and Fig. 3(b) an accumulation histogram from all 31 sub-

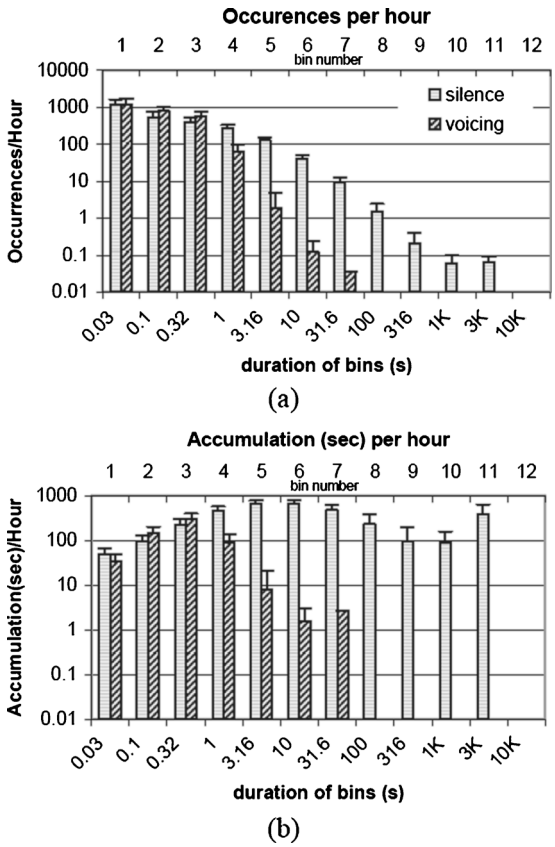


FIG. 3. Ensemble averages (31 teachers) of histograms for (a) voice and silence occurrences/h for specific durations in logarithmic bins, and (b) voice and silence accumulations in s/h for these same occurrences, over the same bin distribution.

jects. Black and white striped bars are for voicing periods and solid gray bars are for silence periods.

Collectively, the data in Fig. 3 represent 5175 h of recording over 412 days (average of 13.3±1.4 days per subject, 9.5±1.2 weekdays, 3.8±0.6 weekend days). An average of 12.5 h per day was recorded per subject, but it was not uncommon for a subject to wear the dosimeter for more than 18 h each day; the average of 12.5 h per day includes some short recording days, attributed to equipment failure or the subjects' choice not to wear the dosimeter. Over the span of a week, the daily vocal activity varied widely, as will be explained later. In Fig. 3, the vertical bar height represents occurrences per hour (part a) and the derived accumulation per hour (part b). The per hour normalization was necessary to do comparisons because recording durations differed for each day. The variables are plotted logarithmically on the vertical axis. Along the horizontal axis is the duration of bins in s, with the beginning and end of each bin labeled at the tic marks. Across the top is the bin number, labeled 1 through 11 for ease of discussion. Error bars indicate standard deviations across subjects.

Each bin in the histograms contains a half decade of logarithmic time. Thus, bin 1 contains all occurrences that are 0.0316–0.10 s long (voicing and silence periods below and up to the phonemic segmental level in speech). Bin 2 contains all occurrences that are 0.1–0.316 s long (voicing and silence periods roughly at the phonemic and syllabic level). Bin 3 contains all occurrences that are 0.316–1.0 s long (voicing and silence periods roughly at the word and phrase boundary level). Bin 4 contains all occurrences that are 1.0–3.16 s long (all-voiced sentences and voicing pauses between sentences). It is tempting to relate these durations to exact speech rhythms and pauses (Boomer and Dittman, 1962; Jaffe and Feldstein, 1970), but the division into half decades is more systematic and computer manageable at this point.

Bin 5 in Figure 3 contains all occurrences that are 3.16–10 s long (sustained phonations, as in singing, and pauses between sentences, as perhaps waiting for a response from a student). Bin 6 contains all occurrences that are 10–31.6 s long (rare exceptionally long phonations and silences typical of frequent turn taking in a dialogue). Bin 7 contains all occurrences that are 31.6–100 s long (no voicing periods, but silences that are part of less frequent turn taking in dialogue). Bin 8 contains all occurrences that are 100–316 s long (typical of a break in class to allow students to ponder or do an in-class assignment). Bin 9 contains all occurrences that are 316–1000 s long (typical of between-class silent preparations for the next class). Bin 10 contains all occurrences that are 1000–3162 s long (typical of full class periods of silence). Finally, bin 11 contains all occurrences that are 3162–10,000 s long, several hours of continuous silence.

Bin 1 deserves some special discussion. Any period (voicing or silence) less than 0.1 s is sampled by less than three consecutive data points, given that 30 ms was the sampling period of the processed data. Because it requires at least two data points to determine the shortest on-off sequence, considerable "sampling noise" may have contaminated bin 1 data. The temptation was to eliminate bin 1, but there is as yet no strong justification for it. Because the durations are so short, the overall voice accumulations were not significantly affected by bin 1. Hence we kept it in the distribution. However, we did not include bin 1 in the final totals and averages.

In bins 2 and 3, voicing occurrences are slightly greater than silence occurrences. These occurrences pertain to the voiced-unvoiced characteristics of a spoken language, in this case American English. Unvoiced consonants and short gaps between words occur almost as often as vowels and voiced consonants in continuous speech, as reported by Löfqvist and Mandersson (1987). In a reading monologue, these authors found that the voicing percentage is about 50%, with the silences being divided into unvoiced segments (15%) and boundary pauses (35%). By adding up bins 1 through 3, a total of 2500 occurrences of these segmental durations/h are derived from Fig. 3(a). In other words, the vocal folds adducted and abducted almost once per second throughout the day.

For bins 4 and higher, the order-of-magnitude differences between voicing periods and silence periods diverge sharply. Continuous voicing is rare for 3 s or more, but silences for 3 s or more are abundant. Silence durations between 3.16 and 10 s (bin 5) occur about 125/h, silences between 10 and 31.6 s (bin 6) occur on the order of 40 times/h, silences between 31.6 and 100 s (bin 7) occur on the order of 10 times/h, silences between 100 and 316 s (bin 8) occur on the order of 1.5/h, and silences above 316 s (more than 5 min, bin 9) occur only on the order of once per day (1–2 every 10 h).

A. At-work and not-at-work comparisons

Each subject kept a daily log of their activities, from which the dosimeter data were subdivided into *at work* and *not at work*. The at-work division included all times at school, meetings after classroom teaching, coaching, performances and all other school-related extracurricular activities. This resulted in over 2170 total recording hours with an average of 70±15.5 h per subject. Not-at-work time was any other time the dosimeter was active, including evenings and weekends, resulting in over 3000 total recording hours with an average of 96.8±22.5 h per subject.

Comparing the occurrences of voicing periods/h at work to voicing periods/h not at work (Fig. 4(a)), the histograms show a 30% reduction in occurrences of short periods, from about 3000 occurrences/h of voicing collectively in bins 1–3 at work to about 2000 occurrences/h of voicing in the same three bins not at work. However, as the bin duration gets larger (e.g., bins 5 and 6), not-at-work occurrences of voicing become equal to or overtake the at-work occurrences.

The larger bins could be influenced by the test utterances we ask the teacher to give several times a day (recall Fig. 2). These utterances include a sustained pitch and a pitch glide (both 2–4 s in length). The test utterances are performed softly and with a high fundamental frequency to test vocal fatigue. Since the subjects produced six voice tests per day on average, there would be a minimum of

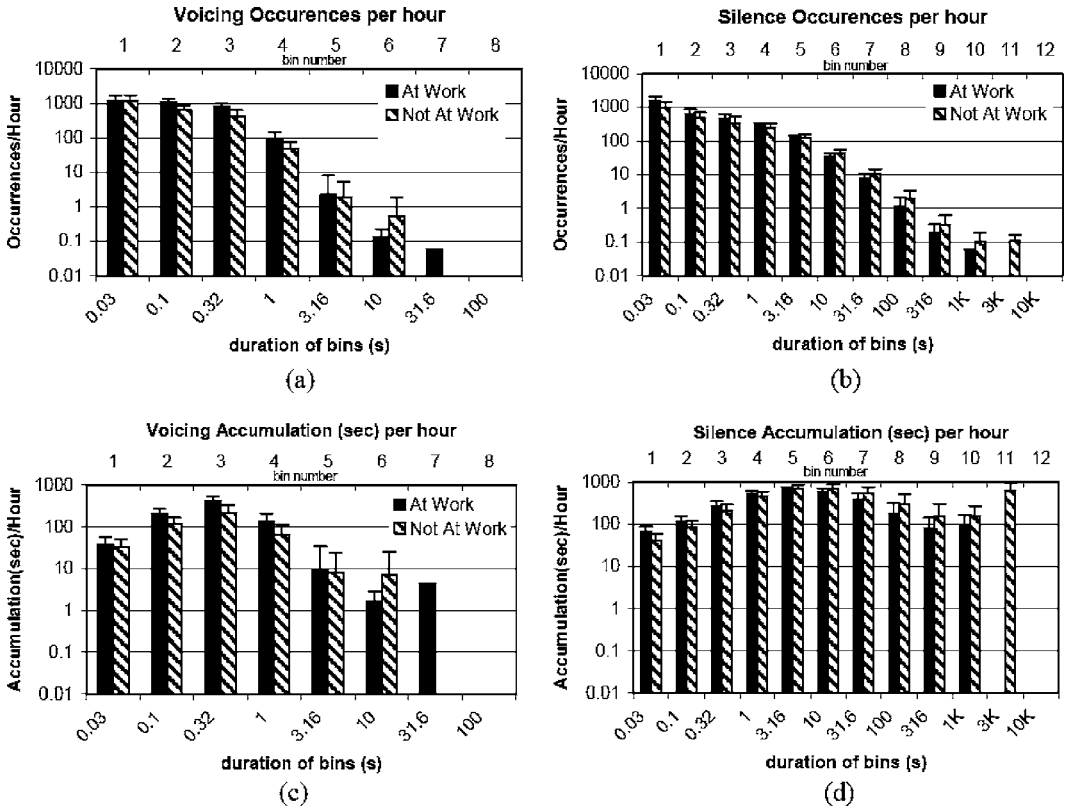


FIG. 4. Ensemble averages (31 teachers) comparing at-work and not-at-work histograms of (a) voice occurrences/h, (b) silence occurrences/h, (c) voice accumulation in s/h, and (d) silence accumulation in s/h.

12 phonations per day (or about one/h) between 2–4 s long (bins 4 and 5). But it is not uncommon for the high-pitched, soft sustained phonations to have voice breaks because of vocal fatigue; therefore, a 3 s sustained pitch with a break may appear as two 1.5 s phonations. This would double the occurrences in bin 4 from test utterances (from 12 to 24/day), but would still not come close to accounting for the 100 occurrences/h measured in bin 4. Thus, we conclude that teachers must be vocalizing long duration speech or nonspeech (e.g., singing, cheering, or shouting) as part of their teaching material, especially for occurrences in bin 5 and higher.

The longer voicing periods in bins 6 and 7 are not typical of normal speech and happen rarely (bin 6, about once every 10 h; bin 7 about every two days). Given that several of the subjects were music and theatre teachers, and that there are instances where long vocalizations may be used in other teaching examples, it is not surprising that there are some instances of long periods. Because of the low occurrence of such voicing durations, however, no statistical differences between at work and not at work in bins 5–7 could be determined. By comparison, differences between not-at-work occurrences and at-work occurrences in bins 2–4 were highly significant, t(31, single tailed), $p<0.001$.

Consider now the silence occurrences at work and not at work (Fig. 4(b)). The most interesting result is the gradual increase of the not-at-work bin height (relative to the at-work bin height) with higher bins. Not-at-work time favors long silence periods, while at-work time favors short silence periods. At bin 5, the occurrences are nearly equal. Differences between not-at-work occurrences and at-work occurrences were different for all bins, t(31), $p<0.01$. Consider the 1–15 min silence periods (bins 8 and 9), which occur twice as frequently not at work as at work. These silences may be important for teachers in terms of muscle fatigue recovery. There were about two of these silences/h not at work, but most of them less than 5 min in length. Occurrences of 5–15 min rest periods (bin 9) occurred only about once every 6 h at work and once every 3 h when not at work. These long durations could be slightly underestimated, however, due to accelerometry artifacts (e.g., from physical activities such as swallowing or bumping one's neck). These artifacts, in the form of disruptions of voicing or silence, cannot be eliminated completely from current dosimetry. Our measurements indicate, however, that accelerometry artifacts account for only about 2% of the detected voicing durations (unpublished).

Since there is a general inverse relation between the number of occurrences of a voicing period and its duration (i.e., all shorter occurrences are more frequent than the longer ones), a reasonable question is whether or not the distributions become flat if the total *integrated* time of phonation or rest is plotted for each bin. This is done in the lower part of Fig. 4 (c and d). The accumulations of voicing and voice rest offer a different perspective. The distributions are not flat, but unimodal up to bin 10. Again, because bin 11 is poorly sampled due to low occurrence, there is less confidence in the shape of the distribution at high bins. For voicing accumulation, the peak is at bin 3 and the distribution is skewed slightly to the left. The 0.316–1.0 s voicing periods

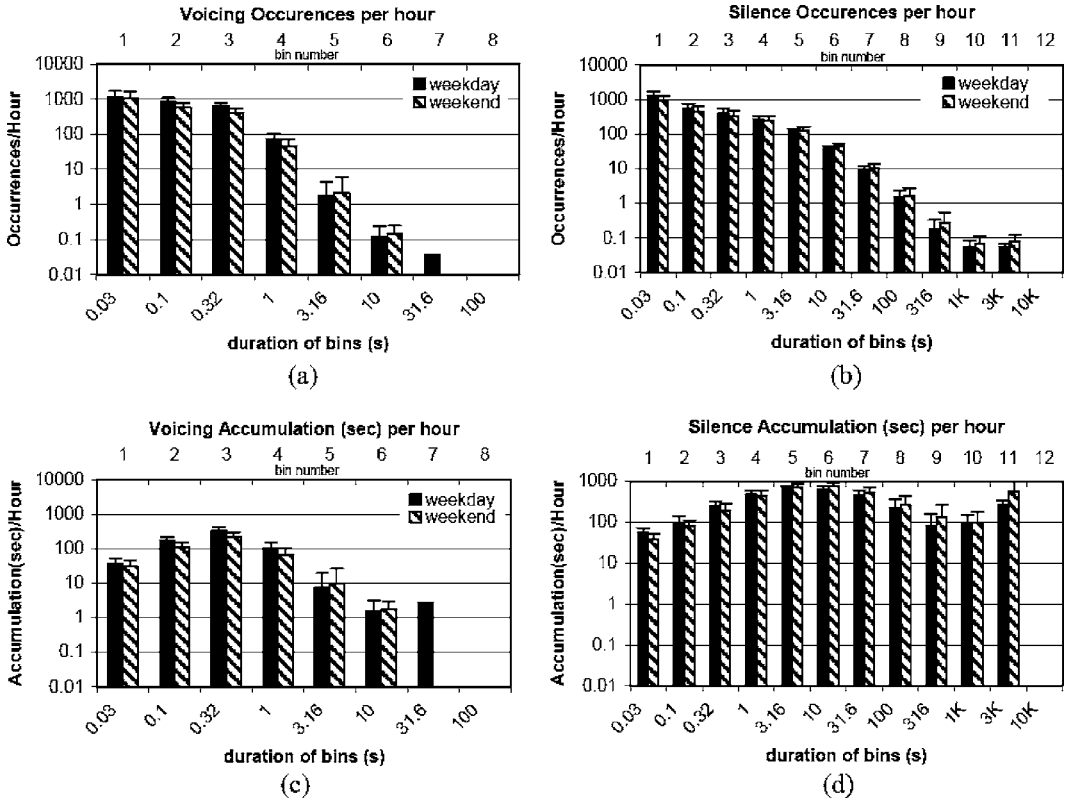


FIG. 5. Ensemble averages (31 teachers) comparing weekday and weekend histograms of (a) voice occurrences/h, (b) silence occurrences/h, (c) voice accumulation in s/h, and (d) silence accumulation in s/h.

in bin 3 (basically strings of vowels and voiced consonants) were responsible for the greatest amount of accumulated voicing. There are 460 s/h of accumulated voicing in bin 3 at work and 350 s/h not at work.

For silence accumulation (Fig. 4(d)), we also see a broad unimodal distribution, but the skewing is slightly to the right. The peak of the distribution is at bins 5–6. The 3.16–31.6 s silence periods in bins 5 and 6, which we deemed typical of dialogue turn taking, provided the greatest amount of accumulated vocal rest. Note also that bins 9–10 are larger than bins 1–2 for not-at-work hours, which is not true for the at-work hours. After work, the teachers get more accumulated rest from the long silences than from the short silences.

B. Workdays and weekend comparisons

Of further interest is the comparison between work days and nonwork days (i.e., *weekdays* and *weekends*). Figure 5 shows this comparison. If we assume that teacher weekend vocal loads are typical of average nonteacher vocal loads, then these numbers provide the internal control for voicing on the job to normal (nonoccupational) voicing. In general, the weekend histograms are similar to the not-at-work weekday histograms (compare all the striped bars in Fig. 5 to all the striped bars in Fig. 4). Weekday occurrences of voicing and weekday voice accumulations are all higher than their weekend counterparts for bins 1–4 (Figs. 5(a) and 5(c)), but for larger voicing periods (bins 5–7) there is a reversal. This may again be attributed to nonspeech vocalizations, or poor sampling of the occurrences of these long periods. Aside

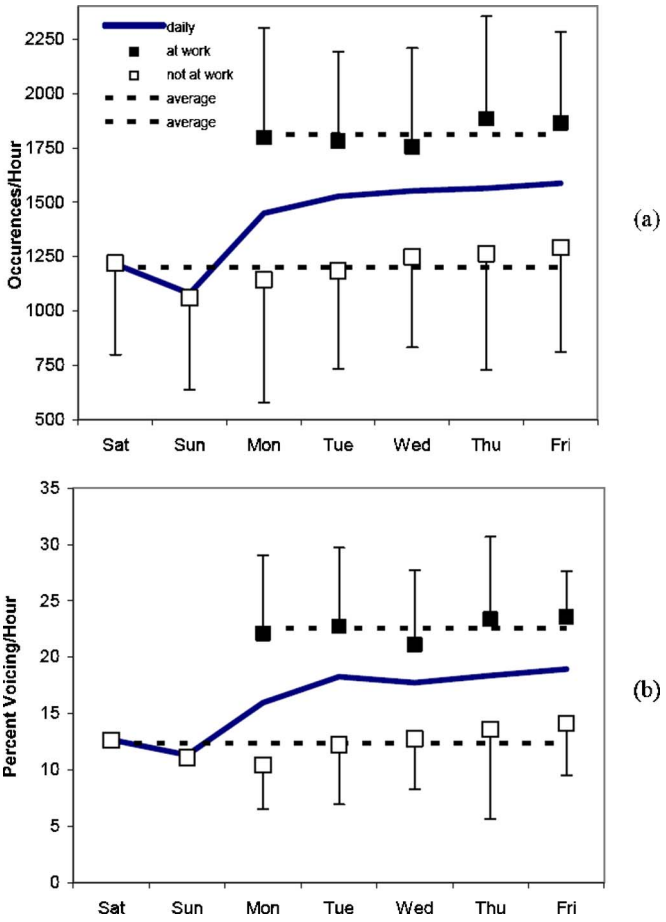


FIG. 6. (Color online) Ensemble averages (31 teachers) of (a) voicing occurrences/h and (b) voicing percentage/h over a period of a week. A two-week stretch was averaged into an equivalent one week stretch.

TABLE I. The results of each subject are presented in terms of vocal occurrences/h of recording (No. occurrences), the number of hours of recording, and number of days the subject participated. Results are displayed in terms of the weekday at work average, weekday not at work average and the weekend average. Averages and standard deviations for the overall group were shown. At Work and Not At Work averages were statistically different ($p < 0.001$) where Not At Work and Weekends were statistically similar.

	At Work		Not At Work		Weekends		No. days of recording	
Subject	No. occurrences	hrs	No. occurrences	hrs	No. occurrences	hrs	Weekday	Weekend days
F002	2712.5	71.6	1735.1	94.8	1610.5	52.8	9	5
F009	3207.8	78.9	1732.9	114.4	1838.9	51.0	10	4
F017	3409.5	35.2	2015.2	104.5	1062.5	40.6	9	4
F020	2366.9	53.6	1551.4	64.7	1296	30.4	8	4
F035	2246.8	76.3	1318.8	63.7	1584.2	46.1	8	4
F061	4353.5	64.1	3093.1	104.3	3105	36.7	10	3
F062	1554.3	78.3	1813.8	58.5	1485.4	24.3	10	4
F064	2911.6	75.6	1649.2	90.3	1549.3	36.9	10	3
F069	2568	38.8	1802.4	62.8	1967.4	29.6	5	3
F076	4161.5	63.2	2188.4	95.0	2088	53.5	10	4
F077	2802.7	100.4	2084.6	75.3	2025	45.6	10	4
F081	3890.5	60.4	1560.5	97.9	1432.9	37.2	10	4
F083	3367.6	70.9	1519.3	75.0	1530.7	27.4	8	2
F086	4505.5	83.0	2650.2	124.5	2842.9	52.4	10	4
F089	2052.8	86.8	1420.2	100.4	1732.8	42.4	10	4
F093	3908.3	87.7	3159.7	103.2	3337.4	47.2	10	4
F094	2467.3	91.3	1775.4	50.9	1960.5	52.3	7	4
F095	2919.1	55.9	2407.8	116.1	2340.7	47.7	10	4
F096	3151.6	72.9	2422.9	77.2	2288	45.7	9	5
F097	3015.9	55.6	2422.6	108.4	2160.8	43.2	10	4
F098	3216.4	56.4	2804.5	124.1	2905.6	44.8	10	4
F100	2201	80.9	1308.1	105.4	1736.8	61.8	9	3
F101	3615.1	58.9	2375.2	108.9	1720.9	51.0	10	4
F102	3050	60.7	2845.7	125.2	3083.3	56.4	10	4
F103	2833.6	73.2	2250.2	87.0	2065.1	37.8	10	4
F104	4634.1	63.5	3493.4	126.8	3791.8	51.6	10	4
M042	3572.7	70.8	2770.2	110.2	1909.1	46.5	10	4
M043	2322.9	85.4	1721.6	103.1	1765.7	46.2	10	4
M044	2655.5	68.6	1578	115.6	1548.6	43.1	10	4
M045	3680.4	99.8	2865.4	77.7	3075.9	29.6	11	3
M047	2711.1	54.4	2940.4	136.6	2776.7	54.1	10	4
avg	*3098.9*	*70.1*	*2170.2*	*96.9*	*2116.7*	*44.1*	*9.5*	*3.8*
std	*750.7*	*15.7*	*611.1*	*22.6*	*676.3*	*9.2*	*1.2*	*0.6*

from this overall lowering of the voicing occurrences and accumulations on weekends, the distributions did not change remarkably.

C. Changes over a week

Figure 6 shows ensemble averages for voicing occurrences/h and voicing percentage (of recorded time) over a one week stretch. The two-week data were averaged to get this equivalent one-week result. Voicing occurrences (Fig. 6(a)) were on the order of 1800/h at work and 1200/h not at work (bin 1 occurrences were not included in these totals and averages). The overall daily average (dark, solid curve) was 17,000 occurrences per work day, computed over an average 12.4 h recording session (7.5 h at work and 5.35 h not at work). If the not-at-work time were extended to 8 h, there would be as many as 20,000 occurrences of voicing per day for an average teacher. In terms of voicing percentage, Fig. 6(b) shows a mean of 23% at work and 13% not at work. On weekends, voicing percentage drops to 12%. However, the 1% difference between not at work and weekend was not statistically significant. The value 12%–13% is considered to be the control value for the population at large, assuming that teachers take on weekends and time away from work the same way as people in other occupations. It is close to the 11% value cited earlier (Watanbe *et al.*, 1987) for a general population of nonteachers.

Details of inter- and intra-subject comparisons, with means and standard deviations, are given in Tables I and II. Table I shows results for occurrences of voicing and silence periods, while Table II shows results for voicing percentage. At-work and not-at-work averages for both the number of occurrences and percent voicing were statistically different ($p < 0.001$), whereas not at work and weekends were not statistically different.

TABLE II. The results of each subject are presented in terms of the percentage voicing/h of recording (% voicing), the number of hours of recording, and number of days the subject participated. Results are displayed in terms of the weekday at work average, weekday not at work average and the weekend average. Averages and standard deviations for the overall group were shown. At work and not at work averages were statistically different ($p < 0.001$) where Not At Work and Weekends were statistically similar.

	At Work		Not At Work		Weekends		No. days of recording	
Subject	% voicing	hrs	% voicing	hrs	% voicing	hrs	Weekday	Weekend days
F002	20.0	71.6	9.9	94.8	8.9	52.8	9	5
F009	27.3	78.9	14.0	114.4	17.1	51.0	10	4
F017	37.5	35.2	14.6	104.5	10.7	40.6	9	4
F020	14.6	53.6	9.1	64.7	7.6	30.4	8	4
F035	20.7	76.3	9.5	63.7	11.7	46.1	8	4
F061	32.1	64.1	18.4	104.3	16.6	36.7	10	3
F062	12.8	78.3	11.9	58.5	7.8	24.3	10	4
F064	26.7	75.6	9.9	90.3	9.5	36.9	10	3
F069	20.9	38.8	13.1	62.8	14.4	29.6	5	3
F076	25.4	63.2	11.2	95.0	11.0	53.5	10	4
F077	18.8	100.4	13.0	75.3	12.8	45.6	10	4
F081	29.3	60.4	8.8	97.9	8.2	37.2	10	4
F083	27.0	70.9	9.2	75.0	10.5	27.4	8	2
F086	20.0	83.0	9.4	124.5	9.9	52.4	10	4
F089	25.6	86.8	12.0	100.4	14.3	42.4	10	4
F093	21.8	87.7	14.6	103.2	16.2	47.2	10	4
F094	20.3	91.3	10.8	50.9	12.9	52.3	7	4
F095	12.9	55.9	11.7	116.1	12.3	47.7	10	4
F096	19.3	72.9	13.8	77.2	12.6	45.7	9	5
F097	25.0	55.6	13.7	108.4	12.6	43.2	10	4
F098	20.9	56.4	14.2	124.1	16.5	44.8	10	4
F100	22.2	80.9	10.0	105.4	14.9	61.8	9	3
F101	27.8	58.9	11.0	108.9	8.1	51.0	10	4
F102	23.6	60.7	15.7	125.2	18.3	56.4	10	4
F103	15.0	73.2	7.6	87.0	6.8	37.8	10	4
F104	26.1	63.5	13.3	126.8	14.4	51.6	10	4
M042	24.2	70.8	14.7	110.2	9.9	46.5	10	4
M043	12.1	85.4	9.9	103.1	8.3	46.2	10	4
M044	20.8	68.6	9.8	115.6	10.0	43.1	10	4
M045	23.1	99.8	11.5	77.7	11.1	29.6	11	3
M047	22.8	54.4	16.9	136.6	15.7	54.1	10	4
avg	22.5	*70.1*	*12.0*	*96.9*	*13.0*	*44.1*	*9.5*	*3.8*
std	5.7	*15.7*	*2.6*	*22.6*	*3.2*	*9.2*	*1.2*	*0.6*

IV. DISCUSSION AND CONCLUSIONS

In a long-range goal of quantifying the amount of vibration exposure that teachers inflict upon themselves by talking, a first step has been taken to determine the distribution of voicing and silence periods. It was found that teachers vibrate their vocal folds 23% of the time that they teach, as opposed to 12% of the time that they do not teach. The total accumulation of voicing time is then about 2 h in an 8 h workday. This agrees with findings of others, who report the voicing percentages of teachers to be between 15% and 40% (Masuda *et al.*, 1993; Rantala *et al.*, 1994; Rantala *et al.*, 2002; Södersten *et al.*, 2002). But because voicing is not continuous for long periods of time, the distribution of voicing periods and silence periods becomes important. Overall, voicing turns on and off about 20,000 times a day for teachers. This in itself may be a fatigue factor because laryngeal adductor/abductor muscles must execute this switching on and off of the voice. There are many short on-off episodes and a few long ones. Voicing periods and silence periods are about equal in number of occurrences, about 1000/h, when their duration is less than 1 s. This is simply a function of spoken language, reflecting the number of voiced and unvoiced segments in speech. Voicing periods in the 3–10 s range are rare, reflecting nonspeech vocalizations such as singing or testing of one's voice. Silences in the 3–10 s range are abundant, on the order of 100/h. These silences, and those in the 10–100 s range, are frequent in dialogue. They represent the greatest accumulation of vocal rest during the workday. When teachers are not at work (evenings and weekends), more of the total vocal rest is attributed to longer periods of silence. Sleep time is, of course, the greatest contributor to this. We have not asked the teachers to wear the dosimeter during sleep, but informally, members of our research team have done so. They report no problem in gathering the data, but results are not included here.

A next step is to determine which of these rest periods has any profound effect on vocal fatigue recovery. The NCVS Dosimeter is designed to query the subject occasionally for perceptual cues of vocal fatigue; while a discussion of fatigue was beyond the scope of this paper, when addressed, fatigue will be analyzed statistically on the basis of the distributed occurrences of voicing and silences. Looking ahead to this in-depth discussion in the future, if fatigue is related to redistribution of fluids in the vibrating portion of the tissue (Fisher *et al.*, 2001), perhaps a rest period of a few seconds will produce some recovery. If blood circulation has been disturbed by tissue vibration, perhaps a few minutes of rest will produce recovery (Švec and Sram, 2001). If muscles have been fatigued in the voice onset-offset gestures, perhaps a fraction of an hour to several hours will produce recovery (Skof and Strojnik, 2006; Augustsson *et al.*, 2006; Albert *et al.*, 2006). Future work will answer these questions at two levels of investigation. First, prescribed levels of vocal duty ratio (e.g., 30% on, 70% off) will be elicited from volunteer vocalists over several hours. Autoperception of vocal fatigue will be tracked (Scherer *et al.*, 1987). Second, using the results from the current paper, *in vitro* studies on engineered vocal fold tissue in a bioreactor (Titze *et al.*, 2004) will be conducted to determine gene expression and corresponding indication of tissue repair due to vibration over specified periods of time. From this gene expression analysis, specific proteins will be identified that may play a role in tissue recovery.

ACKNOWLEDGMENTS

Funding for this work was provided by the National Institute on Deafness and Other Communication Disorders, Grant No. 1R01 DC04224. The authors would like to thank the research team at the National Center for Voice and Speech with many supporting roles in this work, particularly Albert Worley and Andrew Starr.

Airo, E., Olkinuora, P., and Sala, E. (**2000**). "A method to measure speaking time and speech sound pressure level," Folia Phoniatr Logop **52**(6), 275–288.

Albert, W. J., Wrigley, A. T., McLean, R. B., and Sleivert, G. G. (**2006**). "Sex differences in the rate of fatigue development and recovery," Dyn. Med. **5**(2).

Augustsson, J., Thomee, R., Linden, C., Folkesson, M., Tranberg, R., and Karlson, J. (**2006**). "Single-leg hop testing following fatiguing exercise: Reliability and biomechanical analysis," Scand. J. Med. Sci. Sports **16**(2), 111–120.

Boomer, D. S., and Dittman, A. T. (**1962**). "Hesitation pauses and juncture pauses in speech," Lang Speech **5**(4), 215–220.

Buekers, R., Bierens, E., Kingma, H., and Marres, E. H. M. A. (**1995**). "Vocal load as measured by the voice accumulator," Folia Phoniatr Logop **47**(5), 252–261.

Cheyne, H. A., Hanson, H. M., Genereux, R. P., Stevens, K. N., and Hillman, R. E. (**2003**). "Development and testing of a portable vocal accumulator," J. Speech Lang. Hear. Res. **46**, 1457–1467.

Fisher, K. V., Telser, A., Phillips, J. E., and Yeates, D. B. (**2001**). "Regulation of vocal fold transepithelial water fluxes," J. Appl. Physiol. **91**(3), 1401–1411.

Gotaas, C., and Starr, C. D. (**1993**). "Vocal fatigue among teachers," Folia Phoniatr (Basel) **45**(3), 120–129.

Gray, S. D., Titze, I. R., and Lusk, R. P. (**1987**). "Electron microscopy of hyperphonated canine vocal cords," J. Voice **1**(1), 109–115.

Granqvist, S. (**2003**). "The self-to-other ratio applied as a phonation detector for voice accumulation," Logoped. Phoniatr. Vocol. **28**(2), 71–80.

Jaffe, J., and Feldstein, S. (**1970**) *Rhythms of Dialogue* (Academic, New York).

Löfqvist, A., and Mandersson, B. (**1987**). "Long-time average spectrum of speech and voice analysis," Folia Phoniatr (Basel) **39**(5), 221–229.

Masuda, T., Ikeda, Y., Manako, H., and Komiyama, S. (**1993**). "Analysis of vocal abuse: fluctuations in phonation time and intensity in 4 groups of speakers," Acta Oto-Laryngol. **113**(4), 547–552.

Morton, V., and Watson, D. R. (**1998**). "The teaching voice: problems and perceptions," Logoped. Phoniatr. Vocol. **23**(3), 133–139.

Ohlsson, A.-C., Brink, O., and Löfqvist, A. (**1989**). "A voice accumulator—validation and application," J. Speech Hear. Res. **32**, 451–457.

Popolo, P. S., Švec, J. G., Hunter, E. J., Starr, A. C., Rogge-Miller, K., Nix, J., and Titze, I. R. (**2004**). "Teacher's guide to voice dosimetry," NCVS Online Technical Memo No. 4, October 2004: 1–31.

Popolo, P. S., Švec, J. G., and Titze, I. R., (in press). "Adaptation of a pocket PC for use as a wearable voice dosimeter," J. Speech Lang. Hear. Res..

Rantala, L., Haataja, K., Vilkman, E., and Körkkö, P. (**1994**). "Practical arrangements and methods in the field examination and speaking style analysis of professional voice users," Scand. J. Log. Phon. **19**, 43–54.

Rantala, L., Paavola, L., Körkkö, P., and Vilkman, E. (**1998**). "Working-day effects on the spectral characteristics of teaching voice," Folia Phoniatr Logop **50**, 205–211.

Rantala, L., Vilkman, E., and Bloigu, R. (**2002**). "Voice changes during work, subjective complaints and objective measurements for female primary and secondary schoolteachers," J. Voice **16**(3), 344–355.

Roy, N., Merrill, R. M., Thibeault, S., Parsa, R. A., Gray, S. D., and Smith, E. M. (**2004**). "Prevalence of voice disorders in teachers and the general population," J. Speech Lang. Hear. Res. **47**(2), 281–293.

Russell, A., Oates, J., and Greenwood, K. M. (**1998**). "Prevalence of voice problems in teachers," J. Voice **12**(4), 467–479.

Sala, E., Airo, E., Olkinuora, P., Simberg, S., Ström, U., Laine, A., Pentti, J., and Suonpää, J. (**2002**). "Vocal loading among day care center teachers," Logoped. Phoniatr. Vocol. **27**(1), 21–28.

Sapir, S., Keidar, A., and Mathers-Schmidt, B. (**1993**). "Vocal attrition in teachers: survey findings," Eur. J. Disord. Commun. **28**(2), 177–185.

Scherer, R., Titze, I. R., Raphael, B. N., Wood, R. P., Ramig, L. A., and Blager, R. F. (**1987**). "Vocal fatigue in a trained and an untrained voice user," in *Laryngeal Function in Phonation and Respiration*, edited by T. Baer, C. Sasaki, and Katherine S. Harris (College-Hill, Little, Brown, Boston, 1987), pp. 533–555.

Skof, B., and Strojnik, V. (**2006**). "Neuro-muscular fatigue and recovery dynamics following anaerobic interval workload," Int. J. Sports Med. **27**(3), 220–225.

Smith, E., Gray, S. D., Dove, H., Kirchner, L., and Heras, H. (**1997**). "Frequency and effects of teachers″ voice problems," J. Voice **11**(1), 81–87, 88.

Södersten, M., Granqvist, S., Hammarberg, B., and Szabo, A. (**2002**). "Vocal behavior and vocal loading factors for preschool teachers at work studied with binaural DAT recordings," J. Voice **16**(3), 356–371.

Švec, J. G., and Sram, F. (**2001**) "Violet-reddening of the vocal folds and the vocal fatigue," *25th World Congress of the International Association of Logopedics and Phoniatrics*, Aug. 5–9.

Švec, J. G., Popolo, P. S., and Titze, I. R. (**2003**). "Measurement of vocal doses in speech: Experimental procedure and signal processing," Logoped. Phoniatr. Vocol. **28**(4), 181–192.

Švec, J. G., Hunter, E. J., Popolo, P. S., Rogge-Miller, K., and Titze, I. R. (2004). "The calibration and setup of the NCVS dosimeter," NCVS Online Technical Memo No. 2, April 2004, version 2.4: 1–52.

Švec, J. G., Titze, I. R., and Popolo, P. S. (**2005**). "Estimation of sound pressure levels of voiced speech from skin vibration of the neck," J. Acoust. Soc. Am. **117**(3), 1386–1394.

Szabo, A., Hammarberg, B., Hakansson, A., and Södersten, M. (**2001**). "A voice accumulator device: Evaluation based on studio and fields recordings," Logoped. Phoniatr. Vocol. **26**(3), 102–117.

Szabo, A., Hammarberg, B., Granqvist, S., and Södersten, M. (**2003**).

"Methods to study pre-school teachers' voice at work: Simultaneous recordings with a voice accumulator and a DAT recorder," Logoped. Phoniatr. Vocol. **28**(1), 29–39.
Titze, I. R., Hitchcock, R. W., Broadhead, K., Gray, S. D., Li, W., Webb, K., and Tresco, P. A. (**2004**). "Design and validation of a bioreactor for engineering vocal fold tissues under combined tensile and vibrational stresses," J. Biomech. **37**(10), 1521–1529.
Vilkman, E. (**2004**). "Occupational safety and health aspects of voice and speech professions," Folia Phoniatr Logop **56**(4), 220–253.
Watanabe, H., Shin, T., Oda, M., Fukaura, J., and Komiyama, S. (**1987**). "Measurement of total speaking time in a patient with spastic dysphonia," Folia Phoniatr (Basel) **39**, 65–70.
Yiu, E. M. (**2002**). "Impact and prevention of voice problems in the teaching profession: Embracing the consumers' view," J. Voice **16**(2), 215–228.

An *in vitro* setup to test the relevance and the accuracy of low-order vocal folds models

Nicolas Ruty, Xavier Pelorson, and Annemie Van Hirtum
Institut de la Communication Parlée, UMR5009 CNRS/INPG/Université Stendhal, 46 Avenue Félix Viallet, 38031 Grenoble Cedex 01, France

Ines Lopez-Arteaga
Department of Mechanical Engineering, Dynamics and Control, Eindhoven University of Technology, P.O. Box, 513, WH-1.129, 5600 MB Eindhoven, The Netherlands

Avraham Hirschberg
Department of Applied Physics, Gas Dynamics and Aero-acoustics, Eindhoven University of Technology, P.O. Box 513, CC 2.24, 5600 MB Eindhoven, The Netherlands

(Received 28 February 2006; revised 30 August 2006; accepted 7 October 2006)

An experimental setup and human vocal folds replica able to produce self-sustained oscillations are presented. The aim of the setup is to assess the relevance and the accuracy of theoretical vocal folds models. The applied reduced mechanical models are a variation of the classical two-mass model, and a simplification inspired on the delayed mass model for which the coupling between the masses is expressed as a fixed time delay. The airflow is described as a laminar flow with flow separation. The influence of a downstream resonator is taken into account. The oscillation pressure threshold and fundamental frequency are predicted by applying a stability analysis to the mechanical models. The measured frequency response of the mechanical replica together with the initial (rest) area allows us to determine the model parameters (spring stiffness, damping, geometry, masses). Validation of theoretical model predictions to experimental data shows the relevance of low-order models in gaining a qualitative understanding of phonation. However, quantitative discrepancies remain large due to an inaccurate estimation of the model parameters and the crudeness in either flow or mechanical model description. As an illustration it is shown that significant improvements can be made by accounting for viscous flow effects.
[DOI: 10.1121/1.2384846]

PACS number(s): 43.70.Jt, 43.70.Bk [AL] Pages: 479–490

I. INTRODUCTION

Schematically, phonation can be understood as the result of a complex interaction between respiratory airflow and soft tissues. Under certain circumstances, the ongoing fluid/structure interaction can result in self-oscillation of the vocal folds. Vocal fold auto-oscillation produces the main sound source for the vocal tract during the production of voiced sounds.

Physical modeling of this process is obviously needed for many practical applications, especially in the field of voice synthesis and pathology. Although some important advances have been made in the field, full numerical models of phonation, based on finite element models for the tissues, for example (Hunter *et al.*, 2004; Vampola *et al.*, 2005), are still underused. This can be understood firstly because of the numerical complexity of such an approach, which implies, in particular, hours of computation time. Secondly, it must be noted that an accurate description of many physical effects such as the collision of the vocal folds, the way to account for unsteady boundary conditions, or the presence of turbulence in the flow, just to name a few, are far beyond the capabilities of most common numerical software.

For these reasons, low-order models, often called distributed (or lumped) models, are still very popular because of their simplicity. Such models are indeed often used for speech synthesis purposes (Flanagan *et al.*, 1975; Story and Titze, 1995; Kob, 2002), but low-order models have also been used in speech pathology studies (Wong *et al.*, 1991; Herzel *et al.*, 1994; Lous *et al.*, 1998). One major concern about these low-order models is the crudeness of the physics used in the models. This includes the fluid mechanics of the airflow through and past the glottis, the biomechanics of the vocal fold tissues, and the acoustics of the vocal tract as well as the interactions between these physical processes.

Concerning the description of the airflow, many attempts have been performed in order to test theories of varying complexity. While some attempts to measure pressure and flow *in vivo*, i.e., on human speakers, have been performed (Van den Berg *et al.*, 1957; Cranen, 1987) most experiments deal with *in vitro* setups based on a mechanical replica of the larynx. Flow measurements on a rigid nonoscillating replica of the vocal folds have been performed by van den Berg *et al.* (1957), Scherer *et al.* (1983, 2001), Pelorson *et al.* (1995), and Hofmans *et al.* (2003). More complex, and realistic, flow conditions were recently obtained using oscillating replicas of the larynx (Titze *et al.*, 1995; Barney *et al.*, 1999; Deverge *et al.*, 2003; Mantha *et al.*, 2005; Thomson *et al.*, 2005). In particular, self-oscillating replicas are of great interest because they allow us to test not only the flow description, but

also the interaction with a mechanical (eventually deformable) structure and with an acoustical resonator. This type of deformable replica is also used to study lip behavior during buzzing in musical acoustics (Gilbert *et al.*, 1998; Cullen *et al.*, 2000; Vilain *et al.*, 2003).

Another major concern with low-order models deals with the relevance of the model itself with respect to the vocal fold physiology. While some parameters of the models, like the subglottal pressure or the geometry of the glottis, can be directly compared to observations performed on humans, others, like springs or dampers in a two-mass model such as Ishizaka and Flanagan (1972), have no direct equivalent in real life. The existence and the way to derive a relationship between these parameters and human physiology is therefore a crucial problem. Some important studies (Svec *et al.*, 2000; Berry and Titze, 1996; Berry *et al.*, 1994; Kaneko *et al.*, 1983; De Vries, 1999; Childers and Wong, 1994) have been performed using either *in vivo* measurements or numerical simulations of some mechanical properties of human vocal folds. While these data are helpful to derive the mechanical parameters of low-order models, they do not allow for testing the models themselves.

In this paper we present an experimental setup designed to test low-order models of the vocal folds including the mechanics, the fluid mechanics, the acoustics, and their interactions. This setup relies on a self-oscillating replica consisting of a pair of thin latex tubes filled with water. Although it behaves in a comparable way, this replica is not intended to be viewed as a "real" larynx but rather as a test for the relevance of the theoretical models. Compared with "real life" experiments, this setup has the advantage of easy control and quantitative access to most important fluid mechanical and mechanical parameters. In particular, thanks to an optical device, the mechanical response of the replica can be measured. Using the same device, the vibration of the replica can be quantitatively observed. Two low-order models of the vocal folds are tested. First, a modification of the original two-mass vocal fold model of Ishizaka and Flanagan is presented. Second, a further simplification exploiting only one single degree of freedom to describe the tissue mechanics is considered. Next, stability analysis is applied in order to evaluate the performance of the theoretical predictions against the experimental data.

II. MODELS UNDER TEST

Two theoretical models of the complex phonatory apparatus are described. These models are severe simplifications of the fluid structure interactions appearing during human voiced sound production. The interaction between airflow and mechanical models of the vocal folds is described in Sec. II A. The two mechanical models are detailed in Sec. II B. The acoustic interaction between the voiced sound source and the vocal tract, using a linear acoustic approximation, is described in Sec. II C.

A. Flow model

The geometry of the glottal channel is schematically depicted in Fig. 1. For the sake of simplicity, the simplest approximation for the glottal flow has been chosen. This description relies on the assumption of a quasi-steady inviscid and incompressible flow within the glottis. Further, the subglottal pressure P_{sub} is assumed to be constant and independent from the other characteristic values of the system. Lastly, when the glottis forms a diverging channel, the flow is assumed to separate at a position determined by the following *ad hoc* criterion $\text{As}(t)=1.2^{*}\min(A(x,t))$ with $A(x,t)$ being the glottal cross-sectional area and $\text{As}(t)$ the glottal area at the separation point. Using these assumptions one can predict the instantaneous pressure within the glottis, $P(x,t)$, as

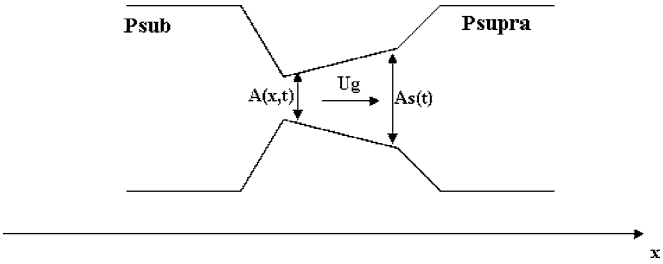


FIG. 1. Geometry used for the fluid mechanical description. Psub is the subglottal pressure, Psupra is the supraglottal pressure, $A(x,t)$ is the transversal area of the glottis along the X coordinate, U_g is the volume flow, and $\text{As}(t)$ denoted the transversal area of the glottis at separation point.

$$P_{\text{sub}} - P(x,t) = \frac{1}{2}\rho U_g^2\left(\frac{1}{A^2(x,t)}\right), \quad \text{if } x < x_s,$$

$$P(x,t) = P_{\text{supra}}, \quad \text{if } x > x_s, \tag{1}$$

where ρ indicates the constant air density, U_g is the volume flow velocity, and P_{supra} is the supraglottal pressure.

Neglecting the effects of wall vibrations, the volume flow velocity, U_g, is a constant that can be predicted from the overall pressure drop at the glottis:

$$U_g = \sqrt{\frac{2(P_{\text{sub}} - P_{\text{supra}})}{\rho(1/A_s^2)}}. \tag{2}$$

Equation (2) can be simplified as

$$U_g = A_s\sqrt{\frac{2(P_{\text{sub}} - P_{\text{supra}})}{\rho}}. \tag{3}$$

Combining Eqs. (3) and (1) the pressure distribution is written as

$$P(x,t) = P_{\text{sub}} - (P_{\text{sub}} - P_{\text{supra}})\left(\frac{A_s^2}{A^2(x,t)}\right), \quad \text{if } x < x_s,$$

$$P(x,t) = P_{\text{supra}}, \quad \text{if } x > x_s. \tag{4}$$

B. Mechanical principles

The mechanical properties of the vocal folds are represented by the distributed model depicted in Fig. 2. It is assumed that the glottis has a constant width, L_g. Applying the second principle of dynamics to these models, the following equations are obtained:

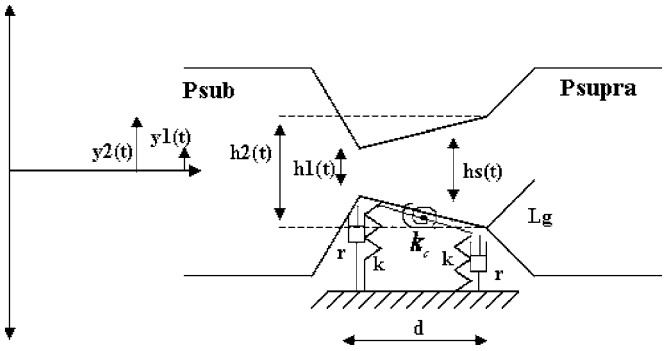


FIG. 2. Two mass model of the vocal folds. y1(t) and y2(t) are the movements on the *Y* coordinate, h1(t) and h2(t) are the apertures of the glottis, hs(*t*) is the aperture of the glottis at separation point, Lg denotes the vocal fold width, d is the length of the vocal folds, k and k_c denote the spring stiffness, and *r* is the damping.

$$\frac{m}{2}\frac{\partial^2 y_1(t)}{\partial t^2} = -k(y_1(t)-y_{10}) - k_c(y_1(t)-y_{10}-y_2(t)+y_{20}) - r\frac{\partial y_1(t)}{\partial t} + F_1(P_{\text{sub}},P_{\text{supra}},H_1,H_2), \tag{5}$$

$$\frac{m}{2}\frac{\partial^2 y_2(t)}{\partial t} = -k(y_2(t)-y_{20}) - k_c(y_2(t)-y_{20}-y_1(t)+y_{10}) - r\frac{\partial y_2(t)}{\partial t} + F_2(P_{\text{sub}},P_{\text{supra}},H_1,H_2), \tag{6}$$

where m is the effective vibrating mass of a vocal fold, k and k_c indicate the spring stiffnesses, r is the damping, d is the length of the glottis, $y_1(t)$ and $y_2(t)$ denote the positions of the two masses, y_{10} and y_{20} are the rest positions of the two masses, and F_1 and F_2 are the pressure forces along the y axis.

The pressure force distribution is derived from the description of Lous *et al.* (1998). We assume a symmetrical movement of the two vocal folds. Consequently, $H(t)=2y(t)$ and Eq. 6 become

$$\frac{\partial^2 H_1(t)}{\partial t^2} = -\frac{2k}{m}(H_1(t)-H_{10}) - \frac{2k_c}{m}(H_1(t)-H_{10}-H_2(t)+H_{20}) - \frac{2r}{m}\frac{\partial H_1(t)}{\partial t} + \frac{4}{m}F_1(P_{\text{sub}},P_{\text{supra}},H_1,H_2),$$

$$\frac{\partial^2 H_2(t)}{\partial t} = -\frac{2k}{m}(H_2(t)-H_{20}) - \frac{2k_c}{m}(H_2(t)-H_{20}-H_1(t)+H_{10}) - \frac{2r}{m}\frac{\partial H_2(t)}{\partial t} + \frac{4}{m}F_2(P_{\text{sub}},P_{\text{supra}},H_1,H_2). \tag{7}$$

This formulation can be simplified if the time delay between the movements of the two masses is assumed to be constant in time. This hypothesis is proposed by Avanzini *et al.* (2001) and as a result Eq. (7) can be simplified since only the movement of the first mass needs to be described while the movement of the second mass is supposed to follow the first one with a fixed time delay t_0. In other words, the coupling between the two masses is now described by the parameter t_0 accounting for the fixed time delay instead of the coupling spring k_c. Consequently, only a single mechanical degree of freedom remains:

$$\frac{\partial^2 H_1(t)}{\partial t^2} = -\frac{2k}{m}(H_1(t)-H_{10}) - \frac{2r}{m}\frac{\partial H_1(t)}{\partial t} + \frac{4}{m}F_1(P_{\text{sub}},P_{\text{supra}},H_1(t),H_1(t-t_0)),$$

$$H_2(t) = H_1(t-t_0). \tag{8}$$

In the following we will refer to this simple description as the delayed mass model. This approach is comparable to the model proposed by Titze (1988) and more recently by Drioli (2005). Note that if the delay t_0 equals zero, the model reduces then to a one-mass model, such as the one described by Flanagan and Landgraf (1968) and by Cullen *et al.* (2000).

A special case occurs when the vocal folds are colliding. Following Ishizaka and Flanagan (1972), this phenomenon is simply modeled as a discrete change in the spring stiffnesses and dampings:

$$k_{1e} = 4*k_1, \quad r_{1e} = r_1 + 2*\sqrt{k_{1e}*m/2} \quad \text{if } H_1 < 0,$$

$$k_{2e} = 4*k_2, \quad r_{2e} = r_2 + 2*\sqrt{k_{2e}*m/2} \quad \text{if } H_2 < 0, \tag{9}$$

where k_{1e}, k_{2e} denote the modified spring stiffness and r_{1e}, r_{2e} are the modified damping.

C. Acoustical description

Acoustical coupling with the vocal tract is only expected to be significant when the fundamental frequency of the vocal folds oscillation becomes close to a formant frequency (typically the first formant for normal speech), as described by Rothenberg (1980). Therefore, for the sake of simplicity, the acoustics of the downstream resonator is assimilated to a single degree of freedom system described by the following equation,

$$\frac{\partial^2 \psi(t)}{\partial t^2} + \frac{\omega_A}{Q_A}\frac{\partial \psi(t)}{\partial t} + \omega_A^2 \psi(t) = \frac{Z_A \omega_A}{Q_A} u, \tag{10}$$

where $\partial\psi(t)/\partial t = p$, with p the acoustic pressure at the entrance of the resonator, ω_A is a resonance pulsation of the resonator, Q_A is the quality factor of this resonance, Z_A is the peak value of impedance at resonance ω_A, and u is the acoustic airflow velocity.

In other words, this description is equivalent to the classical linear theory of the vocal tract (Flanagan, 1972), but accounts only for the first acoustical resonance of the downstream resonator. The vocal tract is simply described by the impedance of an unflanged pipe radiating in free space (Pierce, 1991), but only the first acoustical resonance of the downstream resonator is taken into account.

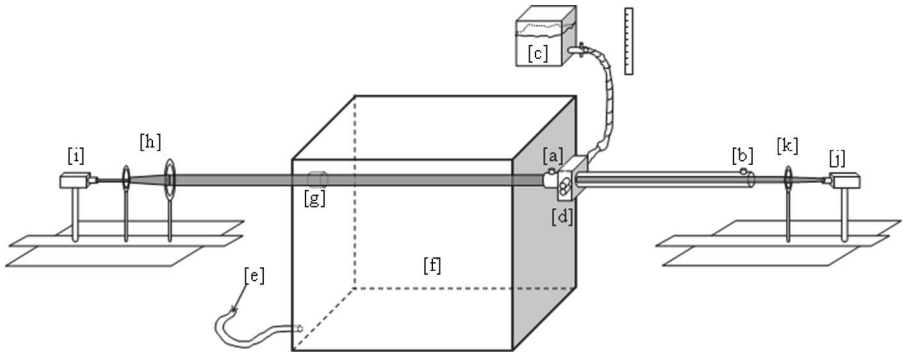


FIG. 3. Experimental setup, replicating the human phonatory apparatus. (a), (b) Pressure transducer plug. (c) Water tank supplying the fixed internal pressure inside the latex tubes of the vocal fold replica. (d) Deformable replica of the vocal folds. (e) Air supply. (f), Pressure tank, walls covered with absorbing foam. (g) Transparency window. (h), (k) Optical lenses. (i) Laser diode. (j) Photodiode.

III. EXPERIMENTAL SETUP

The theoretical models described in the previous section need to be validated. *In vivo* measurements have the advantage of being performed on real human tissues, but the lack of control and reproducibility is a severe drawback, which can harm the validation robustness. For this reason, in the literature, several *in vitro* experimental studies on relevant mechanical replicas have been performed in order to validate different aspects of the interacting theoretical models (fluid mechanical, mechanical, or acoustical). Steady rigid replicas have been used to validate the theoretical description of quasi-steady flow (e.g., Van den Berg *et al.*, 1957; Scherer *et al.*, 1983; Gauffin *et al.*, 1983; Gauffin and Liljencrants, 1988; Scherer and Guo, 1990; Scherer *et al.*, 2001; Pelorson *et al.*, 1994, 1995). More recently, the influence of wall displacement and structure deformation on the airflow have been considered by using respectively rigid mobile larynx replicas (Kiritani *et al.*, 1987; Barney *et al.*, 1999; Alipour and Scherer, 2001; Deverge *et al.*, 2003) and deformable larynx replicas (Titze *et al.*, 1995; Chan *et al.*, 1997; Thomson *et al.*, 2005). The presented setup follows the same approach. In particular, the various interactions between acoustics, mechanics, and fluid mechanics are investigated.

A. Description

The replica of the larynx is mounted in a suitable experimental setup illustrated in Fig. 3. The lungs are represented by a pressure reservoir of 0.75 m^3 fed by a compressor. The walls of the reservoir have been covered with absorbing foam in order to reduce acoustical resonances of the reservoir. The reservoir pressure is controlled by means of a Norgren pressure regulator type 11-818-987 and could be varied from a few Pa up to 3000 Pa.

A vocal fold replica, depicted in Figs. 4 and 5, is connected to the reservoir. This replica, upscaled by a factor 3 compared with real vocal folds, consists of two metal half cylinders of 12.5-mm diameter, covered with latex tubes (Piercan Ltd.) of 11-mm diameter (±0.1 mm) and 0.2 mm thickness (±10%). The two cylinders are filled with water under pressure. The water pressure, henceforth denoted the internal pressure Pc, is controlled by means of a water column. Changing the height of the column allows us to impose an internal pressure Pc up to 10 000 Pa. The value of the internal pressure Pc has an influence on the replica's mechanical characteristics, but also on the initial replica opening. Typically, for low values of Pc (e.g., 2500 Pa), the initial replica aperture is of the order of 2 mm and the latex tension is low, while for larger values of Pc (of the order of 6000 Pa), the two latex tubes are in contact, and the tension increases. The acoustics of the vocal tract can be simulated using an acoustic resonator connected downstream to the replica. For this study, two different tubes of uniform circular section (diameter 25 mm) were used. Their lengths are respectively 500 and 250 mm.

This experimental setup has been carefully designed in order to match as much as possible the order of magnitudes relevant for physical quantities during speech production. The relevance of the experimental design can be assessed using quantitative considerations based on a dimensionless analysis. The Reynolds number $\mathrm{Re}=\rho V_g h_g/\mu$, where ρ is the air density, V_g is the mean flow velocity, h_g is the replica aperture, and μ is the dynamic viscosity coefficient of air, can be used as a measure of the importance of viscosity relative to inertial forces in the flow. The Strouhal number $S_r=fd/V_g$, where f is the oscillation fundamental frequency and d is the longitudinal length of the replica, is a measure

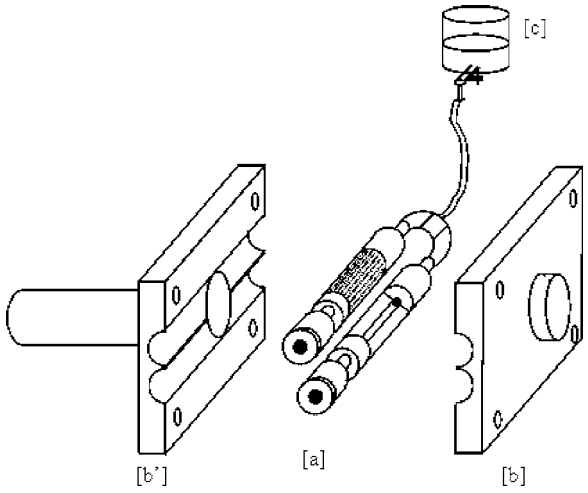


FIG. 4. Deformable replica of the vocal folds (Vilain *et al.*, 2003). Tension is controlled by the water pressure. (a) Brass half-cylinder, covered by latex tubes, and filled with water. (b), (b) Steel support.

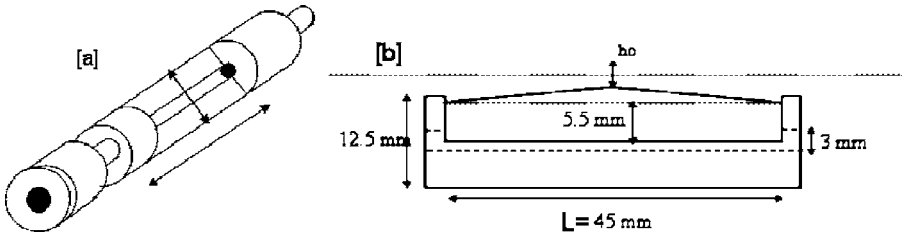


FIG. 5. Details of a vocal fold replica. (a) Isometric view of a half cylinder covered by latex tube. (b) Frontal view used to estimate the water volume filing the latex tube.

for the importance of inertia in the flow compared to convection. The Mach number $\mathrm{Ma}=V_g/c_0$, where c_0 is the speed of sound, provides a measure for the importance of compressibility effects. Other parameters, such as the geometrical aspect or the acoustical resonance frequency, are also to be considered. The correspondence between the replica and *in vivo* data is summarized in Table I.

The experimental setup allows two types of dynamical measurements: pressure and geometry. Pressure measurements are performed by means of two Kulite pressure sensors XCS-0.93-0.35-Bar-G, supplied by a Labor-Netzgerat power supply EA-3005S. They are calibrated against a manometer with typical accuracy of ∓ 5 Pa. The pressure sensors were respectively located just upstream from the vocal fold replica and at the end of the resonator tube. The geometrical measurements are performed using an optical device composed of a laser diode, supplied by a P. Fontaine Dc amplifier FTN2515. The laser beam width is increased by two convergent lenses (L1, $f=50$ mm; L2, $f=100$ mm) and passes through the air reservoir thanks to a small Plexiglas window and proceeds through the vocal folds replica. The transmitted laser beam intensity is therefore modulated by the movement of the vocal folds replica. These variations in light intensity are measured using a photo sensor BPW34, supplied by a Solartron Dc Power Supply. The optical system is calibrated by rectangular apertures with known dimensions of 0.01 mm precision.

Electrical signals are amplified and conditioned using a preamplifier/conditioning board (National Instruments SXCI-1121) connected to a PC through a National Instruments BNC-2080 Card and a National Instruments PCI-MIO-16XE acquisition card. The acquired data are processed using Labview7 software (National Instruments).

B. Determination of the model parameters

In order to compare the theoretical model predictions with the experimental data, the relationship between empirical parameters and model parameters must be known. While this relationship can be derived directly for some parameters, like the geometrical dimensions of the glottis (L_g and d), others need to be estimated.

The mass, m, used to describe the mechanics of the theoretical models is estimated by the quantity of water inside the latex tube covering the metal half-cylinder depicted in Fig. 5. The mass is estimated as

$$m_{cv} = \rho_e L \frac{\pi d_l^2}{8} \tag{11}$$

where ρ_e is the density of water, L is the replica width, and d_l is the diameter of the latex tube (11 mm). Considering the dimensions of the replica, the estimated mass of water inside one latex tube for $dl=11$ mm is $m_{cv}=2.29g$. This mass is in fact the maximum mass able to oscillate.

The spring stiffness and damping are estimated thanks to the direct measurement of the mechanical response of the replica. The experimental procedure follows the one used by Gilbert *et al.* (1998) on artificial lips. A pressure driver unit (ERS TU-100) terminated by a cone is used as an acoustical excitation, as shown in Fig. 6. For each acoustical excitation frequency the replica response is obtained by measuring the variation of the distance between the two latex tubes. The

TABLE I. Comparative table: physiological characteristics of for male adult voice [typical "*in vivo*" data obtained from Hollien and Moore (1960), Hirano *et al.* (1983), Saito *et al.* (1981), Mc Glone and Shipp (1971), and Fitch and HolBrook (1970), and Baken, (1987)] and geometrical, dynamical, and acoustical characteristics of the replica.

	Human voice	Replica
Scale	1	3
Aspect ratio, Lg/hg, mean value	20	20
Aspect ratio, Lg/d, mean value	3.3	2.5
Operating pressure, range of values	300–1000 Pa	200–1000 Pa
Frequency of oscillations F0, range of values	80–200 Hz	110–170 Hz
Reynolds number, Re, range of values	700–1300	800–2500
Strouhal number, Sr, mean value	0.01–0.02	0.03–0.05
March number, Ma, range of values	0.06–0.12	0.05–0.12
First acoustical resonance	300–1000 Hz (range of values)	170 Hz (500 mm) 340 Hz (25 mm) (discrete values for each downstream resonator)

Estimated from subglottal pressure and aspect ratio.

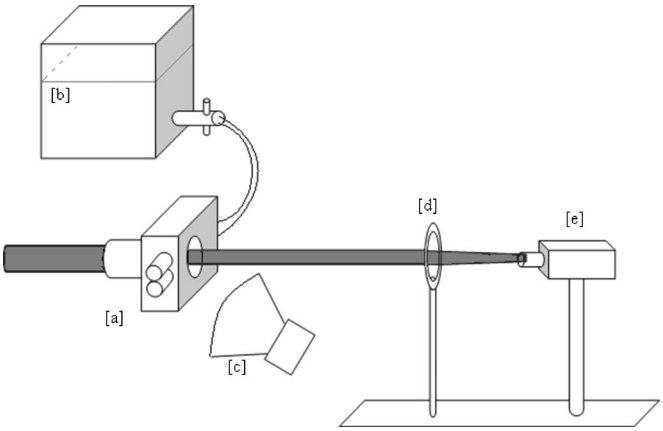


FIG. 6. Experimental setup to measure the mechanical response. (a) Deformable replica of the vocal folds. (b) Water reservoir. (c) Loud speaker, producing the acoustical excitation. (d) Optical lens. (e) Photo diode.

response of the replica is then plotted against the excitation frequency (between 100 and 400 Hz in steps of 1 Hz). This procedure is repeated for each internal pressure Pc between 500 and 6500 Pa in steps of 500 Pa.

A typical example of mechanical response of the replica is depicted in Fig. 7. Note that the retrieved response is quite similar to the mechanical responses observed by Svec *et al.* (2000) on *in vivo* human vocal folds.

From the measured mechanical responses, the resonance frequencies, ω, and associated quality factors, Q, can be easily extracted. The obtained parameters can be related to the natural resonance behavior of the theoretical model of the vocal folds and so to the spring stiffness and damping:

$$\omega_0 = \sqrt{\frac{2k}{m}}, \quad Q_0 = \frac{m\omega_0}{2r}, \tag{12}$$

where ω_0 is the resonance pulsation, Q_0 is the quality factor of this resonance, k is the spring stiffness, $m=m_{cv}/2$, is the

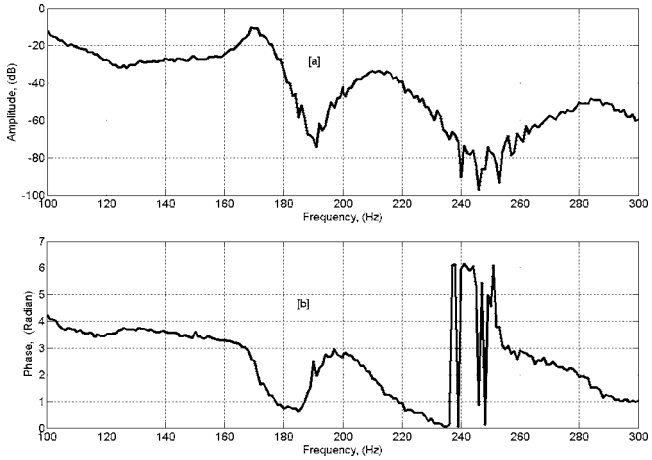


FIG. 7. Mechanical response of the vocal fold replica, internal pressure 2500 Pa. (a) Amplitude of the mechanical response. (b) Phase of the mechanical response.

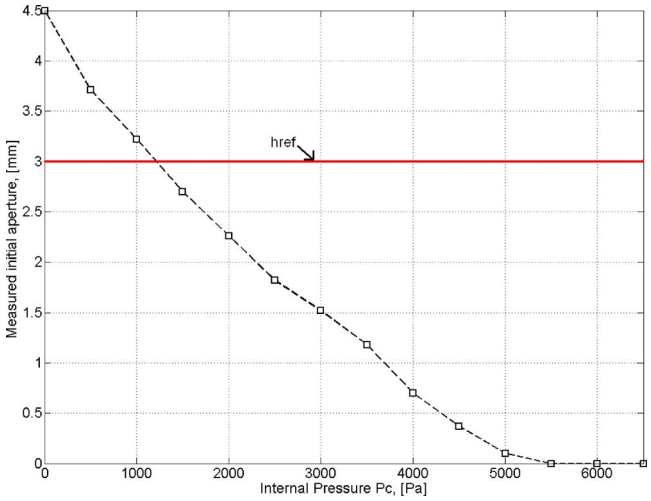


FIG. 8. Measured initial aperture of the vocal fold replica plotted as a function of the internal pressure Pc, varying between 0 and 6500 Pa.

effective mass of a vocal fold, and r is the damping. The coupling spring stiffness k_c is imposed as $k/2$, as described by Lous *et al.* (1998).

When large values of the internal pressure Pc (>5000 Pa) are involved, the frequency response cannot be measured using the above procedure since the two latex tubes are in contact (Fig. 8). For these values of Pc, one of the two latex tubes is removed and the mechanical response of a single tube is measured. The measured response in this case is obviously a very crude approximation since it does not account for the nonlinearity introduced by the contact between the two latex tubes.

IV. RESULTS AND DISCUSSION

In this section, we will detail the comparison between the predictions of the theoretical models and experimental data with respect to two quantitative phonation parameters: the on-set and off-set pressure of oscillation and the fundamental frequency. The measurements are discussed in Sec. IV A and the simulations in Sec. IV B. The comparison between experimental and theoretical results will be discussed in Secs. IV C and IV D.

A. Measurement of the oscillation pressure threshold

The experimental setup described in Sec. III A is able to produce self-sustained oscillations. First, an internal pressure Pc is imposed. The initial aperture, h_0, in absence of upstream pressure, is measured by means of the optical setup. Next, the upstream pressure is continuously increased until self-sustained oscillations of the replica appear. Oscillations are quantitatively detected by a spectrum analysis as detailed in Ruty *et al.* (2005). The threshold pressure is denoted Ponset. The frequency of the oscillation at this threshold is determined by means of a spectrum analysis performed on the acoustic pressure signal. The subglottal pressure is then de-

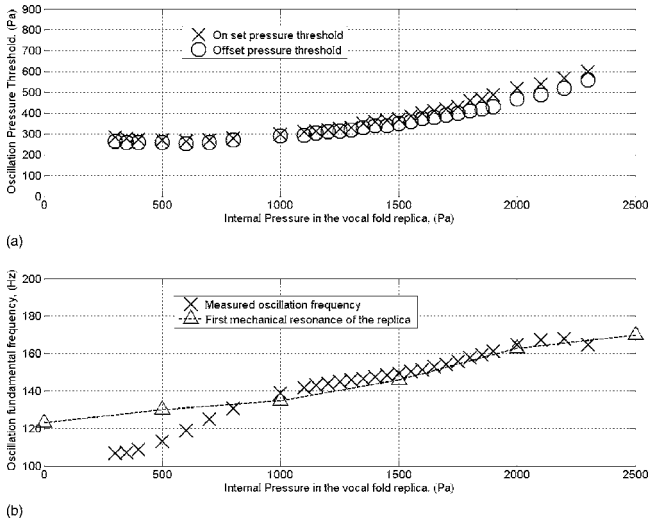


FIG. 9. Measured oscillation pressure thresholds and fundamental frequencies for a downstream resonator of 250 mm length. (a) Pressure thresholds, on-set pressure (X), off-set ($\bigcirc$). (b) Fundamental frequencies of the oscillations (X), first mechanical resonance of the replica (-△-).

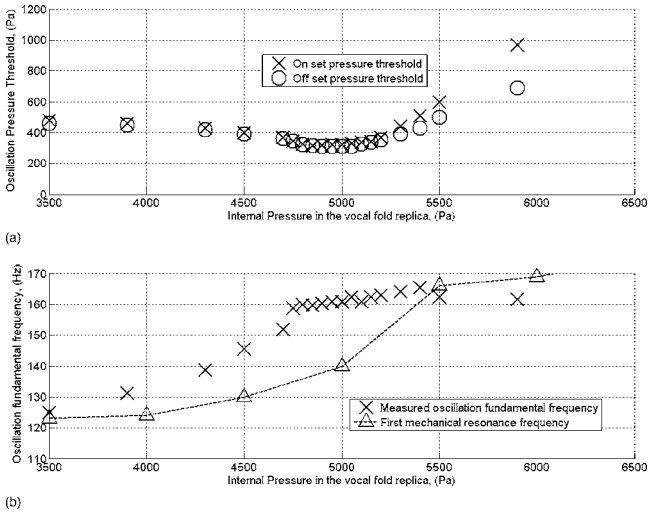


FIG. 10. Measured oscillation pressure thresholds and fundamental frequencies for a downstream resonator of 500 mm length. (a) Pressure thresholds, on-set pressure (X), off-set ($\bigcirc$). (b) Fundamental frequencies of the oscillations (X), first mechanical resonance of the replica (-△-).

creased until the oscillations cease and the associated pressure threshold, Poff-set, is recorded. This operation is repeated for several values of the internal pressure Pc, varying between 500 and 6500 Pa. The results obtained for both downstream resonators are depicted in Figs. 9 and 10.

An interesting observation is that, depending on the downstream resonator length, oscillations appear at comparable thresholds but for different internal pressure Pc. Thus, for the short resonator (250 mm), oscillations appear for low Pc values ($Pc \in [250;2500]$Pa), corresponding to a large initial aperture. For the long resonator (500 mm), oscillations appear for higher values of Pc ($Pc \in [3500;6500]$Pa). In this case, an interesting behavior is observed. An internal pressure Pc of 5000 Pa corresponds to a minimum Pon-set. At this point, the initial aperture is close to zero, i.e., the two latex tubes of the replica are almost in contact. This minimum pressure threshold, also observed by Titze *et al.* (1995), could be related to the optimal configuration to produce voiced sound, as described by Lucero (1998). Finally, in both cases, a hysteresis phenomenon, i.e., Pon-set > Poff-set, as the one described by Lucero (1999), is experimentally observed.

B. Stability analysis of the theoretical equations

In this section, we will examine if the theoretical models described in Sec. II are able to explain the experimental data. Therefore a stability analysis is applied to the proposed theoretical models.

All the variables in the theoretical equations are linearized and hence considered as the sum of an equilibrium value and a fluctuation around this equilibrium, which is denoted as $A=\bar{A}+a$ where $\bar{A}$ is the equilibrium value and a is the fluctuation. Assuming a constant subglottal pressure and a zero downstream pressure P_{supra} at the equilibrium, the following relations are obtained,

$$H_1(t) = \bar{H}_1 + h(t)_1, \quad H_2(t) = \bar{H}_2 + h_2(t), \tag{13}$$

$$P_{\text{sub}} = \bar{P}_{\text{sub}}, \quad P_{\text{supra}} = p \tag{14}$$

$$U_g = \bar{U}_g + u, \quad with\ u = \frac{\partial U_g}{\partial H_1}h_1 + \frac{\partial U_g}{\partial H_2}h_2 + \frac{\partial U_g}{\partial P_{\text{supra}}}p \tag{15}$$

$$F_i(P_{\text{sub}},P_{\text{supra}},H_1,H_2) = F_i(\bar{P}_{\text{sub}},0,\bar{H}_1,\bar{H}_2) + f_i,$$

$$with\ f_i = \frac{\partial F_i}{\partial P_{\text{supra}}}p + \frac{\partial F_i}{\partial H_1}h_1 + \frac{\partial F_i}{\partial H_2}h_2\ i = 1,2. \tag{16}$$

Considering only the fluctuating part, Eqs. (7) and (8) are rewritten as follows:

$$\frac{\partial^2 h_1(t)}{\partial t^2} = -\frac{2k}{m}h_1(t) - \frac{2k_c}{m}(h_1(t) - h_2(t)) - \frac{2r}{m}\frac{\partial h_1(t)}{\partial t} + \frac{4}{m}\left(\frac{\partial F_1}{\partial P_{\text{supra}}}p + \frac{\partial F_1}{\partial H_1}h_1 + \frac{\partial F_1}{\partial H_2}h_2\right),$$

$$\frac{\partial^2 h_2(t)}{\partial t^2} = -\frac{2k}{m}h_2(t) - \frac{2k_c}{m}(h_2(t) - h_1(t)) - \frac{2r}{m}\frac{\partial h_2(t)}{\partial t} + \frac{4}{m}\left(\frac{\partial F_2}{\partial P_{\text{supra}}}p + \frac{\partial F_2}{\partial H_1}h_1 + \frac{\partial F_2}{\partial H_2}h_2\right), \tag{17}$$

$$\frac{\partial^2 \psi(t)}{\partial t^2} + \frac{\omega_A}{Q_A}\frac{\partial \psi(t)}{\partial t} + \omega_A^2\psi(t) = \frac{Z_A\omega_A}{Q_A}\left(\frac{\partial U_g}{\partial H_1}h_1 + \frac{\partial U_g}{\partial H_2}h_2 + \frac{\partial U_g}{\partial P_{\text{supra}}}p\right). \tag{18}$$

Using a state-space representation with

$$x = \left[h_1\ h_2\ \psi\ \frac{\partial h_1}{\partial t}\ \frac{\partial h_2}{\partial t}\ \frac{\partial \psi}{\partial t}\right],$$

we obtained $\partial x/\partial t = Mx$ with M the state-space matrix,

$$M = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ -\frac{2}{m}\left(k + k_c - 2\frac{\partial F_1}{\partial H_1}\right) & \frac{2}{m}\left(k_c + 2\frac{\partial F_1}{\partial H_2}\right) & 0 & -\frac{2r}{m} & 0 & \frac{4}{m}\frac{\partial F_1}{\partial P_{\text{supra}}} \\ \frac{2}{m}\left(k_c + 2\frac{\partial F_2}{\partial H_1}\right) & -\frac{2}{m}\left(k + k_c - 2\frac{\partial F_2}{\partial H_2}\right) & 0 & 0 & -\frac{2r}{m} & \frac{4}{m}\frac{\partial F_2}{\partial P_{\text{supra}}} \\ \frac{Z_A\omega_A}{Q_A}\frac{\partial U_g}{\partial H_1} & \frac{Z_A\omega_A}{Q_A}\frac{\partial U_g}{\partial H_2} & -\omega_A^2 & 0 & 0 & -\frac{\omega_A}{Q_A} + \frac{Z_A\omega_A}{Q_A}\frac{\partial U_g}{\partial P_{\text{supra}}} \end{bmatrix}. \quad (19)$$

Assuming the simplifications made in Eq. (8), the state space vector x reduces to

$$x = \left[h_1 \ \psi \ \frac{\partial h_1}{\partial t} \ \frac{\partial \psi}{\partial t} \right].$$

Assuming $t_0 \ll 2\pi/\omega_0$, and using the Taylor-Young formula, one obtains $H_2(t) = H_1(t - t_0) = H_1(t) - t_0 \partial H_1(t)/\partial t$ and the state space matrix M becomes

$$M = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -\frac{2}{m}\left(k - 2\frac{\partial F_1}{\partial H_1} - 2\frac{\partial F_1}{\partial H_2}\right) & 0 & -\frac{2r}{m} - t_0\frac{4}{m}\frac{\partial F_1}{\partial H_2} & \frac{4}{m}\frac{\partial F_1}{\partial P_{\text{supra}}} \\ \frac{Z_A\omega_A}{Q_A}\left(\frac{\partial U_g}{\partial H_1} + \frac{\partial U_g}{\partial H_2}\right) & -\omega_A^2 & -t_0\frac{Z_A\omega_A}{Q_A}\frac{\partial U_g}{\partial H_2} & -\frac{\omega_A}{Q_A} + \frac{Z_A\omega_A}{Q_A}\frac{\partial U_g}{\partial P_{\text{supra}}} \end{bmatrix}. \quad (20)$$

Studying the eigenvalues of M, one is able to determine the presence or absence of oscillations. More precisely, an unstable equilibrium results in an eigenvalue with a positive real part, i.e., the model parameters (spring stiffness, damping, subglottal pressure, initial geometry) correspond with oscillation. The oscillation frequency is calculated as

$$f = \text{Im}(\lambda)/(2\pi) \quad (21)$$

with λ the considered eigenvalue.

During the experimental protocol, the imposed internal pressure Pc in the vocal fold replica affects both the initial geometry as well as the mechanical properties. Next, for a given Pc, the presence of vocal fold oscillations depends on upstream pressure variations. In a similar way, for a given set of model parameters (spring stiffness, damping, subglottal pressure, initial geometry), the stability analysis is assessed for subglottal pressure values P_{sub} varying between 0 and 1000 Pa. An exemplary result of the stability analysis is illustrated in Fig. 11. For this particular case, oscillations are predicted to appear for a subglottal pressure of 485 Pa. The fundamental frequency of the oscillations is 117 Hz.

This computation is repeated for the set of all model parameters identified experimentally by means of the method described in Sec. III B. For each of the two downstream resonator models, respectively of lengths 250 and 500 mm, and for each set of control parameters (associated with an internal pressure Pc), the onset pressure threshold, Pon-set, is calculated. The pressure threshold and fundamental frequencies obtained for the two models of vocal folds are compared with the experimental results, as depicted in Figs. 12 and 13.

C. Discussion

The results predicted with the theoretical models, for the two downstream resonators, show their capacity to reproduce qualitatively what happens experimentally, in terms of oscillation threshold pressure and fundamental frequency, but within a large error range.

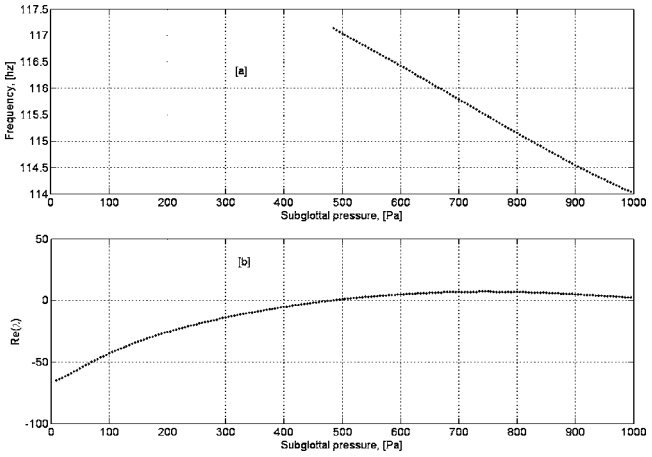


FIG. 11. Graphs resulting from the stability analysis. (a) Frequency of the oscillations, if they exist. (b) Eigenvalue real part.

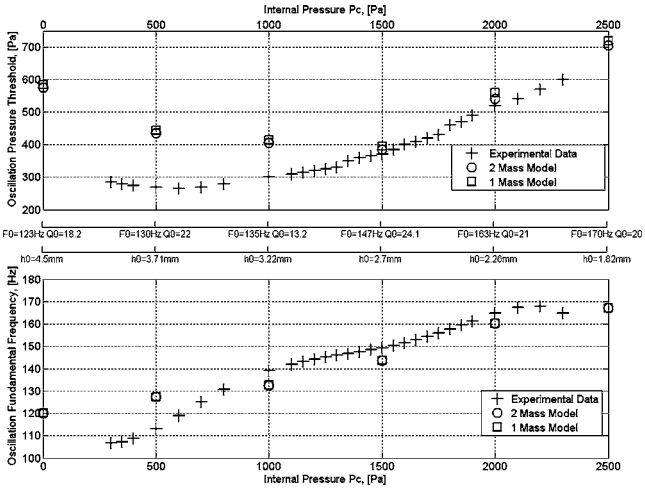


FIG. 12. Comparative study of the stability analysis result for the two theoretical models and the experimental results, for a 250 mm length downstream resonator. (a) Oscillation pressure threshold for experimental data (+), two-mass model (○), and one-mass model (□). (b) Fundamental frequency of the oscillations for experimental data (+), two-mass model (○), and one-mass model (□).

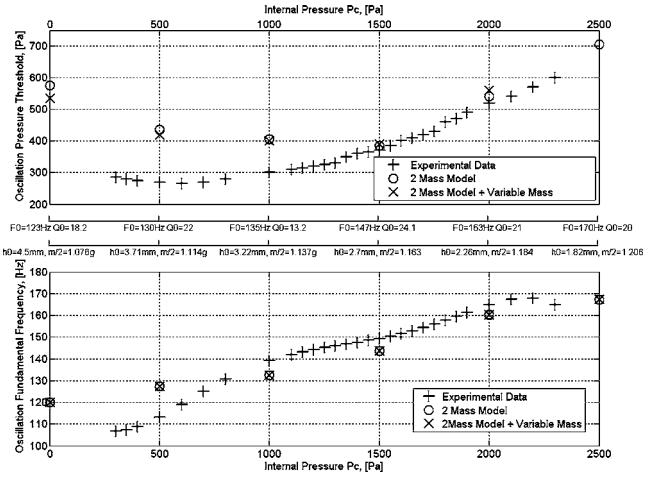


FIG. 14. Comparative study of the stability analysis result for the two-mass model with the two different mass estimations and the experimental results, for a 250 mm length downstream resonator. (a) Oscillation pressure threshold for experimental data (+), two-mass model (○), and two-mass model with a variable mass (X). (b) Fundamental frequency of the oscillations for experimental data (+), two-mass model (○), and two-mass model with a variable mass (X).

More precisely, concerning the fundamental frequency, a good accuracy of all theoretical models can be noted (agreement within 10%). As expected, the fundamental frequency increases when the internal pressure Pc is increased since increasing the internal pressure also increases the tension of the latex replica. This effect is well predicted in the theoretical models since an increase of the internal pressure in the experimental setup is associated with an increase of the spring stiffness in the theoretical models.

Concerning the oscillation threshold, the same order of magnitude and the same global behaviors are obtained. Globally, the same U-shaped behavior of the pressure threshold as a function of the internal pressure, Pc, is predicted by all theoretical models except when the vocal folds are in contact at rest position ($h_0=0$). In such a case, large departures from

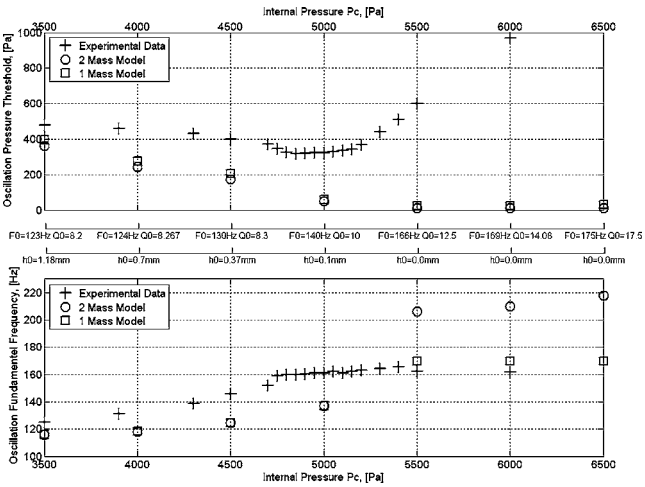


FIG. 13. Comparative study of the stability analysis result for the two theoretical models and the experimental results, for a 500 mm length downstream resonator. (a) Oscillation pressure threshold for experimental data (+), two-mass model (○), and one-mass model (□). (b) Fundamental frequency of the oscillations for experimental data (+), two-mass model (○), and one-mass model (□).

the theoretical predictions can be observed. When the vocal fold are not in contact at rest ($Pc<5500$ Pa), the delayed mass and two-mass models tend to provide comparable estimations for both the oscillation threshold pressure and fundamental frequency.

Quantitatively, although better for the longer acoustical resonator, the agreement between the predicted pressure thresholds and the measured data is rather poor except near the minimum threshold. Note that the value of the delay, t_0, does not change quantitatively the adequacy of the delayed model. These departures can, of course, be due to the theoretical models themselves but, also, to some extent, to a bad estimation of some input parameters. For example, the total mass, m, is estimated in a geometrical way as described in Sec. III B. This estimation assumes that the mass, m, is independent of the internal pressure, Pc, which is not true in the case of the latex replica. Indeed, for low internal pressures Pc, the latex tubes contain less water than for high values of Pc, where the latex tubes are inflated. To account for this effect, an alternative is to adjust the mass estimation by taking into account the geometrical and hence mass variations due to imposing different internal pressures. Considering an approximation of the latex tube deformation, depicted in Fig. 5(b), the estimation of the mass is modified as follows,

$$m = \rho_e L \frac{\pi(d_l + (h_{\text{ref}} - h_0)/4)^2}{8}, \tag{21'}$$

where h_{ref} is a reference aperture of 3 mm, and h_0 is the measured initial aperture of the replica for a given internal pressure Pc, which can be seen in Fig. 8.

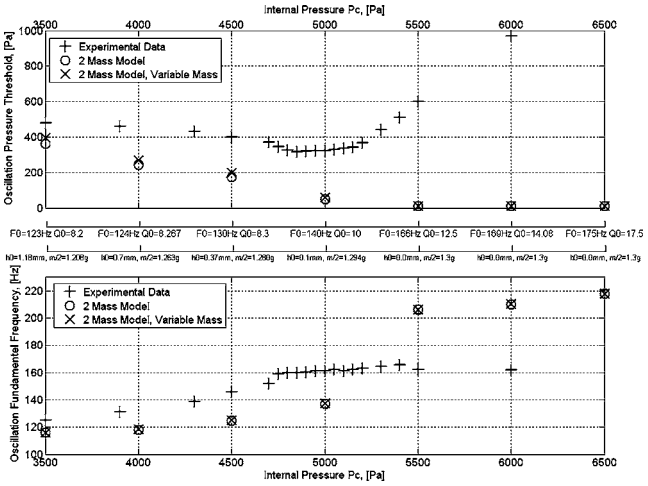


FIG. 15. Comparative study of the stability analysis result for the two-mass model with the two different mass estimations and the experimental results, for a 500 mm length downstream resonator. (a) Oscillation pressure threshold for experimental data (+), two-mass model (○), and two-mass model with a variable mass (X). (b) Fundamental frequency of the oscillations for experimental data (+), two-mass model (○), and two-mass model with a variable mass (X).

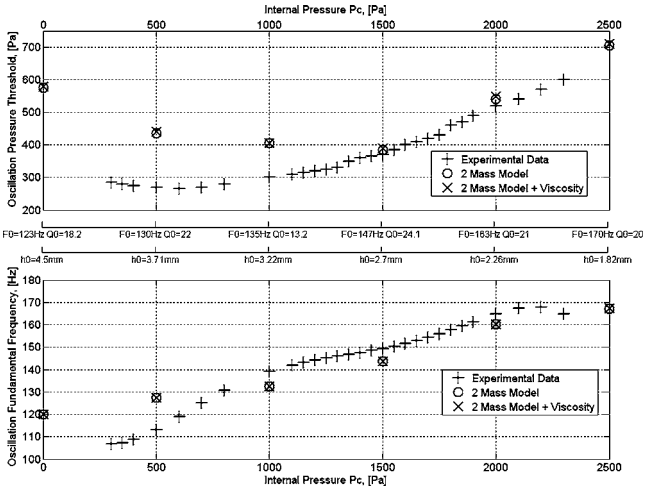


FIG. 16. Comparative study of the stability analysis result for the two mass model with additional viscosity and experimental data, for a 250 mm length downstream resonator. (a) Oscillation pressure threshold for experimental data (+), two-mass model (○), and two-mass model+viscous flow model (X). (b) Fundamental frequency of the oscillations for experimental data (+), two-mass model (○), and two-mass model+viscous flow model (X).

After this correction, the predicted oscillation thresholds are altered within an order of magnitude of 10% as shown on Figs. 14 and 15. If $h_0 > h_{\text{ref}}$, the oscillation threshold is decreased, while if $h_0 < h_{\text{ref}}$, the oscillation threshold is increased. The choice of this parameter does not seem to explain the departures observed between the measured data and the theoretical predictions.

Next the influence of the flow model is considered by taking into account viscosity in the flow model. Adding a Poiseuille term, Eqs. (1) and (2) become

$$P_{\text{sub}} - P(x,t) = \frac{1}{2}\rho U_g^2\left(\frac{1}{A^2(x,t)}\right) - 12\mu L_g^2 U_g \int_{x_0}^{x} \frac{dx}{A^3(x,t)}, \quad \text{if } x < x,$$

$$P(x,t) = P_{\text{supra}}, \quad \text{if } x > x, \tag{22}$$

where μ indicates the viscosity coefficient of the air and

$$U_g = \frac{12\mu L_g^2 \int_{x_0}^{x_s} [dx/A^3(x,t)] + \sqrt{(12\mu L_g^2 \int_{x_0}^{x_s} [dx/A^3(x,t)])^2 + 2(P_{\text{sub}} - P_{\text{supra}})\rho(1/A_s^2)}}{\rho(1/A_s^2)}. \tag{23}$$

The resulting simulations including the suggested modifications are depicted in Figs. 16 and 17. For large initial aperture no striking differences between the two flow descriptions can be found. As expected, the differences are significant in the case of small apertures corresponding to partial or complete closure of the glottis. At these points, the accounting for viscosity increases the accuracy in terms of prediction the pressure thresholds, in particular when the vocal folds are almost in contact, as well as in terms of fundamental frequency prediction. Indeed, at these points, the viscous term becomes dominant because of $1/A^3(x,t)$. However, the proposed corrections seem to not be completely sufficient to explain the quantitative differences observed between theoretical predictions and experimental data, which remain of the order of magnitude of 20%.

V. CONCLUSION

An experimental setup suitable to validate theoretical models of flow-structure-acoustic interaction with applications to voiced sound production is presented. The approach is illustrated on reduced low-order models for which the quantitative relationship between required model parameters, like mass and spring constants, and the physical characteristics of the replica can be fully exploited and controlled in order to validate the models.

Comparison between the low-order model predictions and experimental data can be summarized as follows:

(i) Despite their simplicity, the theoretical model outcomes are found to be qualitatively correct in terms of fundamental frequency and pressure threshold evolution as a function of the internal pressure of the rep-

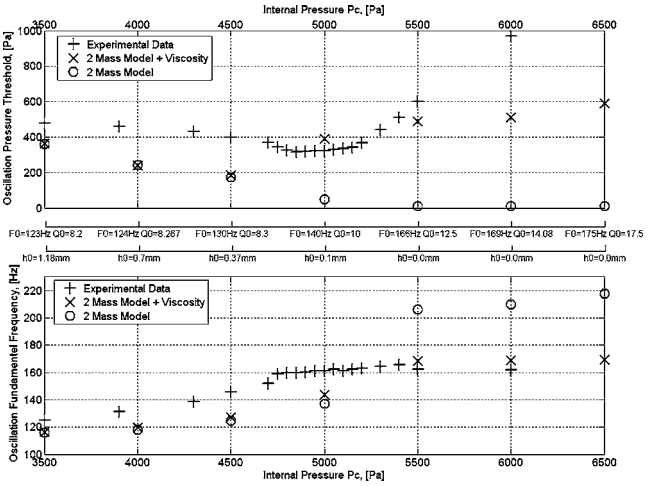


FIG. 17. Comparative study of the stability analysis result for the two-mass model with additional viscosity and the experimental data, for a 500 mm length downstream resonator. (a) Oscillation pressure threshold for experimental data (+), two-mass model (○), and two-mass model+viscous flow model (X). (b) Fundamental frequency of the oscillations for experimental data (+), and two-mass model (○), two-mass model+viscous flow model (X).

lica when the vocal folds are not in contact. In this later case, taking into account viscosity seems necessary.

(ii) The influence of the acoustical coupling is qualitatively well predicted by the theoretical models.

(iii) Viscosity of air appears to be crucial when very small glottal apertures (up to closed glottis) are involved and prediction of the oscillation pressure threshold is aimed.

(iv) While the fundamental frequency of the self-sustained oscillations is well predicted, the error ranges with respect to quantitative prediction of the oscillation pressure threshold is still large. Although an improved model outcome could be yielded by adjusting the model parameters, it is argued that the inaccuracies are mainly due to the nature of the validated reduced models themselves. Accounting for three-dimensional motions of the folds or for a better flow description during the initiation of the oscillation are examples of obvious weaknesses of the described one- and two-mass models.

(v) Lastly, the experimentally observed hysteresis has not been analyzed in the theoretical models, although several studies show that reduced order models are able to predict the observed hysteresis between on and offset pressure necessary to sustain the oscillations. This was the case for a two-mass model analyzed by Lucero and Koenig (2005), and also for a one-mass model with a delay factor presented in two papers by Lucero (2005, 1999).

ACKNOWLEDGMENTS

This research was partially supported by a Ph.D. Grant from the French Ministry of Research and Education and by the French/German project Popaart (CNRS-MAE). We would like to acknowledge Pierre Chardon for his help building the mechanical setup, and Freek van Uittert for his help designing the data acquisition chain.

Alipour, F., and Scherer, R. C. (**2001**). "Effects of oscillation of mechanical hemi-larynx model on mean transglottal pressures and flows," J. Acoust. Soc. Am. **110**, 1562–1569.

Avanzini, F., Alku, P., and Karjalainen, M. (**2001**). "One-delayed-mass Model for Efficient Synthesis of Glottal Flow," Proc. Eurospeech Conf., Aalborg, pp. 51–54.

Baken, R. J. (**1987**). *Clinical Measurement of Speech and Voice* (Allyn and Bacon),

Barney, A., Shadle, C. H., and Davies, P. O. A. L. (**1999**). "Fluid flow in a dynamic mechanical model of the vocal folds and tract. I. Measurements and theory," J. Acoust. Soc. Am. **105**, 444–455.

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., Herzel, H., Titze, I. R., and Krischer, K. (**1994**). "Interpretation of biomechanical simulations of normal and chaotic vocal fold oscillation with empirical eigenfunctions," J. Acoust. Soc. Am. **95**, 3595–3604.

Chan, R. W., Titze, I. R., and Titze, M. R. (**1997**). "Further studies of phonation threshold pressure in a physical model of vocal fold mucosa," J. Acoust. Soc. Am. **101**, 3722–3727.

Childers, D. G., and Wong, C.-F. (**1994**). "Measuring and modeling vocal source-tract interaction," IEEE Trans. Biomed. Eng. **41**, 663–671.

Cranen, B. (**1987**). "*The Acoustic Impedance of the Glottis. Measurement and Modelling*" (Sneldruk Enschede, The Netherlands).

Cullen, J. S., Gilbert, J., and Campbell, D. M. (**2000**). "Brass instruments: linear stability analysis and experiments with an artificial mouth," Acta Acust. **86**, 704–24.

Deverge, M., Pelorson, X., Vilain, C., Lagrée, P.-Y., Chentouf, F., Willems, J., and Hirschberg, A. (**2003**). "Influence of the collision on the flow through *in vitro* rigid models of the vocal folds," J. Acoust. Soc. Am. **114**, 3354–3362.

De Vries, M. P., Schutte, H. K., and Verkerke, G. J. (**1999**). "Determination of parameters for lumped parameter models of the vocal folds using a finite-element approach," J. Acoust. Soc. Am. **106**, 3620–3628.

Drioli, C. (**2005**). "A flow waveform-matched low-dimensional glottal model based on physical knowledge," J. Acoust. Soc. Am. **117**, 3184–3195.

Fitch, J. L., and Holbrook, A. (**1970**). "Model fundamental frequency of young adults," Arch. Otolaryngol. **92**, 379–382.

Flanagan, J. L. (**1972**). "Voices of men and machines," J. Acoust. Soc. Am. **51**, 1375–1387.

Flanagan, J. L., and Landgraf, L. (**1968**). "Self-oscillating source for vocal tract synthesizers," IEEE Trans. Audio Electroacoust. **AU-16**, 57–64.

Flanagan, J. L., Ishizaka, K., and Shipley, K. L. (**1975**). "Synthesis of speech from a dynamic model of the vocal cords and vocal tract," Bell Syst. Tech. J. **54**, 485–506.

Gauffin, J., and Liljencrants, J. (**1988**). "Modeling the Air Flow in the Glottis," Ann. Bull. RILP **22**, 41–52.

Gauffin, J., Binh, N., Ananthapadmanabha, T. V., and Fant, G. (**1983**). "Glottal geometry and volume velocity waveform," in *Vocal Fold Physiology: Contemporary Research and Clinical Issues*, edited by D. Bless and J. Abbs (College-Hill, San Diego CA), pp. 194–201.

Gilbert, J., Ponthus, S., and Petiot, J. F. (**1998**). "Artificial buzzing lips and brass instruments: experimental results," J. Acoust. Soc. Am. **104**, 1627–1632.

Herzel, H., Berry, D., Titze, I., and Saleh, M. (**1994**). "Analysis of vocal disorders with methods from non linear dynamics," J. Speech Hear. Res. **37**, 1008–1019.

Hirano, M., Kurita, S., and Nakashima, T. (**1983**). "Growth, development and aging of human vocal folds," in *Vocal Fold Physiology: Contemporary Research and Clinical Issues*, edited by D. M. Bless and J. M. Abbs, (College-Hill, San Diego, CA), pp. 22–43.

Hofmans, G. C. J., Groot, G., Ranucci, M., Graziani, G., and Hirschberg, A. (**2003**). "Unsteady flow through *in vitro* models of the glottis," J. Acoust. Soc. Am. **113**, 1658–1675.

Hollien, H., and Moore, P. (**1960**). "Measurements of the vocal folds during changes in pitch," J. Speech Hear. Res. **3**, 157–165.

Hunter, E. J., Titze, I. R., and Alipour, F. (**2004**). "A three-dimensional model of vocal fold abduction/adduction," J. Acoust. Soc. Am. **115**, 1747–1757.

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**, 1233–1267.

Kaneko, T., Komatsu, K., Suzuki, H., Kanesaka, T., Masuda, T., Numata, T., and Naito, J. (**1983**). "Mechanical properties of the human vocal fold-Resonance characteristics in living humans and in excised larynges," in *Vocal Fold Physiology: Biomechanics, Acoustics and Phonatory Control*, edited by I. R. Titze and R. C. Scherer (Denver Center for the Performing Arts, Denver, CO), pp. 304–317.

Kiritani, S., Imagawa, H., Imaizumi, S., and Saito, S. (**1987**). "Measurement of air flow pattern through a mechanically driven oscillating slit: a preliminary report," Ann. Bull. RILP **21**, 1–8.

Kob, M. (**2002**). "Physical modeling of the singing voice," Ph.D. thesis, Logos-Verl, Berlin.

Lous, N. J. C., Hofmans, G. C. J., Veldhuis, N. J., and Hirschberg, A. (**1998**). "A symetrical two-mass model vocal-fold model coupled to vocal tract and trachea, with application to prothesis design," Acustica **84**, 1135–1150.

Lucero, J. C. (**2005**). "Bifurcations and limit cycles in a model for a vocal fold oscillator," Commun. Math. Sci. **3**, 517–529.

Lucero, J. C. (**1999**). "A theoretical study of the hysteresis phenomenon at vocal fold oscillation onset-offset," J. Acoust. Soc. Am. **105**, 423–431.

Lucero, J. C. (**1998**). "Optimal glottal configuration for ease of phonation," J. Voice **12**, 151–158.

Lucero, J. C. and Koenig, L. L. (**2005**). "Phonation thresholds as a function of laryngeal size in a two-mass model of the vocal folds," J. Acoust. Soc. Am. **118**, 2798–2801.

Mantha, S., Mongeau, L., and Siegmund, T. (**2005**). "Dynamic digital image correlation of a dynamic physical model of the vocal folds," 4th International Workshop MAVEBA 2005, pp. 125–128.

Mc Glone, R. E. and Shipp, T. (**1971**). "Some physiological correlates of vocal fry phonation," J. Speech Hear. Res. **14**, 769–775.

Pelorson, X., Hirschberg, A., Wijnands, A. P. J., and Baillet, H. (**1995**). "Description of the flow through in-vitro models of the glottis during phonation," Acta Acust. **3**, 191–202.

Pelorson, X., Hirschberg, A., Van Hassel, R. R., Wijnands, A. P.J., and Auregan, Y. (**1994**). "Theoretical and experimental study of quasisteady-flow separation within the glottis during the phonation. Application to a modified two-mass model," J. Acoust. Soc. Am. **96**, 3416–3431.

Pierce, A. D. (**1991**). *Acoustics: An Introduction to its Physical Principles and Applications*, 1989 ed. (Acoustical Society of America, Melville, NY).

Rothenberg, M. (**1980**). "Acoustic interaction between the glottal source and the vocal tract," in *Vocal Fold Physiology*, edited by K. N. Stevens and M. Hirano (Univ. of Tokyo, Tokyo), pp. 305–328.

Ruty, N., Van Hirtum, A., Pelorson, X., Lopez, I., and Hirschberg, A. (**2005**). "A mechanical experimental setup to simulate vocal folds vibrations. Preliminary results," ZAS papers in Linguistics, pp. 162–175⟨http://www.zas.gwz-berlin.de/papers/zaspil/articles/zp40/ruty_final2.pdf⟩.

Saito, S., Fukuda, K., Suzuki, H., Komatsu, K., Kanesaka, T., and Kobayashi, N. (**1981**). "X-ray stroboscopy," in *Vocal Fold Physiology*, edited by K. N. Stevens and M. Hirano (Univ. of Tokyo, Tokyo), pp. 95–106.

Scherer, R. C. and Guo, R. C. (**1990**). "Effect of vocal fold radii in pressure distributions in the glottis," J. Acoust. Soc. Am. **88** (Suppl. 1), S150.

Scherer, R. C., Titze, I. R., and Curtis, J. F. (**1983**). "Pressure-flow relationships in two models of the larynx having rectangular glottal shapes," J. Acoust. Soc. Am. **73**, 668–676.

Scherer, R. C., Shinwari, D., De Witt, K. J., Zhang, C., Kucinschi, B. R., and Afjeh, A. A. (**2001**). "Intraglottal pressure profiles for a symmetric and oblique glottis with a divergence angle of 10 degrees," J. Acoust. Soc. Am. **109**, 1616–1630.

Story, B. H. and Titze, I. R. (**1995**). "Voice simulation with body cover model of the vocal folds," J. Acoust. Soc. Am. **97**, 1249–1260.

Svec, J. G., Horacek, J., Sram, F., and Vesely, J. (**2000**). "Resonance properties of the vocal folds: *In vivo* laryngoscopic investigation of the externally excited laryngeal vibrations," J. Acoust. Soc. Am. **108**, 1397–1407.

Thomson, S. L., Mongeau, L., and Frankel, S. H. (**2005**). "Aerodynamic transfer to the vocal folds," J. Acoust. Soc. Am. **118**, 1689–1700.

Titze, I. R. (**1988**). "The physics of small-amplitude oscillation of the vocal folds," J. Acoust. Soc. Am. **83**, 1536–1552.

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.

Vampola, T., Horáček, J., Veselý, J., Vokřál ∃ (**2005**). "Modelling of influence of velopharyngeal insufficiency on phonation of vowel /a/," 4th International Workshop MAVEBA 2005, pp. 43–46.

Van den Berg, Jw., Zantema, J. T., and Doornenbal, P. (**1957**). "On the air resistance and the Bernoulli effect of the human larynx," J. Acoust. Soc. Am. **29**, 625–631.

Vilain, C. E., Pelorson, X., Hirschberg, A., Le Marrec, L., Op't Root, W., and Willems, J. (**2003**). "Contribution to the physical modeling of the lips. Influence of the mechanical boundary conditions," Acta. Acust. Acust. **89**, 882–887.

Wong, D., Ito, M. R., Cox, N. B., and Titze, I. R. (**1991**). "Observation of perturbations in a lumped-element model of the vocal folds with application to some pathological cases," J. Acoust. Soc. Am. **89**, 383–394.

Measuring tongue motion from tagged cine-MRI using harmonic phase (HARP) processing[a)]

Vijay Parthasarathy and Jerry L. Prince
Department of Electrical and Computer Engineering, The Johns Hopkins University, Baltimore, Maryland 21218

Maureen Stone[b)] and Emi Z. Murano
Department of Biomedical Sciences and Orthodontics, University of Maryland Dental School, Baltimore, Maryland 21201

Moriel NessAiver
Department of Radiology, University of Maryland Medical School, Baltimore, Maryland 21201

(Received 4 March 2005; revised 14 August 2006; accepted 20 September 2006)

A cine series of tagged magnetic resonance (MR) images of the tongue is used to measure tongue motion and its internal deformation during speech. Tagged images are collected in three slice orientations (sagittal, coronal, and axial) during repetitions of the utterance "disouk" (*/disuk/*). A new technique called harmonic phase MRI (HARP-MRI) is used to process the tagged MR images to measure the internal deformation of the tongue. The measurements include displacement and velocity of tissue points, principal strains, and strain in the line-of-action of specific muscles. These measurements are not restricted to tag intersections, but can be calculated at every pixel in the image. The different motion measurements complement each other in understanding the tongue kinematics and in hypothesizing the internal muscle activity of the tongue. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2363926]

PACS number(s): 43.70.Jt, 43.72.Ar [AL] Pages: 491–504

I. INTRODUCTION

Understanding the dynamics of tongue motion is important because the tongue plays a crucial role in speech, swallowing, and breathing. Measuring tongue motion is a challenging problem for a variety of reasons: the tongue's location deep within the vocal tract, high degrees of freedom during motion, and its rapidity of motion during speech and swallowing. A number of measurement techniques have been used to measure tongue motion during continuous speech. Imaging techniques measure the entire tongue surface contour noninvasively (cinefluoroscopy, ultrasound, and cine-MRI), but lack temporal resolution [Stone (1999)]. Point-tracking techniques have better temporal resolution (electromagnetic midsagital articulometer EMMA, electromagnetic articulograph EMA, x-ray microbeam), but sparsely represent the tongue surface.

Electromyography (EMG) is the theoretical ideal for interpreting muscle activity. Bipolar fine-wire electrodes have been used as the method of choice for measuring lingual muscle activity during speech and breathing [Eastwood *et al.* (2003); Honda (1996); Honda and Kusakawa (1997); Miyawaki *et al.* (1997); Sabisky *et al.* (2006); Sauerland and Harper (1976)]. EMG signals, however, are difficult to obtain and interpret, and the relation between action and function of the muscles becomes less clear as task complexity increases. In breathing, genioglossus muscle (GG) is basically considered a dilator muscle. In a slightly more complex movement such as vowel production, GG is subdivided into at least two portions, anterior and posterior, based on its fanlike fiber distribution. In the most complex motions, such as consonants, EMG signals are almost uninterpretable. In other words, the linkage between muscle activity and tongue behavior becomes increasingly ill-defined. Therefore, a method that considers fiber direction, tongue contour, and physiological data during voluntary movements is invaluable and in urgent need.

Measurement techniques that use magnetic resonance imaging (MRI) are proving to be important tools in the analysis of tongue behavior because of MRI's ability to measure not only the tongue's surface but its interior as well. The use of MRI in speech research began with the recording of steady-state vowels. Initial studies imaged one static image plane [cf. Baer *et al.* (1987)]. The applications of static MRI of the oral cavity have included anatomical studies [Lufkin *et al.* (1987); McKenna *et al.* (1990)], examination of cancers and neurological diseases [Cha and Patten (1988); Takashima *et al.* (1989)], language studies [Hoole *et al.* (2001, 2000)], and static vocal tract modeling [Engwall (1999)]. Static MRI has been compared to other tongue measurement techniques such as electropalatography, ultrasound, electromyography, and electromagnetic articulography [Engwall (2003a); Wein *et al.* (1990)]. Most commonly, static MRI of the oral cavity is used to measure vocal-tract boundaries, catalog vocal-tract shapes, and extract airway volumes

[a)]A portion of this work was presented in the proceedings of 15th International Congress of Phonetic Sciences in a paper titled, "Tracking tongue motion from tagged magnetic resonance images using Harmonic Phase imaging (HARP-MRI)."

[b)]Electronic mail: mstone@umaryland.edu

for use in acoustic tube models [cf. Badin *et al.* (2002); Greenwood *et al.* (1992); Lakshminarayanan *et al.* (1991); Moore *et al.* (1992); Story *et al.* (1996, 1998); Sulter *et al.* (1992)].

Static MRI has limited capabilities in that it cannot be used to observe the changes in the tongue's position during speech. Hence, cine-MRI—acquisition of a time series of MR images—is used to observe the motion of the tongue. MR data are collected over multiple repetitions of a speech utterance, and an ensemble combination of the data produces a cine series of images. Multiple repetitions are necessary so that rapid tongue motions can be imaged with adequate spatial and temporal resolution [cf. Masaki *et al.* (1999)]. The images are collected in a specified orientation, usually sagittal, coronal, or axial [cf. Shadle *et al.* (1999)]. Multiple image orientations can be acquired with more repetitions [Stone *et al.* (2001b)]. A cine series of images can also be reconstructed using only a single repetition, but at the cost of reduced temporal and spatial resolution of the images. For example, using a single repetition per slice, Demolin *et al.* (2000) collected images at 4 Hz; Mady *et al.* (2002) studied glossectomized speech at 8 Hz; and Engwall (2003b) examined words at 9 Hz. Recently, Narayanan *et al.* (2004) have applied spiral MRI acquisition to do cine imaging collecting short sentences at 24 Hz with a single repetition.

A basic problem in speech research is to understand the relation between the dynamic forces of the muscles in the tongue's body and the kinematics of the tongue's surface. While cine-MRI has proven to be useful to study kinematics of the tongue surface, it does not give rich enough information about deformation of muscles in the tongue's body. The drawback of cine-MRI to provide information about the muscle deformation has been overcome through the use of MR tagging, which is a complementary method to EMG. MR tagging temporarily marks or "tags" a plane of tissue within the tongue. As the planes of tagged tissue move, the internal deformations can be observed. During early tagging experiments, reference tags were applied prior to a movement and deformed tags were captured either during or after the movement. Niitsu *et al.* (1994) separately captured vertical and horizontal tags of the head, first scanning the rest position followed by a vowel position held for 2 s. Niitsu's work first demonstrated the ability to visualize internal tongue deformation. Napadow *et al.* (1999a, b) extracted principal strains in the tongue's interior during nonspeech motions (protrusion and swallowing) by repeatedly measuring the rest and deformed positions. By progressively changing the time position of the deformed frame, several time frames were collected over multiple repetitions, and the data were combined into a pseudosequence. The tags could be hand-traced in the reference and deformed images, and subsequently used to hypothesize active muscle activity [Niitsu *et al.* (1992)].

Formation of a pseudosequence of tagged images and the subsequent hand-tracing of the tags is a manually intensive way of visualizing the progression of the internal deformation of the tongue. Naturally, the combination of cine-MRI with tagged MRI to create tagged-cine-MRI (tMRI) was the next substantial improvement in the measurement of the tongue's internal deformation. In tMRI, tags are applied prior to movement of the tissue, and cine-MRI is then collected throughout the entire motion. Tagged cine-MRI is a critical step forward in research on tongue behavior because it enhances the use of MRI in three ways. First, the tags can be tracked throughout the entire period of the speech utterance. Therefore, subtle deformations of the tongue can be visualized and quantified temporally. Second, while the existing tissue point tracking methods (x-ray microbeam, and electromagnetic articulography) track only the oral surface of the tongue, tMRI tracks the tongue surface completely from the tip to the root. Third, since the entire body of the tongue can be tracked, local tissue compressions can be calculated in the interior of the tongue. These local tissue compressions can be used to infer muscle activity, which can either be used in representational models of tongue control [Stone *et al.* (2000)] or as inputs to predictive models that map muscle contraction patterns to tongue surface deformations [Dang and Honda (1997); Iny *et al.* (2001)]. Local tissue compressions can also be used to test physiologically based models of tongue motion such as finite element models (FEM). Three-dimensional FEMs have been used to predict tongue behavior from muscle activity or vice versa [Dang and Honda (2004); Hashimoto and Suga (1986); Kakita *et al.* (1985); Kiritani *et al.* (1976); Levine *et al.* (2005); Wilhelms-Tricarico (1995)]. Local compressions from tMRI can also provide information about muscle activity that can serve as the bases for initial predictions for such models.

In order to quantify motion using tMRI images, tags need to be tracked and then combined to form motion estimates. A variety of image processing techniques have been proposed to estimate motion and deformation from tMRI images [Denney and McVeigh (1997); Guttman *et al.* (1994); Kerwin *et al.* (2000); Kerwin and Prince (2002); Ozturk and McVeigh (2000); Radeva *et al.* (1997); Young *et al.* (1995)]. These techniques include several semiautomatic tools to extract the tags from the tagged images and then combine the tag displacements into 2D or 3D motion measurements. All these methods were first developed to measure the motion of the heart from tagged MR images. In fact, harmonic phase MRI (HARP-MRI) was also first introduced as a fast automatic method to compute the deformation of the heart [Osman *et al.* (2000); Osman and Prince (2000)]. Even though there are considerable differences in the motion of the heart and the tongue, some of these methods could be adapted to measure tongue motion. Dick *et al.* (2000) adapted a 4D B-spline model to measure 3D strains and motion tracks in the tongue.

In the same vein, this work adapts HARP-MRI to measure tongue motion. In comparison to the existing methods of tagged image processing, HARP-MRI is faster and works with the least human intervention. Using the unique properties of HARP-MRI, we propose a fast MR imaging scheme that reduces the number of repetitions required by a factor of 2. We also develop algorithms to calculate motion measures that are unique to the tongue. These motion measures are subdivided into two categories: user-driven measures and data-driven measures. User-driven measurements include tracking of handpicked tissue points and calculating the

stretch between two hypothesized muscle endpoints. Data-driven measures, on the other hand, are calculated on all tissue points. These include velocities and principal strains. Data-driven measurements are objective measurements of internal tissue behavior and can be used to corroborate the user-driven measurements. They can also be used to develop and validate kinematic and dynamic models of the tongue.

II. THEORY

A. MR Imaging methods

MR tagging is traditionally done using the SPAMM (spatial modulation of magnetization) imaging protocol [Axel and Dougherty (1989a, b)]. A popular category of SPAMM tagging, which comes standard in many MR scanners, is called higher-order SPAMM (for example 1–4–6–4–1). This type of SPAMM tagging produces thin sharp black lines on the image. Even though these sharp tags are excellent for visualization, they have been found to be unsuitable for HARP-MRI processing [Kuijer *et al.* (2001)]. HARP-MRI performs better on smoothly varying tagging patterns produced by a technique called CSPAMM (complementary spatial modulation of magnetization) [Fischer *et al.* (1993)]. CSPAMM produces smoothly varying sinusoidal tags rather than sharp tags. Each CSPAMM tagged image is a combination of two tagged images, which are obtained separately using 1–1 SPAMM tagging. The 1–1 SPAMM tagging can be understood as the multiplication of the magnetization of the anatomy with a two-dimensional sinusoid. The two tagging sinusoids in a CSPAMM pair differ by a phase shift of π radians, which can be seen by comparing the two images in a vertically tagged CSPAMM pair in Figs. 1(a) and 1(b). Figures 1(d) and 1(e) show a horizontally tagged CSPAMM pair. Each of these images are acquired using a 1–1 SPAMM protocol. The combination of the images in each pair results in an improved tagged image that has twice the tag contrast than the individual 1–1 SPAMM images [Fischer *et al.* (1993)] and a tag persistence up to 1200 milliseconds.

The traditional way of combining CSPAMM images requires a complex subtraction of the two images in each pair. This subtraction is problematic since the phases of these images are corrupted by magnetic field inhomogeneity. In this work, we have used a recent method called MICSR (magnitude image CSPAMM reconstruction) to combine the pair of CSPAMM images [NessAiver and Prince (2003a)]. The MICSR method bypasses the use of complex data sets by using magnitude images only. If $A(x,y)$ and $B(x,y)$ represent two images in a CSPAMM pair, the MICSR reconstruction formula is given as

$$I_{\mathrm{MICSR}}(x,y) = |A(x,y)|^2 - |B(x,y)|^2. \quad (1)$$

The MICSR-reconstructed CSPAMM images are shown in Figs. 1(c) and 1(f). The MICSR method produces tagged images with pure sinusoidal tags without the constant bias that is characteristic of 1–1 SPAMM tagged images. To visualize this, notice that the magnitude images in Figs. 1(a), 1(b), 1(d), and 1(e) have a black background with crests of the rectified sinusoid seen as white. The sinusoids are rectified because only the magnitude is displayed, whereas in MICSR images the background is zero (gray), with both positive (white) and negative (black) excursions of the sinusoid. Since the output of MICSR reconstruction is pure sinusoidal tags, the use of MICSR is ideal for HARP-MRI processing. Also, improved tag contrast and tag persistence of these MICSR images improve our ability to acquire and analyze longer speech samples.

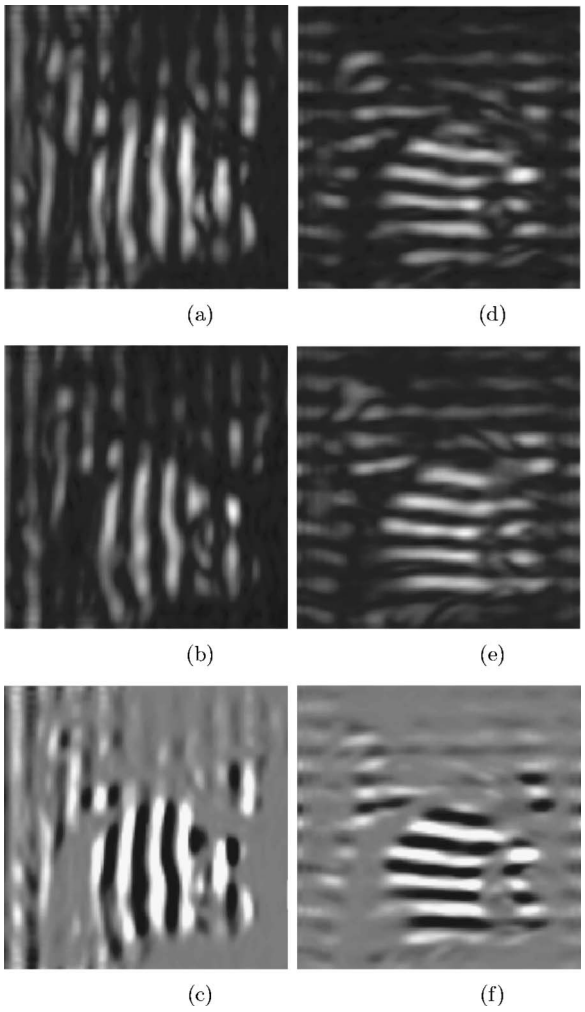


FIG. 1. Images (a) and (b) form a pair of vertically tagged CSPAMM images with phase shifted tags, and (c) is the MICSR combination of images (a) and (b). Images (d) and (e) form a pair of horizontally tagged CSPAMM images and (f) is the MICSR combination of images (d) and (e). Intensity of MICSR images (c) and (f) have been thresholded for better visualization of the tags.

B. Theory of HARP-MRI

Harmonic phase MRI (HARP-MRI) is used to estimate motion from sinusoidally tagged images [Osman *et al.* (1999)], an example of which is shown in Fig. 2(a). The tag lines, which were straight when initially applied, are now bent due to the local deformations in the tissue. The multiplication of the underlying magnetization of the anatomy with the tagging sinusoid produces two harmonic spectral peaks in the Fourier spectrum of the image [Fig. 2(b)]. The premise of HARP-MRI is that most of the spectral energy corresponding to the motion of the tissue is localized around

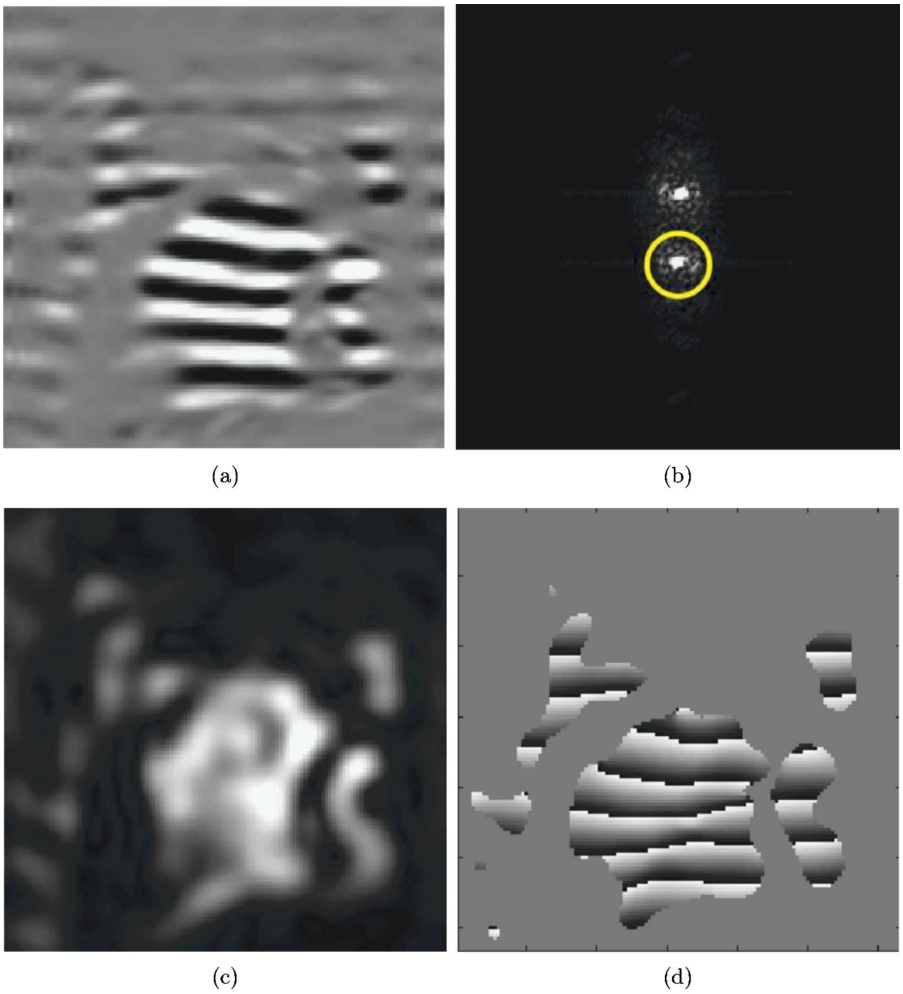


FIG. 2. (a) Horizontally tagged image produced by MICSR combination; (b) magnitude of its Fourier spectrum with the HARP bandpass filter; (c) harmonic magnitude image; and (d) harmonic phase image.

one of the harmonic spectral peaks. A bandpass filter [see the circle in Fig. 2(b)] is used to filter one of the harmonic peaks and the rest of the spectrum is zero padded. A *harmonic image* is constructed by calculating the inverse Fourier transform of the zero-padded spectrum. The resulting complex harmonic image can be expressed as

$$I(\mathbf{x},t) = D(\mathbf{x},t)e^{j\phi(\mathbf{x},t)}, \tag{2}$$

where $\mathbf{x}=[x_1\ x_2]^T$ is the position of the tissue point within an image and t is the time elapsed after the application of the tags [Osman *et al.* (2000)].

The *harmonic magnitude image* $D(\mathbf{x},t)$ [Fig. 2(c)] is proportional to the transverse magnetization at position $\mathbf{x}$ and time t. It represents an image of the anatomy, which is blurred by the bandpass filter. The basic concept of HARP-MRI is that the motion of the tissue is embedded in the *harmonic phase* (HARP) of the tagging sinusoid, which is measured as the phase $\phi(\mathbf{x},t)$ of the complex harmonic image. Figure 2(d) shows the image of the harmonic phase corresponding to the tagged image in Fig. 2(a). Since the phase of a complex image is constrained to lie between $-\pi$ and π, phase wraps occur in the HARP image. Note that these phase wraps exactly coincide with the troughs of the tag pattern, which indicates that the HARP phase has information about the motion of the tags. Mathematically, the harmonic phase $\phi(\mathbf{x},t)$ is linearly related to 2D tissue displacement $\mathbf{u}(\mathbf{x},t)$ as follows:

$$\phi(\mathbf{x},t) = -\boldsymbol{\omega}_0^{\mathrm{T}}\mathbf{u}(\mathbf{x},t) + \boldsymbol{\omega}_0^{\mathrm{T}}\mathbf{x}, \tag{3}$$

where $\boldsymbol{\omega}_0$ is the tagging vector, which corresponds to the orientation and frequency of the tag pattern in the image plane. In order to completely characterize displacement in two dimensions, we combine the harmonic phases from both the horizontal and vertical tags.

III. MATERIALS AND METHODS

A. Speaker and speech material

We demonstrate the application of HARP-MRI on one normal speaker. The speaker was a 27-year-old male, nonnative speaker of Tamil accented English. The speaker did not have any dental fillings that might interfere with the MRI magnetic field. The speech material was the utterance */disuk/*, which was selected because it involves large movements of the tongue. The tongue makes an alveolar /d/, releases downward into /*is*/, moves backward for the /u/, upward into the velar stop /k/, and then returns forward for the next repetition.

B. Tagging and imaging parameters

Tagged MR images were collected on a Marconi 1.5 Tesla Eclipse scanner. The speaker laid supine in the MR scanner with the TMJ (temporomandibular joint) phased array coil positioned to image the area from the lower nasal

cavity to the upper trachea. Images in three different orientations were collected: sagittal, coronal, and axial. Eight sagittal slices, nine coronal slices, and nine axial slices were collected with a 5-mm slice thickness. On each slice, two sets of tagged images were obtained, one horizontal and the other vertical. Tagging is done using a 1–1 SPAMM tagging protocol with a tag spacing of 5 mm, which corresponds to a 10-mm period of the sinusoidal tag modulation. Imaging was done using a gradient echo imaging sequence with the following parameters: echo time of 2.2 milliseconds (ms), pulse repetition time of 4.13 ms, and a rectangular field of view of 200×200 mm. The data acquisition matrix was 64×22, zero filled to 128×128 sample points in Fourier space. Thus, the spatial resolution of the interpolated spatial voxel is $1.56 \times 1.56 \times 5$ mm. More details about the data acquisition matrix can be found in Appendix A.

C. Spatial resolution

The data acquisition matrix of size 64×22 in Fourier space corresponds to a tagged MR image with low spatial resolution (not tag resolution). This is unlike the other studies of the tongue that acquire MRI images with the primary objective of obtaining good spatial resolution. Therefore, it is natural to ask how does HARP-MRI work on such low resolution data. The key is to understand that HARP-MRI uses a bandpass filter to extract the complex harmonic image and needs only a small part of the Fourier spectrum to reconstruct motion. Therefore, it is sufficient to collect a small number of samples in Fourier space. Sampath *et al.* (2003) have validated this claim by showing that HARP-MRI estimates motion accurately even in cases of low resolution data. This unique property of HARP-MRI is critical in reducing the number of repetitions of the utterance required to reconstruct one tagged MR image.

D. Temporal resolution

The frame rate of the cine series is 18 Hz. Each cine series was collected over multiple repetitions, each repetition being 1 s long. The one second repetition time contained two parts: 667-ms recording time and 333-ms setup time for the next recording. With a frame rate of 18 Hz and 667-ms recording time, we collected 12 images for each cine series. In the rest of this paper we refer to each image in the cine tMRI series as one time frame.

E. Number of repetitions and repeatability

In order to use HARP-MRI on one 2D slice, we need four cine sequences of tagged images—two CSPAMM image sequences in two tag directions. In this research, we acquire each tagged cine series in four repetitions of the utterance. Therefore, we need 16 repetitions to track tongue motion on one slice. When compared to previous protocols that require 32 repetitions to acquire similar kind of data [Stone *et al.* (2001a)], this protocol represents a twofold improvement in the speed of data acquisition.

Since each tagged MR image is a combination of four repetitions, the speaker's ability to repeat the utterance precisely is critical to image quality. Variability across repetitions causes blurring of images and crossing of tags after combination. In order to ensure least variability across repetitions, each potential speaker is pretested and trained before the MRI experiment. Subjects are selected based on this pretest, which consists of repeating the proposed speech material (and others) to a metronome. The metronome is a combination of a repeating rhythmic tone [Shimada *et al.* (2002)] and an MR gradient sound that was recorded during a previous imaging session. This pretest is virtually identical acoustically to the experience in the scanner during tagged image acquisition. The speech material is repeated at 1-s intervals and the subject is asked to time his/her repetitions to this gradient sound cycle. The acoustic data are subsequently analyzed to measure the variability in repetition time. A variability threshold of 50 ms is set, above which the speaker is not used. The speaker used in this paper was very precise, with a standard deviation of 20 ms. Following the pretest, qualified subjects are trained by further synchronizing their utterance to the repeating rhythmic tone. The number of repetitions for the subject to get into rhythm is determined and later incorporated into the MR imaging protocol. The subject presented in this paper needed three repetitions to get into rhythm, and felt comfortable when the number of beats in the rhythmic tone was equal to the number of syllables in the utterance, which is two in this case.

F. Acquisition of cine-MRI and high-resolution static MRI

In addition to tagged MR images, untagged cine-MRI images were collected for the same speech task. The cine-MRI protocol requires eight repetitions per slice to reconstruct the cine image series. The data acquisition matrix is 66×64 (almost three times the spatial resolution of the tagged images) over a rectangular field of view of 200×200 mm, zero filled to 256×256 sample points in Fourier space. Thus, the spatial resolution of the interpolated spatial voxel is $0.75 \times 0.75 \times 5$ mm. Figures 3(a)–3(c) show images of one particular time frame of the coronal, sagittal, and axial slices, respectively. The tongue surface is seen better on the untagged images than on the tagged images because of its better spatial resolution. The untagged images are overlaid on the tagged images to visually correlate the changes in the tongue surface with the tongue's internal deformations. The untagged images are also used to create an articulation table (Table I) for */disuk/* by correlating the motion of the vocal-tract structures with the respective sounds. This table is later used to interpret the motion measurements from HARP-MRI.

Finally, in order to delineate the directions of specific muscles for this particular subject, a high-resolution static MRI is also acquired. An example of the high-resolution static MRI of a midsagittal slice is shown in Fig. 3(d). Anatomical landmarks delineated in the static MRI are used to outline the muscle directions in the tagged and untagged images. For example, in this study, the high-resolution sagittal image is used to find the origin and the insertions of the genioglossus muscle in order to find the strain along its line-of-action.

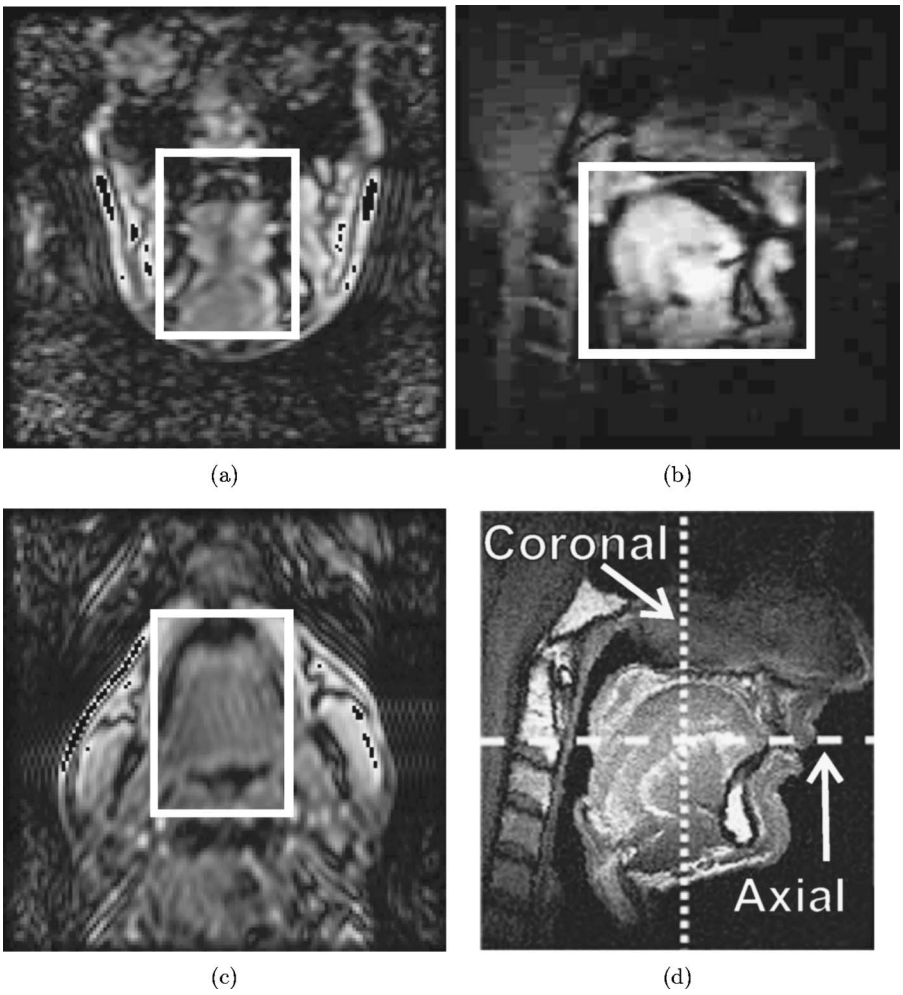


FIG. 3. Untagged image of one time frame of (a) coronal slice; (b) sagittal slice; (c) axial slice. White box shows region of interest around the tongue. (d) High-resolution static image of the anatomy to identify the muscle directions. White box shows a region of interest (ROI) around the tongue. The motion measures will be shown only for this ROI in order to avoid clutter of measurements in the surrounding tissues.

In this study, the motion is visualized on all three slice orientations. Of the eight sagittal slices, the results will be shown only on the midsagittal slice. The coronal and axial slices that are selected for visualization are marked in Fig. 3(d). While the selected coronal slice runs through the middle of the first upper molar, the axial slice runs parallel to the crown of the upper molars.

IV. RESULTS: VISUALIZATION AND DATA ANALYSIS

A. Checkerboard visualization

It is often helpful to visualize the local deformation of the tissue in addition to quantifying it. In order to better visualize the local tongue deformations, we introduce the checkerboard visualization [Fig. 4(a)], which is obtained by multiplying the horizontally and vertically tagged MICSR images and thresholding the resulting image [NessAiver and Prince (2003a), (2003b)]. Local elongation, compression, and shear of the tissue can be visualized as different shapes of the deformed squares. The images are upsampled using sine interpolation to increase tag crispness. The checkerboard display is purely a visualization tool and does not have any bearing either on the data analysis or on the resolution of HARP-MRI.

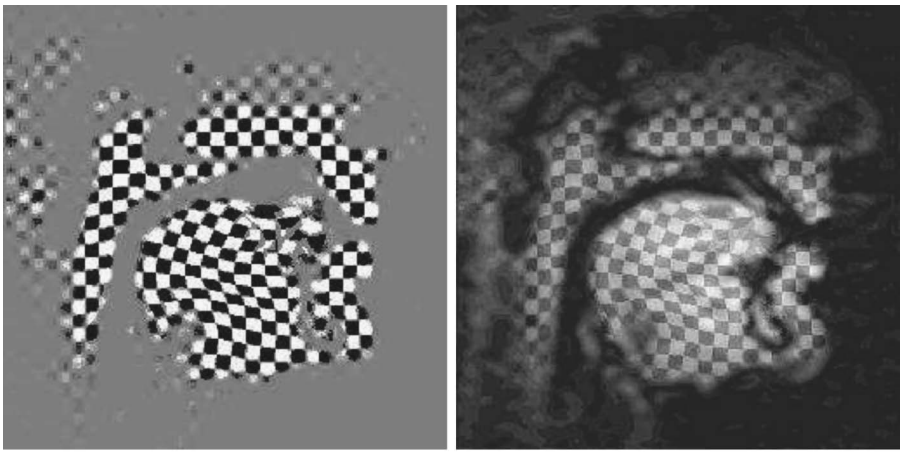
TABLE I. Articulation tabel for */disuk/* based on the shape of tongue and lips as seen in the untagged cine-MRI images.

Frames	Articulation
1–5	/*d*/
5–8	/*d*/−/*s*/
8–9	/*s*/−/u/
9–10	/u/−/k/
10–11	/k/
11–12	Preparation for next utterance

The checkerboard display in Fig. 4(a) shows the position of the tongue in the sixth time frame of the utterance */disuk/*. A considerable shearing and rotation can be seen in the body of the tongue. On the other hand, there is clear lack of deformation in the lips and the velum. The surface of the tongue is smeared due to reduced spatial resolution in the tagged images. In order to improve the visualization of the tongue surface, we overlay the checkerboard image onto the untagged image as shown in Fig. 4(b). This overlay visualization helps to better correlate the tongue's internal deformations with movements of the tongue surface. Note the marked improvement in the visualization of the tongue tip and tongue blade in Fig. 4(b) when compared to Fig. 4(a).

B. HARP-MRI analysis, visualization, and data interpretation

Equation (3) shows the linear relationship between the harmonic phase ϕ and tissue point displacement $\mathbf{u}$. From this



(a) (b)

FIG. 4. (a) MICSR trinary display. (b) Overlay of MICSR checkerboard and untagged image; note the better edge definition near the tongue tip in the overlay image.

estimate of displacement, the following motion measures can be calculated: (1) trajectories of tissue points; (2) velocities of tissue points; (3) principal strains; (4) strain between two points that are hypothesized to be the endpoints of a particular muscle. Even though two of these measures [(1) and (2)] have been reported in the context of the heart, in this research, we adapt the measures to the more complex movements of the tongue. In addition, our method to calculate principal strains and strains in the line-of-action of a muscle are unique to the analysis of the tongue.

The above quantities measured using HARP-MRI can be split into two categories: user-driven and data-driven. User-driven quantities require some manual intervention. The trajectories of tissue points and the line-of-action strains are user-driven quantities because they require the user to pick specific points in the tongue. On the other hand, the principal strains and the velocities are completely data driven and do not require any manual intervention. In the next few sections, we explain the calculation and interpretation of both kinds of measurements.

1. User-driven measurements

a. Trajectories of tissue points. The trajectory of a tissue point tracks its position at different times during a utterance. A tagged tissue point has two HARP values, one from the horizontal tags and another from the vertical tags. Let the horizontal HARP value be ϕ_H and the vertical HARP value be ϕ_V. Let the overall HARP vector be $\boldsymbol{\phi}=[\phi_H\ \phi_V]^T$. The HARP vector is a material property of the tissue. Therefore, as a tissue point moves, its HARP vector moves along with it. Hence, by tracking the HARP value through a cine series of images, the trajectory of a tissue point can be estimated [Osman *et al.* (1999)]. Consider a material point located at $\mathbf{x}_N$ at time t_N. If $\mathbf{x}_{N+1}$ is the position of this point at time t_{N+1}, then since HARP is a material property, we must have

$$\boldsymbol{\phi}(\mathbf{x}_{N+1},t_{N+1}) = \boldsymbol{\phi}(\mathbf{x}_N,t_N). \tag{4}$$

This relationship provides the basis for tracking $\mathbf{x}_N$ from time t_N to time t_{N+1}. Our goal is to find $\mathbf{x}$ that satisfies

$$\boldsymbol{\phi}(\mathbf{x},t_{N+1}) - \boldsymbol{\phi}(\mathbf{x}_N,t_N) = 0, \tag{5}$$

and then set $\mathbf{x}_{N+1}=\mathbf{x}$. This *HARP tracking* algorithm is repeated for consecutive time frames to track a particular point through the entire cine series and the trajectory is visualized as a path line.

b. Visualization. Local motion of the tongue is visualized by displaying trajectories of multiple handpicked tissue points (see Fig. 5). The small circle represents the positions of the tissue points in the first time frame, and the plus symbols represent the positions of the points in the last time frame. The trajectories are split into six (progressively darker) shades of gray. Each shade of gray corresponds to two of the 12 time frames. It is worth recalling that the locations of these tissue points need not be tag intersections. HARP tracking is capable of tracking arbitrary points on the image, even between pixels.

c. Interpretation. The timing and directions of the surface point trajectories corroborate the inhomogeneity of the

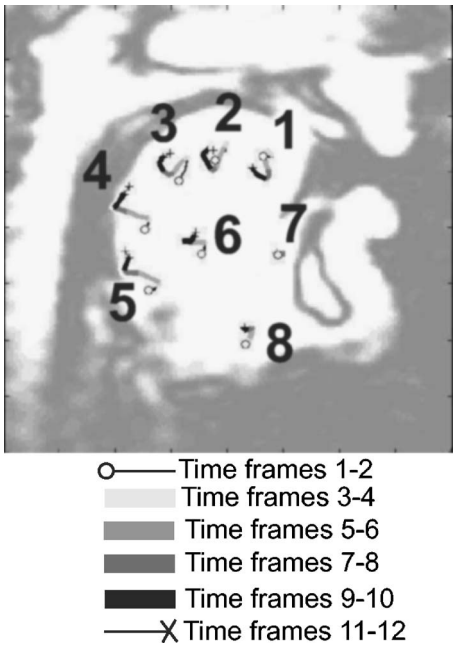
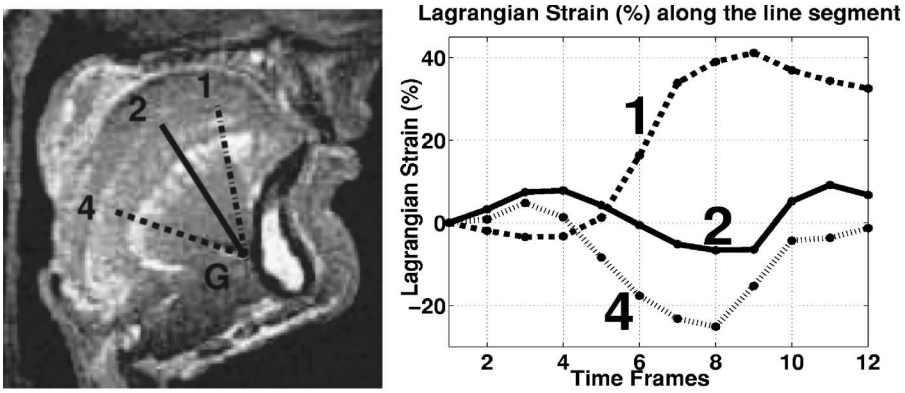


FIG. 5. Motion tracks for /*disuk*/ in a midsagittal slice. Black circles represent the position of the tissue point at time frame 1. The progression of shades from gray to black indicates the trajectory of the tissue point. In the electronic version, please zoom in more than 500% to see the trajectories clearly.



(a) (b)

FIG. 6. (a) Genioglossus muscle directions delineated on high-resolution static MRI. (b) Lagrangian strain in the three lines of action of genioglossus.

tongue surface movements (see Fig. 5). The most interesting differences occur in the middle of the motion. Point 1 first moves down and forward, and then begins to move back. Points 2 and 3 move down and back, and then begin to move back and up. Points 4 and 5 move back continuously, and then move slightly upward. These differences reflect at least three distinct regions of motion on the tongue surface. The internal points indicate further inhomogeneity. Point 6 moves straight back. Point 7, which is close to the origin of genioglossus, does not move at all, implying that the anterior lowering of point 1, 2, and 3 is due to verticalis, and not genioglossus. Point 8 shows an upward motion primarily, then a slight backward motion in the jaw muscles, despite the lack of jaw motion seen in point 7. This implies that jaw muscle contraction elevates the hyoid, and leads to thickening of the floor muscles, elevating this point and in turn the tongue.

The trajectories in Fig. 5 also reflect the coordination needed to execute the specific phonemes in the word. Note the differences in direction, extent, and timing of the trajectory motions. For points 4 and 5, the change from backward to upward motion occurs abruptly at time frame 9, when the tongue moves from /*u*/ to /*k*/. For point 1, the upward turn is slower (time frames 8–10) and is preceded by a lowering motion from the /*s*/ into the /*i*/. Points 2 and 3 turn upwards during time frames 8–10, ending with a forward slide during time frames 11 and 12 to get ready for the upcoming /*d*/ of the next repetition. Point 6 shows minimal upward/downward motion of the internal tongue.

d. Strain along the line-of-action of a muscle. The use of HARP tracking can be extended to calculate the change in length of a line segment by simply tracking the endpoints of the line segment. By choosing two endpoints on an image where we hypothesize a particular muscle to be, we can calculate the strain along the line-of-action of that muscle. Consider two points with reference positions $\mathbf{x}^1_{\text{ref}}$ and $\mathbf{x}^2_{\text{ref}}$ at time t_{ref}. At time t_N, we can track the positions of these points using HARP tracking as $\mathbf{x}^1_N$ and $\mathbf{x}^2_N$. The strain between these two points at time t_N with respect to the reference time can be calculated as

$$e = \frac{\|\mathbf{x}^2_N - \mathbf{x}^1_N\|}{\|\mathbf{x}^2_{\text{ref}} - \mathbf{x}^1_{\text{ref}}\|} - 1. \tag{6}$$

The strain is zero if the distance between the points remains unchanged, negative if there is shortening, and positive if there is lengthening. The reference time, t_{ref}, can be set at any time frame. In this work, all strain along the line-of-action are evaluated using the first time frame as the reference. Since this measure involves tracking of specified material points, the resulting strain is a Lagrangian measure.

e. Visualization. Figure 6(a) shows three lines hypothesized to be the lines of action of the genioglossus muscle. Figure 6(b) shows the strains calculated along these deforming lines. The origin and insertions of the genioglossus muscle were selected based on the high-resolution static image shown in Fig. 6(a). The origin of the genioglossus muscle (point *G*) was chosen as the midpoint of the flat inner surface of the mandibular bone in the midsagittal slice. Three insertion points were selected in anterior (point 1), middle (point 2), and posterior (point 4) compartments of genioglossus. The numbers correspond to the points that were tracked in the previous section. Note that the insertion of the genioglossus is picked a fraction below the tongue surface, even though the true insertion points are at the tongue surface. This is done because of the reduced spatial resolution at the tongue surface in the tagged image, which often leads to errors in tracking at the surface.

f. Interpretation. Since there is no movement at point 7 (as seen from Fig. 5), the Lagrangian strains are entirely due to the surface motion of tissue points 1, 2, and 4. As a result, these compressions and extensions reflect the subtle differences of adjacent tongue surface regions. Such subtle difference are not captured in existing surface measuring modalities, and are unique to HARP-MRI. The length of line 4 (dotted) increases from frame 1 to frame 3, and then starts to reduce by frames 4 and 5. These first five frames correspond to the utterance /*di*/. From frame 5 till frame 8, line 4 continues to compress. Frames 5 through 8 correspond to the start of /*s*/. The maximum compression takes place at frame 8, after which line 4 starts to expand until the last time frame. Line 2 (solid), although similar to line 4, has greater expansion and lesser compression than line 4. Moreover, the expansion into /*u*/ begins one time-frame later for line 2 and is in a more vertical direction, indicating substantial differences between the motion of these two surface regions. Line 1 (dashed) is quite different. Its length is relatively unchanged during the utterance /*d*/, but there is dramatic increase in length from frames 5 through 9, which corresponds

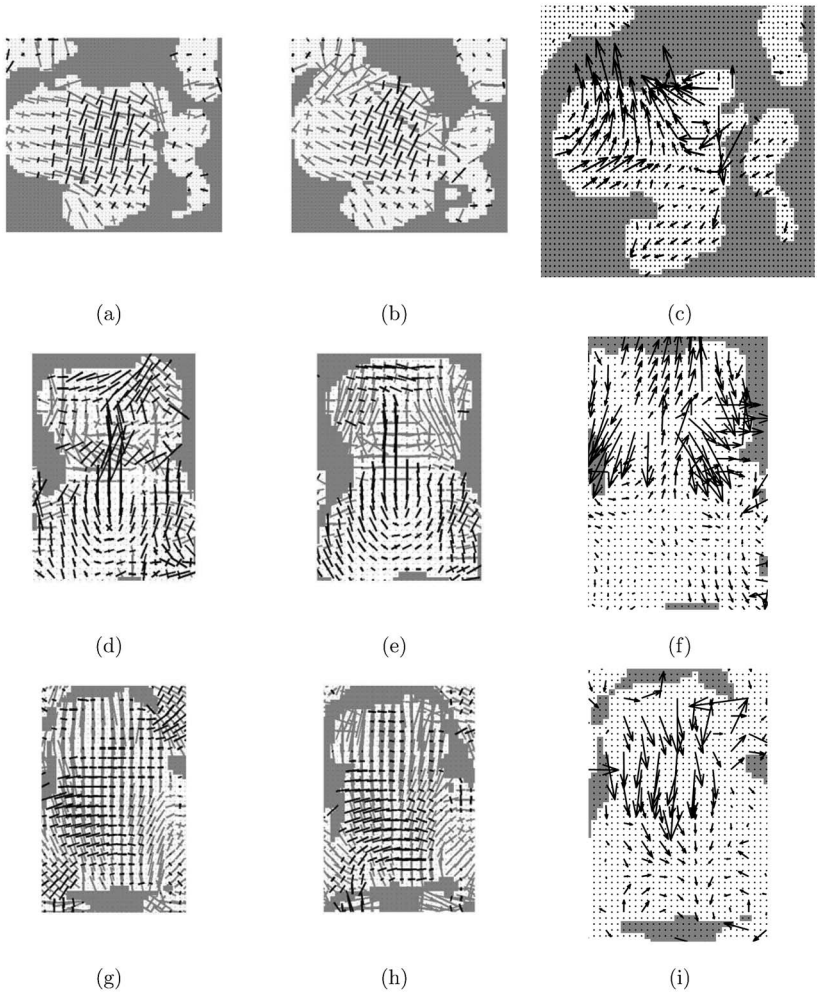


FIG. 7. The first two columns show the principal strains at time frames 9 and 10, respectively; black denotes compression and gray denotes expansion. The third column shows the velocity of tissue points calculated between time frames 9 and 10. While the first row shows the measurements for a midsagittal slice, the second and third rows show the measurements for an anterior coronal slice and a superior axial slice, respectively.

to /*su*/. This is consistent with the elevation of the entire tongue body, including the pharyngeal region. The maximum expansion occurs during /*u*/ at time frame 9, after which there is a decrease in length.

The user-defined measurements are useful to test hypotheses and to investigate relationships between the local surface and internal movements. For this data set the direction and rate of motion were consistent with three distinct surface regions containing independent spatiotemporal patterns of movement.

2. Data-driven measurements

Principal strains and tissue velocities are calculated from the data without any explicit modeling or user intervention. The data-driven measures help to corroborate the user-driven measures and to generate testable hypotheses.

a. Principal strains. Local stretch and compression in the interior of the tongue are objectively captured by the principal strains. The two principal directions are orthogonal and the strains along these directions capture the amount of maximal stretch or compression of a local tissue element. The strains are calculated on a regular grid on each pixel in the deformed tissue configuration. The principal strains can be derived from the Eulerian strain tensor E^* [Lai *et al.* (1996)], which is defined as follows:

$$E^* = \frac{1}{2}(I - B^{-1}), \tag{7}$$

where $B = FF^{\mathrm{T}}$ is the left Cauchy-Green strain tensor, $F = (\nabla \boldsymbol{\phi}(\mathbf{x},t))^{-1}\Omega^{\mathrm{T}}$ is the deformation gradient tensor, and $\Omega([\boldsymbol{\omega}_1 | \boldsymbol{\omega}_2]^{\mathrm{T}})$ is a matrix whose columns are two linearly independent tag direction vectors, namely horizontal and vertical. The matrix $\nabla \boldsymbol{\phi}(\mathbf{x},t)$ can be proven to be always invertible and hence E^* always exists. The eigenvectors of E^* represent the principal directions of strain. Using the definition of E^* [Lai *et al.* (1996)], we define strain in the principal direction as

$$S_i = 1 - \sqrt{\beta_i}, \quad i = [1,2], \tag{8}$$

where β_i is the ith eigenvalue of B^{-1}. Therefore, S_i represents the strains in the principal directions. More details about Eq. (8) can be found in Appendix B.

b. Visualization. The first and second columns in Fig. 7 show the principal strains corresponding to the motion from /u/ to /k/ and during /k/, respectively. The three rows correspond to the sagittal, coronal, and axial orientations. The principal strains are plotted in the form of a cross, with black denoting compression and gray denoting expansion. The length of the lines in the cross denotes the amount of strain, and the direction of the lines point to the directions of maxi-

mal strain. Although the strains are calculated at all pixels, strain is downsampled in the display to improve clarity.

c. Interpretation. In the midsagittal strains shown in the top row, there is a distinct difference in the local strain pattern in the anterior and posterior parts of the tongue. In the anterior and central parts, vertical compression is accompanied by either horizontal or oblique expansion. In contrast, in the posterior and the velar parts of the tongue there is expansion in both principal directions, though primarily horizontal in time frame 9. An interesting observation is that, even though the image is in the midsagittal plane, the direction of compression lines does not correspond to the direction of the genioglossus muscle. To better interpret the compressions an anterior coronal slice is considered, the position of which is shown in Fig. 3(d). The middle row of Fig. 7 shows the principal strain in the coronal slice. In both the coronal images, especially in time frame 10, one can see a midline vertical compression surrounded by a bilateral expansion. This pattern follows the anatomical distribution of the genioglossus muscle surrounded by the bilateral intrinsic muscles. This is consistent with genioglossus activation. There is a reduction in genioglossus compression between frames 9 and 10; the vertical lines are shorter and fewer in number. At the same time, expansion is seen laterally in the direction of verticalis. Both of these are consistent with the tongue elevation into /*k*/. The oblique lines below the tongue are consistent with the effects of mylohyoid contraction, which would further elevate the tongue into the /*k*/.

Figures 7(g) and 7(h) show the principal strains in time frames 9 and 10 for the axial slice, which is marked in Fig. 3(d). The lateral-to-medial compression in the axial images can be hypothesized to be because of the contraction of the transverse muscle, which can in turn explain the horizontal expansion in the sagittal plane and the bilateral vertical expansion in the coronal plane. Also, compression patterns around the pharynx resemble the activity of the styloglossus and the palatoglossus muscles. The combined visualization of the principal strains in the three orthogonal planes clearly corroborates the idea that the tongue is a muscular hydrostat, where a compression in one direction will result in an expansion in the another direction.

d. Velocity fields. The velocity of a tissue point at time t can be calculated from the HARP vector as follows:

$$\mathbf{v}(\mathbf{x},t) = -\left[\nabla \boldsymbol{\phi}(\mathbf{x},t)\right]^{-1} \frac{\partial \boldsymbol{\phi}}{\partial t}. \tag{9}$$

The calculation of velocity involves both the spatial gradient of phase within an image, and also the temporal gradient of phase across two time frames. In both cases, the forward difference operator is used.

e. Visualization. Figures 7(c), 7(f), and 7(i) show the velocity vectors between time frame 9 and 10, in the sagittal slice, coronal slice, and axial slice, respectively. These are the same slices that were considered for the strain computations. The length and direction of the arrow heads denote the magnitude and the direction of the velocity. Like the display of the principal strains, the vectors are displayed only on every other pixel for the sake of clarity.

f. Interpretation. The principal strains and velocity fields are complementary to each other. The principal strains indicate expansion and compression without indicating direction of motion, since the stretch may occur in either or both directions. On the other hand, velocity fields indicate direction and extent of motion, but do not indicate whether the tissue point motion was due to translation, expansion, or both.

Between time frames 9 and 10, the tongue was moving from /*u*/ to /*k*/. Examination of the sagittal velocity field in Fig. 7(c) shows the most salient features to be large upward motion of the upper tongue surface. The coronal image in Fig. 7(f) indicates that the upward motion occurs medially and that the sides of the tongue move laterally and downward. These patterns create an explicit map of the local tongue behaviors that comprise the surface elevation. For example, the motion from /*u*/ to /*k*/ requires tongue elevation and contact of the anterior-dorsum with the velum. There are a number of ways this could be executed. The tongue is a volume-preserving muscular hydrostat with interdigitated three-dimensional muscle architecture that allows complex deformations. Therefore, although the tongue can move rigidly by using the extrinsic muscles, local motion is another feasible option. Nonrigid local motion is an advantageous way to accomplish the constriction requirements of the target sound, while incorporating coarticulatory constraints on noncritical regions. The present data set indicates that the /*u*/ to /*k*/ movements actively incorporate features of nonrigid motion. First, the sagittal image shows that the tongue elevation is not uniform in direction between /*u*/ and /*k*/. Each region moves optimally to elevate the anterior-dorsum, which results in different directions of motion throughout the tongue. The posterior tongue actually, moves inward, an excellent volume-preserving mechanism to further facilitate anterior-dorsal elevation.

In the coronal slice, the upward motion is revealed to be primarily in the medial regions of the tongue, presumably because the palate is arched in the center and lower laterally, thus requiring less upward motion from the lateral margins of the tongue. Further, Fig. 7(c) shows that as the tongue contacts the palate, the tongue edges lower and spread out, suggesting recoil, or features of the next sound.

3. Complementary analysis between tissue motion and tissue compression

Comparison of the velocity fields with the principal strains reveals interesting relationships between translation and stretch of specific muscles. The two sagittal images [Figs. 7(a) and 7(b)] indicate homogeneous regions of strain in the anterior and posterior tongue from /*u*/ to /*k*/, with compression occurring only in the anterior tongue. The strain directions rotate from vertical/horizontal in direction [Fig. 7(a)] to oblique [Fig. 7(b)], and the upper surface of the tongue expands considerably between Figs. 7(a) and 7(b). These two regions translate into a velocity pattern [Fig. 7(c)] with at least three regions of surface motion (from front to back: backward, upward, and inward), as well as two horizontal divisions, the upper moving upward and the lower moving forward.

Frames 9 and 10 in the axial slice [Figs. 7(g)–7(i)] show patterns that are consistent with the sagittal data. Lateral-to-

medial shortening is consistent with both active contraction of transverse muscle and passive compression due to contact with the palate. The vector field in the axial plane [Fig. 7(i)] shows little in-plane motion in the posterior tongue between frames 9 and 10, consistent with motion orthogonal to the axial plane.

The coronal images add more information. The principal strains [Fig. 7(d) and 7(e)] show compression at the location of genioglossus (midline) and at the location of mylohyoid (bottom). Active contraction of genioglossus would oppose the upward motion seen in Fig. 7(f). The velocity field shows a reduction, or checking, in upward motion at the location of the GG compression. The tongue surface moves upward medially [Fig. 7(f)] and compresses left-to-right. This compression may be the passive response to the large lengthwise expansion of superior longitudinal required for the tongue surface to reach the palate. Compression seen in the mylohyoid region beneath the tongue [Figs. 7(d) and 7(e)] could reflect active contraction; mylohyoid has been found to contract during production of /k/ [Hirose (1971)]. The lateral tongue lowering in the velocity field [Fig. 7(f)] maps to the regions of expansion in the lateral tongue [Fig. 7(e)].

V. DISCUSSION

A. Advantages of using HARP-MRI

There are several technical advantages of analyzing tagged images using HARP-MRI. First, the motion measurements are not restricted to the points where tag lines intersect. Motion can be measured at the same resolution as the image resolution at every pixel. Second, since HARP-MRI uses only a small part of the Fourier spectrum of the tagged image to reconstruct motion. The ability to reconstruct motion from such small amount of Fourier spectral data is key in reducing the number of repetitions and also to improve the frame rate of the cine series. Third, since all these different measures are obtained from the same modality in the same scanning session, registration between modalities becomes unnecessary. Finally, HARP-MRI requires minimal human intervention and is computationally very fast. Tracking of one tissue point can be done within a second, and the calculation of strains and velocities for one tagged image can be measured within one minute.

Apart from the technical advantages, HARP-MRI brings a new perspective in the understanding of workings of the tongue. The combined analysis of the different motion measurements is a powerful tool in understanding tongue dynamics, and we clearly see three possible uses. First, plausible muscle activity can be inferred for testing with models. Second, models of motor control can be tested against the strain and velocity patterns. Third, multiplanar corroboration of muscle behavior can be observed.

B. Temporal resolution

The time resolution of 18 Hz, which is used in this study, is low when compared to the rate of normal speech. Even though we had the potential to collect data at nearly twice the rate (30 Hz), the speaker's temporal imprecision resulted in poor image quality and faded tags. At higher rates of data collection, even the slightest speaker imprecision between repetitions tends to cause improper ensemble combination. By slowing down the frame rate, we were effectively averaging over larger time units, thereby reducing the variability.

Reduction in frame rate is only a temporary solution to the speaker imprecision problem. We believe that data collected at higher frame rates have numerous advantages. Therefore, in spite of its current drawback with regards to speaker variability, we believe tMRI should be pursued. These drawbacks could be approached from three different directions. First, an image processing approach where a spatiotemporal registration of the four tagged data sets could be used. Second, an MR imaging approach could be used, where either the imaging sequences are designed to collect tagging data in one repetition (for example, grid tags with spiral acquisition) or the data from separate repetitions could be combined using retrospective gating methods. Third, an MR imaging approach where prospective gating could be done using the onset of speech.

C. Selection of speech materials

Temporal resolution is a key factor in deciding the speech materials that can be used for this study. HARP tracking makes a small motion assumption, which means that HARP tracking will be unable to track very large interframe motion. In order for HARP-MRI to track motion of a tissue point, the motion should be less than one half the tag separation between two consecutive frames. Therefore the temporal resolution, which defines the time interval between two consecutive frames, should be adequate for the particular utterance. Our experience has shown that tracking error is also likely to occur at the surface of the tongue because these points are not between two tags.

D. Future work

In this work, we have used 16 repetitions to fully track the motion of the tongue. The number of repetitions, however, needs to be further reduced in order to extend the scope of this study to nonexpert speakers and patients. In order to do this, research into the theoretical limits of HARP is ongoing [Parthasarathy and Prince (2004)]. Understanding of these issues will help predict whether longer and more complex speech patterns can be measured using HARP-MRI.

Measuring the motion and deformation is only the first part in understanding tongue dynamics. In this work, tissue compressions were used to hypothesize muscle activity. We plan to use these compressions in representational models of tongue control. It is important, however, to understand the assumptions made in moving from a pure kinematic measurement to a dynamic model. The observed compressions were assumed to be contractions and the effects of muscle contractions were also assumed to be direct. We acknowledge that there are many possible strategies of motor equivalence. Future work on representational and predictive models should include multiple synergistic patterns to achieve the compressions and surface shapes seen in the data.

VI. CONCLUSION

In this paper, we have presented a new method to measure the motion of the tongue during speech. We have demonstrated the use of HARP-MRI on tagged MRI images. HARP-MRI is an automatic tag tracking technique that estimates motion and deformation at every pixel in a tagged image. The motion measurements include trajectories of tissue points, principal strains at every pixel, strains in the line-of-action of a particular muscle, and tissue velocities. The different motion measurements can be visualized in a variety of ways and are complementary to each other. These measures were calculated on a 3D data set with images in the sagittal, axial, and coronal planes. Even though our data were collected in 3D, the measurements are still on 2D planes. We are actively pursuing 3D imaging and image processing approaches to measure 3D motion using tagging. The final goal is to be able to input these 3D motion measures into representational models in order to better understand the overall muscular compressional patterns in the production of speech.

ACKNOWLEDGMENTS

This research was funded in part by National Institute on Deafness and Other Communication Disorders (NIDCD) Grant R01 DC01758. We also acknowledge Rao Gullapalli, Jiachen Zhuo, and Harsh Agarwal for their contribution in collecting the tagged MR data.

APPENDIX A: DATA ACQUISITION MATRIX

The field of view is square with 200 mm along the tagging direction and 200 mm in the perpendicular direction. The period of each tag cycle is 10 mm. Therefore, there are 20 (200/10) tag cycles across the field of view. This implies that in the tagging direction the harmonic peaks will be centered at ±20 in Fourier space. If it is assumed that HARP processing needs only ±10 Fourier space samples on either side of the harmonic peak, then all we need to sample is 60 points, from −30 to +30. We sample −32 to +31, a total of 64 points (a power of 2) to make the inverse fast Fourier transform (FFT) algorithm efficient. Essentially, we divide Fourier space into three parts along the tagging direction. While the central third is centered around the center of Fourier space, the negative and positive thirds are centered around the negative and positive harmonic peaks respectively. Along the direction perpendicular to the tags, 23 (−11 to 11) Fourier points are sampled.

APPENDIX B: STRAIN IN PRINCIPAL DIRECTION

In strict continuum mechanical sense, the term principal strains are not defined for finite deformation. In this paper, we use the term principal strain to mean strains in the directions that have no shear, i.e., principal directions. The terms "principal strain" and "strain in the principal directions" are used interchangeably.

Let F be the deformation gradient tensor. Using polar decomposition, $F=VR$, where V is a symmetric tensor called the right stretch and R is an orthogonal tensor that characterizes the rotational component. The left Cauchy-Green strain tensor B is defined as $B=V^2$.

Let λ_i, $i=[1,2]$ be the ith eigenvalue of V. Then, the eigenvalues of B^{-1} are

$$\beta_i = \frac{1}{\lambda_i^2}. \tag{B1}$$

The Eulerian strain tensor, E^* is defined as, $E^*=\frac{1}{2}(I-B^{-1})$. Therefore, the eigenvalues of E^* are given as $\frac{1}{2}(1-\beta_i)$, $i=[1,2]$. Since B is symmetric, E^* is also symmetric and hence can be diagonalized when E^* is expressed along its eigenvectors e_1 and e_2 as its basis. Given the definition of the diagonal terms in the E^* [Lai *et al.* (1996)], we get

$$\frac{\|ds_i\|^2 - \|dS_i\|^2}{\|ds_i\|^2} = (1-\beta_i), \tag{B2}$$

where dS_i is an infinitesimally small piece of tissue in the ith direction when the tags were applied and ds_i is the same tissue at a later time after deformation. From the above relation, the unit Eulerian strain along the principal direction, S_i, can be derived as

$$S_i = \frac{\|ds_i\| - \|dS_i\|}{\|ds_i\|} = 1 - \sqrt{\beta_i}, \quad i=[1,2]. \tag{B3}$$

Axel, L., and Dougherty, L. (**1989a**). "Heart wall motion: Improved method of spatial modulation magnetization for MR imaging," Radiology **172**, 349–350.

Axel, L., and Dougherty, L. (**1989b**). "MR imaging of motion with spatial modulation of magnetization," Radiology **171**, 841–845.

Badin, P., Bailly, G., Reveret, L., Baciu, M., and Segebarth, C. (**2002**). "Three-dimensional linear articulatory modeling of tongue lips and face, based on MRI and video images," J. Phonetics **30**(3), 533–554.

Baer, T., Gore, J., Boyce, S., and Nye, P. (**1987**). "Application of MRI to the analysis of speech production," Magn. Reson. Imaging **5**(5), 1–7.

Cha, H., and Patten, B. M. (**1988**). "Amyotrophic lateral sclerosis abnormalities of the tongue on magnetic resonance imaging," Am. Neurological Association **25**, 468–472.

Dang, J., and Honda, K. (**1997**). "A physiological model of the tongue and jaw for simulating deformation in the midsagittal and parasagittal planes," J. Acoust. Soc. Am. **102**, 3167(A).

Dang, J., and Honda, K. (**2004**). "Construction and control of a physiological articulatory model," J. Acoust. Soc. Am. **115**, 853–870.

Demolin, D., Metens, T., and Soquet, A. (**2000**). "Real time MRI and articulatory coordinations in vowels," in Proceedings of International Speech Production Seminar.

Denney, T. S. Jr., and McVeigh, E. R. (**1997**). "Model-free reconstruction of three-dimensional myocardial strain from planar tagged MR images," J. Magn. Reson Imaging **7**, 799–810.

Dick, D., Ozturk, C., Douglas, A., McVeigh, E., and Stone, M. (**2000**). "Three-dimensional tracking of tongue motion using tagged-MRI," in International Society for Magnetic Resonance in Medicine, 8th Scientific Meeting and Exhibition, Denver, Co.

Eastwood, P. R., Allison, G. T., Shepherd, K. L., Szollosi, I., and Hillman, D. R. (**2003**). "Heterogenous activity of the human genioglossus muscle assessed by multiple bipolar fine-wire electrodes," J. Appl. Physiol. **94**, 1849–1858.

Engwall, O. (**1999**). "Modeling of the vocal tract in three dimensions," in Proceedings of the Eurospeech Meeting, pp. 113–116.

Engwall, O. (**2003a**). "Combining MRI, EMA and EPG measurements in a three-dimensional tongue model," Speech Commun. **41**(2–3), 303–329.

Engwall, O. (**2003b**). "A revisit to the application of MRI to the analysis of speech production—testing our assumptions," in 6th International Seminar on Speech Production.

Fischer, S. E., McKinnon, G. C., Maier, S. E., and Boesiger, P. (**1993**).

"Improved myocardial tagging contrast," Magn. Reson. Med. **30**, 191–200.

Greenwood, A. R., Goodyear, C. C., and Martin, P. A. (**1992**). "Measurements of vocal tract shapes using magnetic resonance imaging," in IEE Proceedings, pp. 5553–5560.

Guttman, M. A., Prince, J. L., and McVeigh, E. R. (**1994**). "Tag and contour detection in tagged MR images of the left ventricle," IEEE Trans. Med. Imaging **13**(1), 74–88.

Hashimoto, K., and Suga, S. (**1986**). "Estimation of the muscular tensions of the human tongue by using a three-dimensional model of the tongue," J. Acoust. Soc. Jpn. **7**, 39–46.

Hirose, H. (**1971**). "Electromyography of the articulatory muscles: Current instrumentation and technique," Haskins Lab Status Rep Speech Res. **25–26**, 73–86.

Honda, K. (**1996**). "Organization of tongue articulation for vowels," J. Phonetics **24**, 39–52.

Honda, K., and Kusakawa, N. (**1997**). "Compatibility between auditory and articulatory representations of vowels," Acta Oto-Laryngol., Suppl. **532**, 103–105.

Hoole, P., Geng, C., and Winkler, R. (**2001**). "Towards a speaker-independent representation of tongue-posturing for speech," in 4th International Speech Motor Conference, edited by B. Maassen, W. Hulstijn, R. Kent, H. Peters, and P. Van Lieshout, pp. 138–141.

Hoole, P., Wismuller, A., Leinsinger, G., Kroos, C., Geumann, A., and Inoue, M. (**2000**). "Analysis of tongue configuration in multi-speaker, multi-volume MRI data," in Proceedings of 5th Speech Production Seminar, pp. 157–160.

Iny, D., Lundberg, A., Levine, W. S., and Stone, M. (**2001**). "The dynamics and control of the tongue during speech," in Proceedings of the American Control Conference, pp. 246–251.

Kakita, Y., Fujimura, O., and Honda, K. (**1985**). "Computation of mapping from the muscular contraction pattern to formant in vowel space," in *Phonetic Linguistics; Essays in honor of Peter Ladefoged*, edited by V. Fromkin (Academic Press, NY).

Kerwin, W. S., and Prince, J. L. (**2002**). "Kriging filter for space-time interpolation," Adv. Imaging Electron Phys. **124**, 139–193.

Kerwin, W. S., Osman, N. F., and Prince, J. L. (**2000**). "Image processing and analysis in tagged cardiac MRI," in *Handbook of Medical Image Processing and Analysis*, edited by I. N. Bankman (Academic Press), pp. 375–392.

Kiritani, S., Miyawaki, K., Fujimura, O., and Miller, J. (**1976**). "A computational model of the tongue," Ann. Bull. RILP Univ. Tokyo **10**, 243–251; URL⟨http://www.umin.ac.jp/memorial/rilp-tokyo/ ⟩. Last accessed 12/6/06.

Kuijer, J. P., Jansen, E., Marcus, J. T., Rossum, A. C. V., and Heethaar, R. M. (**2001**). "Improved harmonic phase myocardial strain maps," Magn. Reson. Med. **46**(5), 993–999.

Lai, W. M., Rubin, D., and Krempl, E. (**1996**). *Introduction to Continuum Mechanics* (Butterworth-Heinemann, London).

Lakshminarayanan, A. V., Lee, S., and McCutcheon, M. J. (**1991**). "MR Imaging of the vocal tract during Vowel Production," J. Magn. Reson Imaging **1**, 71–76.

Levine, W. S., Torcaso, C. E., and Stone, M. (**2005**). "Controlling the shape of a muscular hydrostat: A tongue or tentacle," *New Directions in Control Theory and Applications*, edited by W. P. Dayawansa, A. Lindquist, and Y. Zhou, Lecture Notes in Control and Information Sciences, Vol. 321 (Springer Publishing Co.), pp. 207–222.

Lufkin, R., Christianson, R., and Hanafee, W. (**1987**). "Normal magnetic resonance imaging anatomy of the tongue, oropharynx, hypopharnyx and larnyx," Dysphagia **1**, 119–l27.

Mady, K., Sader, R., Zimmermann, A., Hoole, P., Beer, A., Zeilhofer, H., and Hanning, C. (**2002**). "Assesment of consonant articulation in glossectomee speech by dynamic MRI," in Proceedings of ICSLP 2002.

Masaki, S., Tiede, M., Honda, K., Shimada, Y., Fujimoto, I., Nakamura, Y., and Ninomiya, N. (**1999**). "MRI-based speech production study using a synchronized sampling method," J. Acoust. Soc. Jpn. **20**, 375–379.

McKenna, K. M., Jabour, B. A., Luftkin, R., and Hanafee, W. (**1990**). "Magnetic resonance imaging of the tongue and oropharynx," Top Magn. Reson Imaging **2**, 49–59.

Miyawaki, K., Hirose, H., Ushijima, T., and Sawashima, M. (**1997**). "A preliminary report on the electromyographic study of the activity of lingual muscles," Ann. Bull. RILP Univ. Tokyo **9**, 91–1066; URL⟨http://www.umin.ac.jp/memorial/rilp-tokyo/ ⟩. Last accessed 12/6/06.

Moore, C. C., O'Dell, W. G., McVeigh, E. R., and Zerhouni, E. A. (**1992**). "Calculation of three-dimensional left ventricular strains from biplanar tagged MR images," J. Magn. Reson Imaging **2**, 165–175.

Napadow, V. J., Chen, Q., Wedeen, V. J., and Gilbert, R. J. (**1999a**). "Biomechanical basis for lingual muscular deformation during swallowing," Am. J. Physiol. **277**, G695–G701.

Napadow, V. J., Chen, Q., Wedeen, V.J., and Gilbert, R. J. (**1999b**). "Intramural mechanics of the human tongue in association with physiological deformations," J. Biomech. **322**, 1–12.

Narayanan, S., Nayak, K., Lee, S., Sethy, A., and Byrd, D. (**2004**). "An approach to real-time magnetic resonance imaging for speech production," J. Acoust. Soc. Am. **115**, 1771–1776.

NessAiver, M., and J. L. Prince (**2003a**). "Magnitude image CSPAMM reconstruction (MICSR)," Magn. Reson. Med. **50**(2), 331–342.

NessAiver, M., and Prince, J. L. (**2003b**). "Visualization of myocardial motion using MICSR trinary checkerboard display," in Inf. Processing in Medical Imaging **18**, 573–585.

Niitsu, M., Kumada, M., Niimi, S., and Itai, Y. (**1992**). "Tongue movement during phonation: A rapid quantitative visualization using tagging snapshot MRI imaging," Annual Bulletin of the Research Institute of Logopedics and Phoniatrics, University of Tokyo, **26**, 149–156; ⟨URLhttp://www.umin.ac.jp/memorial/rilp-tokyo ⟩. Last accessed 12/6/06.

Niitsu, M., Kumada, M., Campeau, G., Niimi, S., Riederer, S. J., and Itai, Y. (**1994**). "Tongue displacement: Visualization with rapid tagged magnetization-prepared MR imaging," Radiology **191**, 578–580.

Osman, N. F., and Prince, J. L. (**2000**). "Visualizing myocardial function using HARP MRI," Phys. Med. Biol. **45**(6), 1665–1682.

Osman, N. F., McVeigh, E. R., and Prince, J. L. (**2000**). "Imaging heart motion using harmonic phase MRI," IEEE Trans. Med. Imaging **19**(3), 186–202.

Osman, N. F., Kerwin, W. S., McVeigh, E. R., and Prince, J. L. (**1999**). "Cardiac motion tracking using CINE harmonic phase (HARP) magnetic resonance imaging," Magn. Reson. Med. **42**, 1048–1060.

Ozturk, C., and McVeigh, E. R. (**2000**). "Four-dimensional B-spline based motion analysis of tagged MR images: Introduction and *in vivo* validation," Phys. Med. Biol. **45**, 1683–1702.

Parthasarathy, V., and Prince, J. L. (**2004**). "Dynamic range of harmonic phase magnetic resonance imaging (HARP-MRI)," in Proceedings of the International Society of Biomedical Imaging, pp. 884–887.

Radeva, P., Amini, A., and Huang, J. (**1997**). "Deformable B-solids and implicit snakes for 3D localization and tracking of MRI-SPAMM data," Comput. Vis. Image Underst. **66**(2), 163–178.

Sabisky, J. P., Butler, J. E., Fogel, R. B., Taylor, J. L., Trinder, J. A., Davis, D. P., and Gandevia, S. C. (**2006**). "Tonic and phasic respiratory drives to human genioglossus motoneurons during breathing," J. Neurophysiol. **95**(4), 2213–2221.

Sampath, S., Derbyshire, A., Osman, N. F., Atalar, E., and Prince, J. L. (**2003**). "Real-time imaging of two-dimensional cardiac strain using a fast HARP pulse sequence," Magn. Reson. Med. **50**(1), 154–163.

Sauerland, E. K., and Harper, R. M. (**1976**). "The human tongue during sleep: Electromyographic activity of the genioglossus muscle," Exp. Neurol. **51**, 160–170.

Shadle, C., Mohammad, M., Carter, J., and Jackson, P. (**1999**). "Dynamic magnetic resonance imaging: New tools for speech research," in *proceedings of the 14th Int. Cong. Phon. Sci.*, pp. 623–626.

Shimada, Y., Fujimoto, I., Takemoto, H., Takano, S., Masaki, S., Honda, K., and Takeo, K. (**2002**). "4D-MRI using the synchronized sampling method," Jpn. J. Radiol. Technol. **53**(12), 1592–1598.

Stone, M. (**1999**). "Laboratory techniques in speech science," in *A Handbook of Phonetic Science*, edited by W. Hardcastle, A. Marchal, and N. Hewlett, Blackwell Handbooks in Linguistic Series (Blackwell, Ltd., Oxford), pp. 11–32.

Stone, M., Dick, D., Davis, E., Douglas, A., and Ozturk, C. (**2000**). "Modelling the internal tongue using principal strains," in *Proceedings of the 5th Speech Production Seminar, Kloster-Seeon*, pp. 133–136, Germany.

Stone, M., Davis, E. P., Douglas, A. S., NessAiver, M., Gullapalli, R., Levine, W. S., and Lundberg, A. (**2001a**). "Modeling the motion of the internal tongue from tagged cine-MR images," J. Acoust. Soc. Am. **109**(6), 2974–2982.

Stone, M., Davis, E. P., Douglas, A. S., NessAiver, M., Gullapalli, R., Levine, W., and Lundberg, A. J. (**2001b**). "Modeling tongue surface contours from cine-MRI images," J. Speech Lang. Hear. Res. **44**(5), 1026–1040.

Story, B. H., Titze, I. R., and Hoffman, E. A. (**1996**). "Vocal tract area functions from magnetic resonance imaging," J. Acoust. Soc. Am. **10**(3),

537–554.
Story, B. H., Titze, I. R., and Hoffman, E. A. (**1998**). "Vocal tract area function for an adult female speak based on volumetric imaging," J. Acoust. Soc. Am. **104**(1), 471–487.
Sulter, A. M., Miller, D. G., Wolf, R. F., Schutte, H. K., Wit, H. P., and Mooyaart, E. L. (**1992**). "On the relation between the dimensions and resonance characteristics of the vocal-tract—a study with MRI," Magn. Reson. Med. **10**(3), 365–373.
Takashima, S., Ikezoe, J., Harada, K., Akai, Y., Hamada, S., Arisawa, J., Morimoto, S., Masaki, N., Kozuka, T., and Maeda, H. (**1989**). "Tongue cancer: Correlation of imaging and sonography with pathology," American Journal of Neuroradiology **10**, 419–424.
Wein, B., Drobnitzky, M., and Klajman, S. (**1990**). "Magnetresonanztomographie and Sonographie bei der Lutbildung (Magnetic Resonance Imaging and Sonography During Phonation)," Fortschr Roentgensr **153**, 408–412.
Wilhelms-Tricarico, R. (**1995**). "Physiological modeling of speech production: Methods for modeling soft-tissue articulators," J. Acoust. Soc. Am. **97**, 3805–3898.
Young, A. A., Kraitchman, D. L., Dougherty, L., and Axel, L. (**1995**). "Tracking and finite element analysis of stripe deformation in magnetic resonance tagging," IEEE Trans. Med. Imaging **14**(3), 413–421.

Effects of masking noise on vowel and sibilant contrasts in normal-hearing speakers and postlingually deafened cochlear implant users

Joseph S. Perkell[a)]
Speech Communication Group, Research Laboratory of Electronics, and Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Room 36-511, 50 Vassar Street, Cambridge, Massachusetts 02139 and Department of Cognitive and Neural Systems, Boston University, Boston, Massachusetts 02215

Margaret Denny
Speech Communication Group, Research Laboratory of Electronics, Massachusetts Institute of Technology, Room 36-511, 50 Vassar Street, Cambridge, Massachusetts 02139

Harlan Lane
Department of Psychology, Northeastern University, Boston, Massachusetts 02115 and Speech Communication Group, Research Laboratory of Electronics, Massachusetts Institute of Technology, Room 36-511, 50 Vassar Street, Cambridge, Massachusetts 02139

Frank Guenther
Department of Cognitive and Neural Systems, Boston University, Boston, Massachusetts 02215 and Speech Communication Group, Research Laboratory of Electronics, Massachusetts Institute of Technology, Room 36-511, 50 Vassar Street, Cambridge, Massachusetts 02139

Melanie L. Matthies
Department of Speech, Language and Hearing Sciences, Boston University, 635 Commonwealth Avenue, Boston, Massachusetts 02215 and Speech Communication Group, Research Laboratory of Electronics, Massachusetts Institute of Technology, Room 36-511, 50 Vassar Street, Cambridge, Massachusetts 02139

Mark Tiede
Speech Communication Group, Research Laboratory of Electronics, Massachusetts Institute of Technology, Room 36-511, 50 Vassar Street, Cambridge, Massachusetts 02139 and Haskins Laboratories, New Haven, Connecticut 06511

Jennell Vick
Department of Speech and Hearing Sciences, University of Washington, Seattle, Washington 98015

Majid Zandipour
Speech Communication Group, Research Laboratory of Electronics, Massachusetts Institute of Technology, Room 36-511, 50 Vassar Street, Cambridge, Massachusetts 02139 and Department of Cognitive and Neural Systems, Boston University, Boston, Massachusetts

Ellen Burton
Johns Hopkins School of Public Health and the Maryland Association of County Health Officers, Baltimore, MD 21205

(Received 28 February 2006; revised 9 October 2006; accepted 9 October 2006)

The role of auditory feedback in speech production was investigated by examining speakers' phonemic contrasts produced under increases in the noise to signal ratio (N/S). Seven cochlear implant users and seven normal-hearing controls pronounced utterances containing the vowels /i/, /u/, /ɛ/ and /æ/ and the sibilants /s/ and /ʃ/ while hearing their speech mixed with noise at seven equally spaced levels between their thresholds of detection and discomfort. Speakers' average vowel duration and SPL generally rose with increasing N/S. Average vowel contrast was initially flat or rising; at higher N/S levels, it fell. A contrast increase is interpreted as reflecting speakers' attempts to maintain clarity under degraded acoustic transmission conditions. As N/S increased, speakers could detect the extent of their phonemic contrasts less effectively, and the competing influence of

[a)]Electronic mail: perkell@speech.mit.edu

 0001-4966/2007/121(1)/505/14/$23.00

economy of effort led to contrast decrements. The sibilant contrast was more vulnerable to noise; it decreased over the entire range of increasing N/S for controls and was variable for implant users. The results are interpreted as reflecting the combined influences of a clarity constraint, economy of effort and the effect of masking on achieving auditory phonemic goals—with implant users less able to increase contrasts in noise than controls. © *2007 Acoustical Society of America.*
[DOI: 10.1121/1.2384848]

PACS number(s): 43.70.Mn, 43.70.Bk, 43.70.Dn, 43.66.Ts, 43.70.Fq [BHS] Pages: 505–518

I. INTRODUCTION

There is a substantial body of research concerning the effects of adverse speaking conditions on speech parameters. When the speaker perceives a deterioration in signal to noise ratio, either because of reduced signal levels or increased noise levels, that speaker will increase speaking sound level (Lane and Tranel, 1971; Van Summers *et al.*, 1988; Black, 1951; Hanley and Steer, 1949; Tartter *et al.*, 1993) and segmental duration (Van Summers *et al.*, 1988; Hanley and Steer, 1949). Utterances produced under such adverse conditions are more intelligible than those produced under optimal transmission conditions (Van Summers *et al.*, 1988; Dreher and O'Neill, 1958; Peters, 1955; Draegert, 1951).

These changes in sound level and durations under adverse conditions are consistent with those produced under instructions to speak clearly, as Lane *et al.* (1997) and Van Summers *et al.* (1988) noted. (There are, however, considerable differences among talkers: Hazan and Markham, 2004; Ferguson, 2004; Perkell *et al.*, 2002; Gagné and Tye-Murray, 1994.) Under clear speech instructions, vowel amplitudes and durations increase (Picheny *et al.*, 1986; Liu *et al.*, 2004). Furthermore, like speaking under adverse conditions, clear speech is also more intelligible than conversational speech (Picheny *et al.*, 1985, 1986; Chen *et al.*, 1983; Liu *et al.*, 2004; Krause and Braida, 2003; Payton *et al.*, 1994; Ferguson and Kewley-Port, 2002). Inference from similarities between clear speech and speaking under adverse conditions suggests that speakers may respond to clear speech instructions as though they were speaking under adverse conditions.

There is, however, an important difference between the changes in speech induced by instructions to speak clearly and those induced by adverse speaking conditions. Under clear speech instructions, phonemic contrasts are enhanced (Chen, 1980; Chen *et al.*, 1983; Moon and Lindblom, 1989; Picheny *et al.*, 1986). However, under more adverse speaking conditions, phonemic contrasts are characteristically degraded. For example, when the noise to signal ratio (N/S) is increased by subjecting normal-hearing speakers to loud masking noise, the speakers' vowel contrasts are reduced (cf. Bond *et al.*, 1989; Van Summers *et al.*, 1988). Likewise, if little or no signal can be heard, as in profound late-onset hearing loss, vowel contrasts (cf., Waldstein, 1990; Smyth *et al.*, 1991; Richardson *et al.*, 1993; Plant, 1984; Langereis *et al.*, 1997; Lane *et al.*, 2005) and sibilant contrasts (Lane and Webster, 1991; Matthies *et al.*, 1994) are also reduced compared to speakers with normal hearing (also see Kishon-Rabin *et al.*, 1999 on vowels).

The preceding considerations lead us to expect that speaking sound level will increase monotonically with N/S, whereas phoneme contrast distance will show an initial increase followed by a decline. These relations are schematized for vowels in Fig. 1. We define contrast distance for vowels as the average of Euclidean distances between all possible vowel pairs in acoustic (mel 1 × mel 2) space. In Fig. 1(A), for speakers with normal hearing, the monotonic growth of SPL with N/S is shown by the solid line, while a hypothesized inflected phoneme contrast function is shown by a dotted-dashed line. We expect somewhat different contrast-distance functions in persons who have become profoundly deaf postlingually and then have had hearing partially restored with a cochlear implant. Their experience while deaf is likely to have led to reduced vowel contrasts (see above references). Moreover, the somewhat distorted hearing they receive from their cochlear implants may make them less able, compared to speakers with normal hearing, to use auditory feedback to help increase contrasts. Figure 1(B) schematizes hypothetical relations from cochlear implant users, with phoneme contrast at one-month postimplant shown by a dashed function, and at 1 year, by a dotted function. The differences between these two functions in overall levels of contrast and in the noise levels at which the inflection points occur reflect prior observations that experience with an implant can lead to contrast improvements (cf. Perkell *et al.*, 2001; Langereis *et al.*, 1997; Kishon-Rabin *et al.*, 1999). Note that with the present state of our knowledge we have no basis for specifying particular functional shapes for the three contrast-distance functions in Fig. 1—beyond the claims that the functions have (a) different overall levels, (b) a downturn, and (c) a certain ordering of that downturn for the three experimental conditions (controls, implant users at 1 month, and 1 year).

The kind of functional relations illustrated in Fig. 1 between N/S and phonemic contrast have not been reported previously, nor have speaking sound level and phonemic

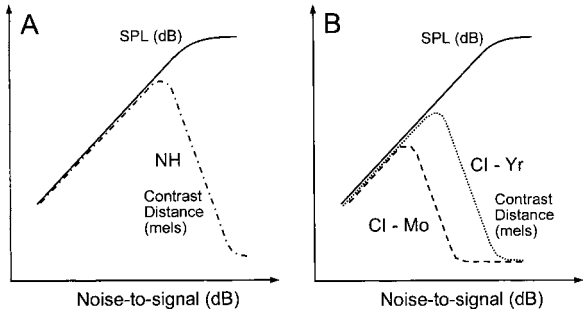


FIG. 1. Schematic illustration of hypotheses for normal-hearing speakers (panel A) and cochlear implant users (panel B) recorded at 1 month and 1 year post-implant.

contrast been examined together under this type of intervention.[1] In determining these functions for the cases of vowel and sibilant contrasts, the present experiment seeks to answer three general questions. First, we ask whether, in accord with Fig. 1, sound level functions are monotonic whereas contrast-distance functions are inflected, revealing a N/S threshold at which speakers typically stop increasing contrast and start reducing it. Second we ask what will be the behavior of speakers with late-onset deafness who have had some hearing restored recently by cochlear prosthesis. Several studies were cited above indicating that implant users produce reduced vowel contrasts. Consequently, the vowel contrasts produced by the implant users in this experiment may prove to be less robust in the face of increasing N/S than those produced by speakers with normal hearing. In other words, the function relating phoneme contrast to N/S may inflect at lower values of N/S for implant users than hearing controls. If so, will the function inflect at lower values in implant users when measured at 1 month than when measured after a year's experience with the implant (as postulated in Fig. 1)? Our third question is whether the findings for vowels and sibilants will differ from one another, and if so, how.

Clarity, effort, and the reduction in contrast: We speculate that the contrast distance functions schematized in Fig. 1 are the product of two underlying functions, one with positive slope, the other with negative slope, resulting in the inflection of the combined function. The function with positive slope would reflect a clarity effect—greater contrast distance under increasingly adverse conditions for communication. Underlying it is presumably the speaker's aim to communicate successfully despite increasing N/S. Underlying the negative slope—reduced phonemic contrast with increasing noise—is hypothetically a principle of least effort (Lindblom, 1990). As N/S increases, speakers are less and less able to hear the auditory consequences of their articulations, so without feedback their contrast distances fall, hypothetically due to a predominating effect of least effort. Consistent with these ideas, Perkell *et al.* (2002) found that when subjects were instructed to speak clearly some of them increased effort, as indexed by peak speed of articulation.

Economy of effort is a principle that guides speech motor control in the DIVA model of speech motor planning ("Directions into Velocities of Articulators;" Guenther, 1995; Guenther *et al.*, 1998, 2006; Perkell *et al.*, 2000), which we use as a framework for elaborating the current hypotheses and interpreting the results. In the model, phonemic goals are regions in auditory-temporal and somatosensory-temporal spaces. The goal regions are acquired and maintained with the use of auditory feedback, and speech movements from one goal to the next are programmed by feedback and feedforward subsystems. In the early stages of speech acquisition, feedback control predominates. As the sensory goals and feedforward commands are refined, movements become controlled mainly by the feedforward subsystem (although the feedback system remains available in case it is needed).

We noted above that when speakers encounter adverse speaking conditions (such as a noisy environment), they tend to speak more slowly, louder, and if possible, with increased contrast, in order to maintain intelligibility. Slower speech allows time for increased engagement of the feedback subsystem, which can help enhance or maintain contrast by providing information about the auditory consequences of the speech movements and allowing for feedback-based error correction if goals are not reached under feedforward control.

When hearing is lost in adulthood, the goal regions and feedforward commands may deteriorate gradually, leading to some diminution of phoneme contrasts. However, because the goals and feedforward commands are relatively robust and because somatosensory goals remain largely intact, basic phonemic identity is preserved.

The preceding reasoning leads to the prediction that under increasing N/S, vowel contrast will grow at the cost of expending more articulatory effort for as long as contrasts can be perceived by the speakers; then, as contrast perception and the use of auditory feedback control become more difficult, a threshold will be reached at which the balance begins to increasingly favor least effort and contrasts diminish (cf. Fig. 1).

Thus, the current study was designed to test the following hypotheses:

Hypothesis (1): In agreement with previous findings, over some range of increasing N/S values, speakers will not only speak more loudly and more slowly, they will also hyperarticulate, increasing contrast distance up to a point.

Hypothesis (2): At intermediate levels of noise, when speakers' auditory feedback of their own speech is masked to an extent that presumably begins to compromise their perception of phoneme contrasts, those speakers will continue to increase SPL toward a maximum level but their contrast distances will begin to fall.

Hypothesis (3): Contrast distance functions will inflect at higher N/S values for normal-hearing speakers than those for cochlear implant users, indicating that the normal-hearing speakers are less vulnerable to degradation of transmission conditions, presumably because they have robust and intact auditory goals and feedforward control mechanisms (cf. Guenther *et al.*, 1995, 2006; Perkell *et al.*, 2000) as well as better hearing.

Hypothesis (4): Implant users' contrast distance functions will inflect at higher N/S values after a year's implant use than after a month, indicating that their speech is less vulnerable to degradation of transmission conditions after experience with the implant—presumably because of retuning of auditory feedback and phonemic contrasts.

II. METHODS

A. Participants

The hypotheses set forth above were tested with two groups of participants. An experimental group comprised of five male and two female postlingually deaf, adult volunteers, who received cochlear implants at an average age of 52. The implant was either the Clarion (Advanced Bionics, CIS strategy; Wilson *et al.*, 1995) or the nucleus device (Cochlear Corporation; Blamey *et al.*, 1987; McKay and McDermott, 1993). The implant users were referred to our labora-

TABLE I. Participant characteristics.

1	Implant User ID	MM	MO	FJ	MK	FI	MJ	MP
2	Etiology	Noise (WWII)	Blood clot	Infection	Unknown	Auto-immune	Hereditary	Hereditary
3	Age at onset of hearing change	20	60	5	18	19	Early 30s	Birth
4	Age at onset of profound loss	72	67	45	28	54	43	26
5	Age at implantation	78	72	46	28	56	49	36
6	Hearing aid used pre-CI: L, R, both	Both	Left	None	None	None	Left	Both
7	Implant; ear	Clarion; R	Clarion C11 HiFocus; R	Clarion C11 HiFocus	Clarion Auria	Nucleus-24	Clarion C11 HiFocus	Clarion C11 HiFocus; R
8	Processing strategy: 1 month	CIS	CIS	SAS (8 chan.) Input dyn. range: 55 dB	HiRes-S	(not available)	CIS	CIS
9	Processing strategy: 1 year	CIS (no changes)	HiRes-P	HiRes-S Input dyn. range: 60 dB	HiRes-S (changed rate, pulse width)	SPEAK	HiRes-S	CIS (no changes)
10	Vowel perception (pre, year) % correct	32, 77	51[a]	26, 55	25, 95	29, 62	35, 96	27, 74
11	Consonant perception (pre, year) % correct	32, 50	49[a]	25, 57	24, 92	26, 44	21, 74	21, 74
12	Noise level: 1, 7; range, 1 mo. (dB)	61, 82; 21	45, 95; 50	38, 90; 52	39, 75; 36	51, 75; 24	61, 88; 27	40, 94; 54
13	Noise level: 1, 7; range, 1 yr. (dB)	53, 93; 40	53, 94; 41	70, 95; 25	51, 81; 30	59, 91; 32	59, 83; 24	73, 95; 22
14	Change in dynamic range (dB)	19	−9	−27	−6	8	−3	−32
15	Normal-Hearing Speaker ID	FNH5	FNH3	MNH2	FNH7	FNH6	FNH4	MNH1
16	Noise level 1, 7; range (dB)	23, 95; 72	17, 95; 78	23, 95; 72	23, 95; 72	23, 95; 72	15, 95; 80	21, 95; 74

[a]Data could not be obtained.

tory by the Massachusetts Eye and Ear Infirmary or the University of Massachusetts Memorial Medical Center and were paid for their participation. Table I presents pertinent characteristics of these individuals. Rows 10 and 11 of the table show that the implant users demonstrated substantial gains in vowel and consonant perception between measures made preimplant and at 1 year post implant.

A control group consisted of five female and two male paid volunteers, mean age 42, range 24–58, with no reported difficulties with speech or hearing. The control group initially consisted of four male and five female participants. Those over age 40 took a screening test, in order to determine approximate thresholds at 0.5, 1, 2, and 4 kHz. After a practice tone at 50 dB HL, sound pressure was increased in 5 dB increments from 0 to 25 dB HL. The series was presented twice at each of the four frequencies to each ear. Subjects who failed to report hearing the tone in any of the 16 series (2 ears, 4 frequencies, 2 trials) were excluded, which resulted in the exclusion of two males and left seven subjects in the control group. (On a test of phoneme recognition the remaining hearing participants scored at least 95% correct in vowel recognition and 90% correct in consonant recognition.) Consequently, the hearing controls are mainly female, while the implant users are mainly male. As explained below, in order to average data across subjects regardless of gender, and to compare patterns of averaged data between groups, the parameters of each speaker's vowel productions were converted to standard scores before averaging.

B. Procedures

1. Effects of masking noise

Two kinds of phoneme contrasts were examined, vowel and sibilant. The vowels /i/, /u/, /ɛ/ and /æ/ were elicited in the words *peet, poot, pet, and pat*, in the carrier phrase, "It's a _ please." The sibilants /s/ and /ʃ/ were elicited in the words *sot* and *shot*, in the carrier phrase "Say _ please." (The carrier phrase was changed for the sibilants to avoid subjects' tendency to elide the "a" in "It's a _.")

Implant users were recorded in two sessions following activation of the speech processors of their cochlear implants, at approximately 1 month postactivation and 1 year postactivation. For the normal-hearing controls, a single time sample was recorded. For all subjects, ten tokens of each of the four /pVt/ and the two /Sɑt/ ("S" = sibilant) words were elicited in random order within each of eight noise to signal levels. One N/S level had ambient noise only, which we call the "quiet" condition; the other seven had noise added at levels ranging between each subject's approximate thresholds for detection of the noise and for discomfort (described below). The presentations were blocked according to increasing levels of N/S.[2]

Participants were seated in a single-walled sound-attenuating booth (Eckel Industries) in a comfortable chair. A head-mounted electret microphone (Audio-Technica, model AT803B) was placed at a fixed distance of 20 cm from the participant's lips. The microphone was connected to a custom-built "feedback controller" (Technical Collaborative,

Lexington MA), which mixed the subject's speech signal with specific calibrated levels of noise under computer control. The noise was approximately speech-shaped, with a spectral envelope that rolled off at 6 dB per octave.

Both subject groups received auditory feedback from the output of the feedback controller. For the implant users, the controller's output was connected to a laboratory speech processor, which was loaded with the settings currently in use on the subject's prosthesis. The implant users were allowed to adjust the overall gain to a level that felt "normal" to them before the recording commenced. For the normal-hearing participants, the device was connected to calibrated TDH-39 headphones with ear cushions and the volume set at a comfortable level for the subject.

The following procedure was used to establish an approximate "dynamic range" for each subject. In each session, for most of the subjects, the lowest noise level was determined by gradually increasing the noise level generated by the device until it was just detectable by the subject. The highest level was set to be just below that considered by the subject to be uncomfortable (see lowest and highest levels in Table I, rows 12, 13, and 16).[3] The same upper limit was used for all normal-hearing subjects, 95 dB SPL. For them as for the implant users, the five intermediate noise levels added were set at equal increments of SPL between the lowest and highest levels.

To provide a reference sound level for calculation of the sound pressure level of the subjects' produced vowels, a calibration signal was recorded while the subject remained silent. The signal was generated by an electrolarynx (Cooper-Rand Sound Source; Luminaud, Inc.; Mentor, OH) placed in front of the speaker's lips while an experimenter observed the sound pressure level on a sound level meter (C scale) placed next to the microphone.

For the recording, the subject's speech and the output of the feedback controller (containing the subject's speech mixed with noise, called the "mixed signal") were low-pass filtered at 7.2 kHz. The resulting signals were digitized directly to computer disk, each at a 16 kHz sampling rate.

2. Loudness-target control experiment

The changes in the speaker's contrast distance due to clarity and masking effects are potentially confounded with any changes in contrast distance due to simply speaking louder. In order to control for the effects on contrast distance of speaking louder (cf. Pickett, 1956), most of the participants served in a control experiment in which they read the same words they had read in the masking experiment, and contrast distance was measured without masking noise. Each speaker was asked to reproduce four of the seven speech sound levels he or she had produced in the masking experiment (called "target levels").[4] There were 10 repetitions of each of the utterances at each of the four target levels. To guide the speaker in this reproduction, the recording software generated a real-time visual display of the subject's sound level in the form of a moving bar graph, with a 4 dB wide target region. The centers of the target levels displayed were derived from the subject's productions in the masking-noise experiment at noise-added levels 1, 3, 5, and 7. All the subjects were readily able to follow the instruction to keep the bar approximately within the target region. Aside from these conditions, the recording and data extraction procedures were the same as in the masking experiment.

C. Data extraction

1. Vowels

Working with a display of the digitized speech signal of each utterance, an experimenter placed markers at the start and end of the vowel in each /pVt/ token. The parameters F1 and F2 were extracted algorithmically (while monitored and corrected, if necessary, by the experimenter) from an LPC spectrum around mid-vowel. A 40 ms analysis window for F0 and a 25 ms window for the formants were used. The LPC filter order was chosen to optimize formant delineation for each subject. (For further details of procedures for formant extraction, see Lane *et al.*, 2005.) Vowel duration was calculated from the labeled start and end times, and SPL was calculated from the RMS over the entire vowel duration as a log ratio with the RMS of the calibration signal. Values in mels for each formant, M1 and M2, were calculated from the formula

$$M = 2595 \times L_{10}(1 + (F/700)).$$

An overall measure of vowel contrast, average vowel spacing (AVS), was calculated as the mean Euclidean distance separating members of all possible pairs of vowels in the $M1 \times M2$ space for each repetition, averaged across repetitions (Lane *et al.*, 2001).

To provide a basis for calculating the N/S for each subject's utterances, the SPL of the noise at each of the seven noise levels presented to that subject was derived from the RMS of a portion of the mixed signal recorded when the subject was not speaking (i.e., containing only the noise). Then, the N/S of each token was calculated as the difference (in dB) between the SPL of the noise and the SPL of that token's recorded speech signal.

2. Sibilants

The start and end of the sibilant frication noise were determined by visual inspection of the waveform and the spectrogram, and the spectral mean was extracted algorithmically at the mid-point of each sibilant.[5] A measure of sibilant contrast distance was calculated as the difference between the spectral means of /s/ and /ʃ/ for each repetition, averaged across repetitions (Matthies *et al.*, 1994; Perkell *et al.*, 2004).

III. RESULTS

A. Vowels

1. Normal-hearing speakers

Figure 2 shows plots of average values of the parameters AVS (labeled A, in mels), duration (D, ms) and SPL (S, dB) as a function of average (N/S, dB) obtained at each of the seven noise levels. The error bars show one standard error about the mean. For plotting purposes, the value of N/S for the quiet condition was arbitrarily set to be less than the

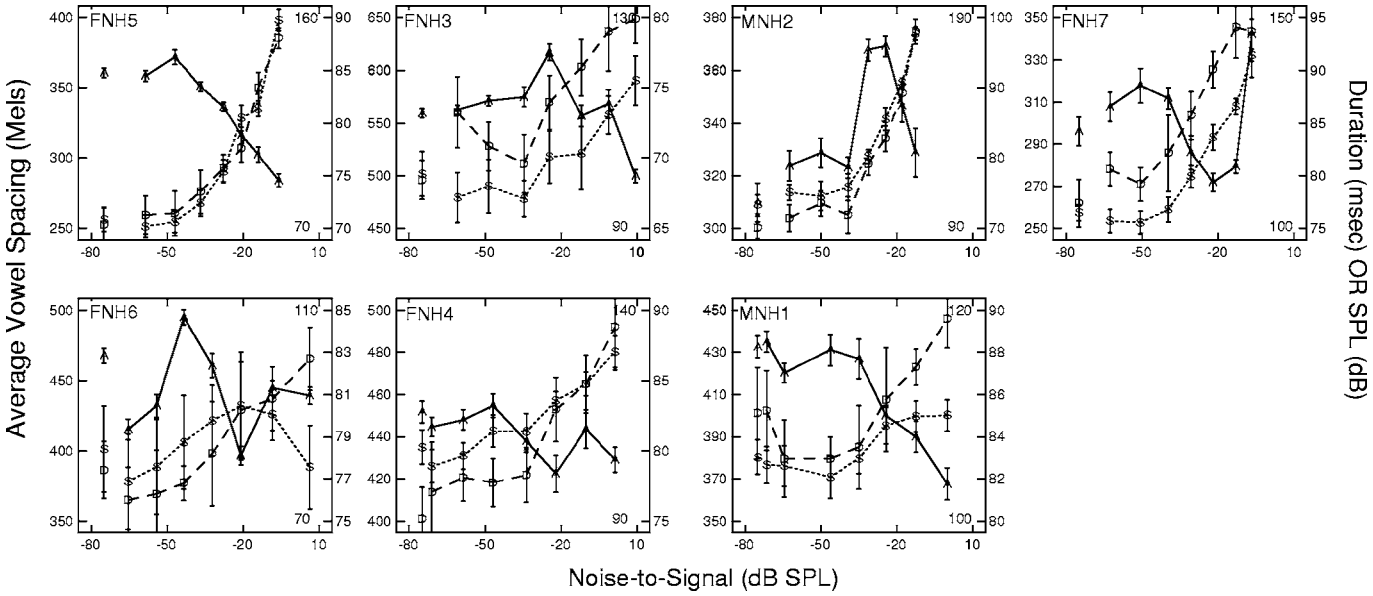


FIG. 2. AVS (A-mels), Duration (D-ms) and SPL (S-dB) vs. noise-to-signal ratio (dB) for each of the seven speakers who had normal hearing. For each panel, values of AVS are on the left vertical axis; the range of durations is shown by numbers in the upper and lower right-hand corners; values of SPL are on the right vertical axis. Values of N/S were obtained by multiplying extracted S/N by −1 and binning the resulting values into seven equally-spaced intervals. The values on the x-axis are located at the centers of the class intervals. Results from the quiet conditions are shown separately at the left of each plot (N/S is arbitrary.) Subject designations: H=hearing; M/F=male/female. NH=normal hearing. Error bars: standard error about the mean.

lowest value of N/S (from noise level 1). In each panel, values of AVS are displayed along the left vertical axis (mels); the range of durations is indicated by numbers in the upper and lower right-hand corners (ms), and values of SPL are displayed along the right vertical axis (dB). The first set of values in each panel is from the quiet condition and the remaining values, connected by lines, are from the seven N/S levels with added noise. Each panel shows data from a control participant with normal hearing. Note that the range of N/S is the same in all the panels, −80 to +10 dB. In order to show values for individual subjects and make the shapes of the functions as observable as possible, the scales of AVS, duration, and SPL vary across panels; they have been set so that the functions fill out the vertical space (which results in varying aspect ratios).

For speakers with normal hearing, the dynamic ranges displayed—the N/S range for each subject from his or her noise detection to noise discomfort levels—are roughly similar to one another, occupying most of the horizontal range in each panel. With a few exceptions and some variation among the subjects, the plots of each of the three dependent variables show similar trends across the subjects. SPL and duration grow approximately monotonically as N/S increases. The shapes of the AVS functions are more irregular; however most of them show an increase followed by a decrease. The most obvious exception to this pattern is AVS for subject FNH7 (top row, fourth panel), which increases substantially at the highest N/S level. Close examination of this subject's vowel spectra and listening to the utterances produced at the highest N/S level revealed that there was no discernable value of F2 for /u/ within its expected range (based on normative data from the literature) and the vowel sounded fronted, unlike an American English /u/. As a result, the algorithmically detected values of F2 were in the normal region of F3 for /u/, which caused inflated values of AVS. Subject FNH7 exhibited the same behavior in the loudness-target control experiment, so her data are not included in the plots or statistics of vowel data averaged across subjects (see below).

In order to examine group trends, Fig. 3 shows the data from six of the speakers with normal hearing (excluding FNH7) averaged across subjects. As noted above in Sec. II A, because subjects differed in the ranges of variables measured, values of each subject's dependent variables and N/S were rescaled by converting them to standard scores before averaging. Except for the quiet condition without

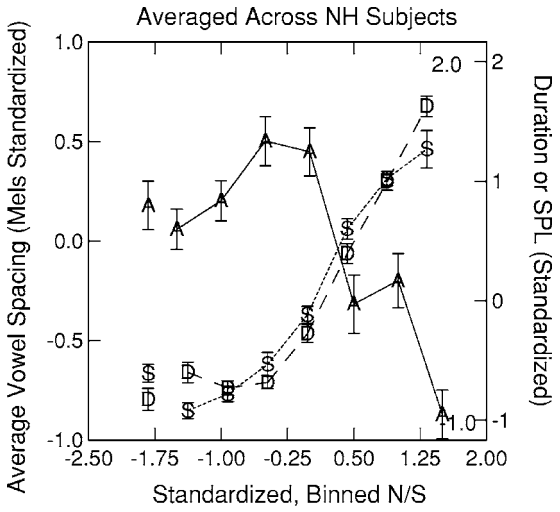


FIG. 3. Data averaged across six speakers with normal hearing (values in standard deviations). To compensate for inter-subject differences in data averages and ranges, each subject's data were standardized and the standardized values were averaged across subjects. See caption for Fig. 2 for remaining details. Data for Subject FNH7 were not included.

TABLE II. Summary of ANOVA results. $\eta^2 \times 100$=percentage of variance accounted for by the effect. η^2 (eta-squared) is calculated from [F=($\eta^2/1-\eta^2$)×(df error/df means)] (Young, 1993). NL: Noise level.

Row	Sample	Variable	Source	F	DF	p	$\eta^2 \times 100$
1	NH	AVSML	Subj	473	5, 33	<0.001	98
2			Noise level	19.5	6, 198	<0.001	37
3			NL×Subj	6.0	30, 198	<0.001	48
4			NL3>NL1	9.9	6, 33	<0.001	64
5			NL7<NL1	22.8	6, 33	<0.001	81
6		SPL	Subj	1130	5, 33	<0.001	99
7			NL	476	6, 198	<0.001	94
8			NL×Subj	92	30, 198	<0.001	93
9		DUR	Subj	128	5, 33	<0.001	95
10			NL	221	6, 198	<0.001	87
11			NL×Subj	23	30, 198	<0.001	78
12	CI—month	AVSML	Subj	156	6, 47	<0.001	95
13			Noise level	3.6	6, 282	<0.001	7
14			NL×Subj	1.6	36, 282	<0.01	17
15			NL7<NL1	3.6	7, 47	<0.001	34
16		SPL	Subj	251	6, 47	<0.001	96
17			NL	27.7	6, 282	<0.001	37
18			NL×Subj	4.9	36, 282	<0.001	38
19		DUR	Subj	328	6, 47	<0.001	97
20			NL	19	6, 282	<0.001	28
21			NL×Subj	13	36, 282	<0.001	62
22	CI—1 year	AVSML	Subj	410	6, 51	<0.001	98
23			Noise level	10	6, 306	<0.001	17
24			NL×Subj	3.5	36, 306	<0.001	30
25			NL4>NL1	3.8	7, 51	<0.05	34
26			NL7<NL1	16	7, 51	<0.001	69
27		SPL	Subj	1719	6, 51	<0.001	99
28			NL	176	6, 306	<0.001	78
29			NL×Subj	19	36, 306	<0.001	69
30		DUR	Subj	2146	6, 51	<0.001	99
31			NL	7.5	6, 306	<0.001	13
32			NL×Subj	8.0	36, 306	<0.001	48

added noise, the values of N/S were binned into seven equally spaced intervals, corresponding to the seven presented noise levels. The averaged standardized values of the dependent variables were plotted as a function of the average standardized value of noise level presented, expressed as the N/S class interval in which that noise level fell.[6] The centers of the seven N/S class intervals are shown on the abscissa in Fig. 3.

The resulting plot more clearly shows the trends observed in the individual data. SPL and duration grow monotonically as N/S increases. AVS increases then drops with increasing N/S. The local minimum at N/S=0.50 is due to downward fluctuations at the fifth point in the individual curves for FNH6, FNH3, and FNH4 (Fig. 2).

To test for the significance of observations made from this and subsequent plots of values averaged across subjects, two-way repeated-measures ANOVAs (with subjects as a category variable) and selected post hoc comparisons were computed on the original (unstandardized) data, separately for the normal-hearing speakers and for the implant users at 1 month and at 1 year. The ANOVAs were calculated on the unstandardized data to avoid possible violations of the assumptions underlying analysis of variance. The ANOVA results are presented in Table II and are indexed in the following text by the corresponding row numbers. The column labeled "$\eta^2 \times 100$" gives the percentage of variance accounted for by the effect. For the normal-hearing speakers and each of the three dependent variables, the effects of subject, noise level and their interaction were significant at $p < 0.001$ (AVS: Table II, rows 1–3, SPL: rows 6–8; Duration: rows 9–11). Values of $\eta^2 \times 100$ show that most of the effects accounted for large amounts of the variance, except for the effect of noise level and the interaction of subject with noise level on AVS (rows 2 and 3), possibly due to the irregular shapes of the individual AVS functions seen in Fig. 2.

Post hoc comparisons showed that the highest value of AVS (at the third noise level) was reliably greater than at the first noise level (Table II, row 4); AVS at the seventh noise level was reliably less than at the first noise level (row 5).

2. Cochlear implant users

Figure 4 shows individual plots for the seven implant participants. This figure is like Fig. 2, except in this case, there are two panels for each subject. The upper one is from data collected at 1 month postimplant; the lower, from data

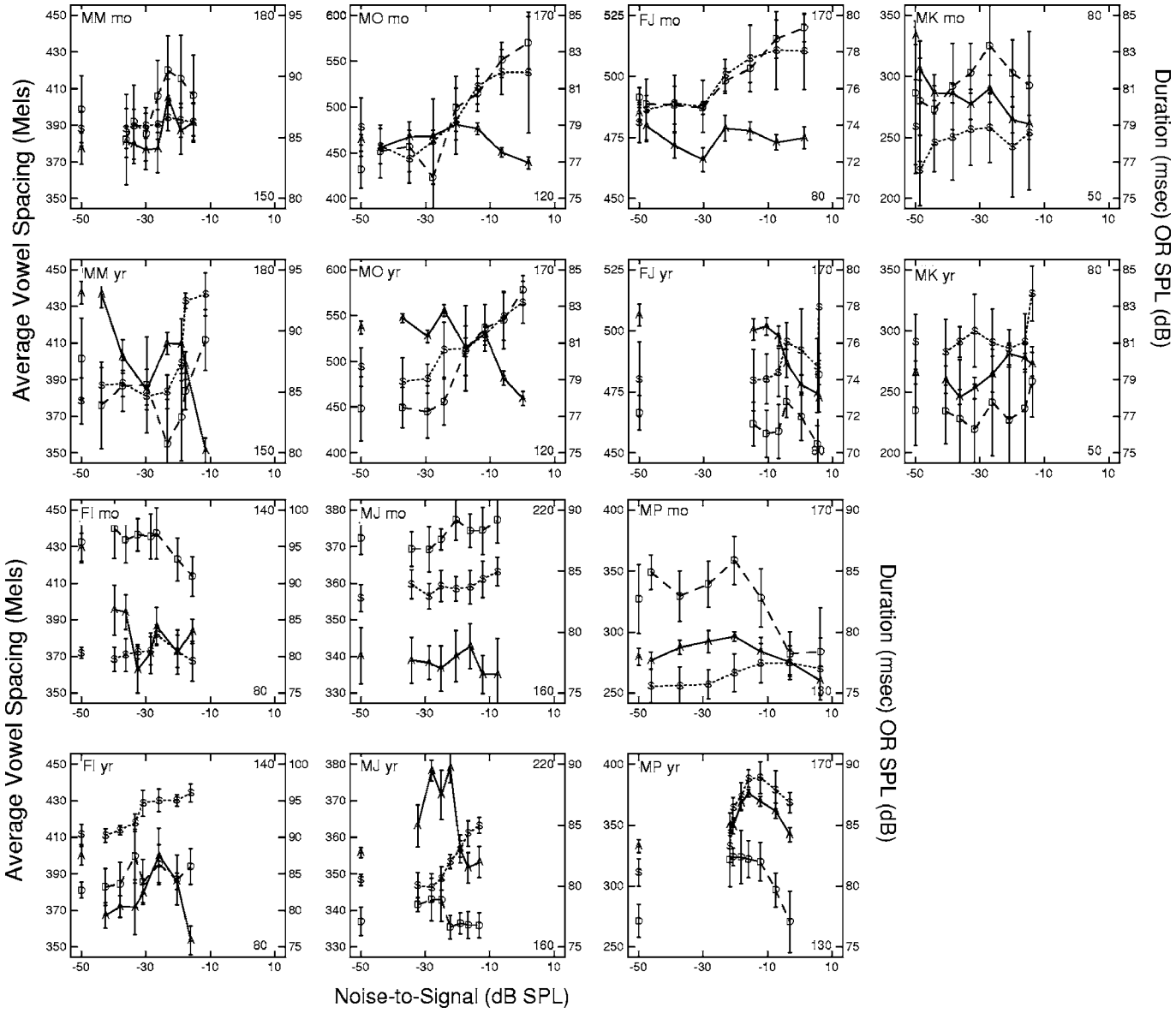


FIG. 4. AVS (A-mels), Duration (D-ms) and SPL (S-dB) vs. noise-to-signal ratio (N/S-dB) for each of the 7 implant users. There are two plots for each implant user, an upper one of data from one month and a lower one from one year post-implant. Subject designations: M/F=male/female. For further detail, see caption of Fig. 2.

collected at 1 year postimplant. The N/S range is the same in all the panels, −50 to 10 dB. The ranges of the dependent values are adjusted for each subject, but they are the same in each subject's 1 month and 1 year plots, to show changes between the two recording sessions.

The overall range of N/S displayed in the plots is 60 dB, 30 dB less than that for the speakers with normal hearing. This observation is consistent with the reduced dynamic range that implant users are known to have in general (cf. Hong *et al.*, 2003). For two of the implant users, FJ and MP, the dynamic range was considerably smaller at 1 year than 1 month (Table I, row 14) and the compression is in the direction of higher values. The opposite is true of subject MM, whose range is considerably greater at 1 year than 1 month (Table I, row 14). As can be seen from rows 8 and 9 in Table I, these changes cannot be accounted for by available information on implant processing strategies since MM and MP's strategies did not change between the two time samples.[7] Although there may be some overall regularity among the shapes of the functions for the dependent variables, the regularities are difficult to discern because of considerable difference among the subjects in the function shape (e.g., location of the inflection point) and sometimes large values of standard error about the mean.

Figure 5 shows data averaged across the seven implant users, at 1 month (panel A) and at 1 year (panel B) postimplant. As explained above, to combine individual speaker's data despite differences in the ranges of their variables, and to compare overall differences in levels of the dependent variables between 1 month and 1 year, before averaging over repetitions and speakers, all measures for each subject, including N/S, were converted to standard scores pooling across the two time samples. The functions in Fig. 5 show some similarities to those of the speakers with normal hearing in Fig. 3—more so at 1 year than at 1 month; however, there are also several differences.

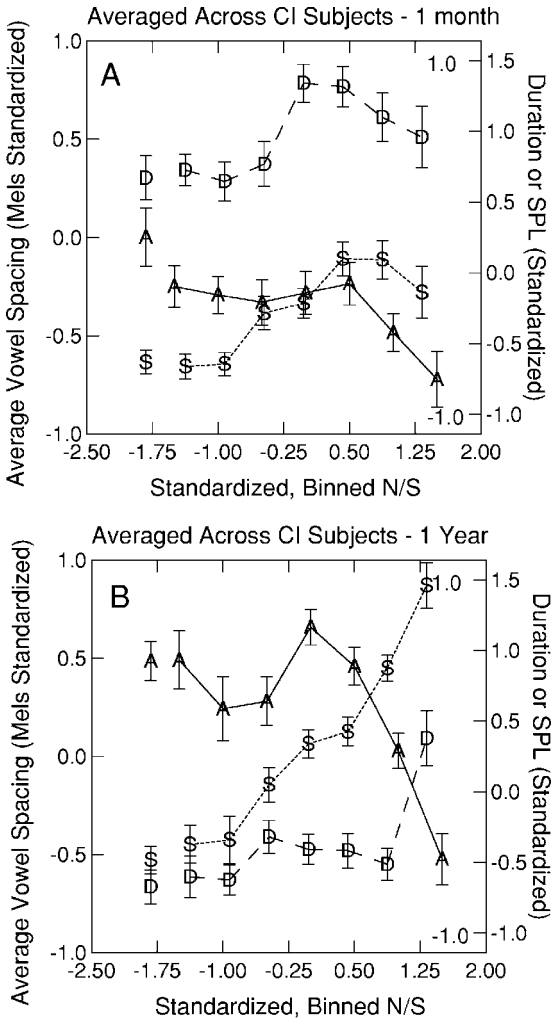


FIG. 5. Data averaged across the seven implant users, at one-month (panel A) and one-year (panel B) post-implant. See caption of Fig. 3 for remaining details.

a. AVS An ANOVA was computed on the unstandardized data for each time sample separately, with subject as a category variable and noise level as the treatment variable. The ANOVA results show main effects of subject and noise level on AVS at 1 month (Table II, rows 12 and 13) and at 1 year (rows 22 and 23) and a significant interaction between noise level and subject at both time samples (rows 18 and 24). In the 1 month data [Fig. 5(A)], AVS stays essentially flat with increased masking until it begins to drop at the sixth noise level. Effect sizes in Table II show that noise level accounts for 7% of the variance in AVS for the implant users at 1 month postimplant (row 13). At 1 year postimplant it accounts for more than twice as much of the variance (17%, row 23) but still less than for the controls (37%, row 2). AVS clearly begins at a higher mean standard score in the 1 year data [Fig. 5(B)] than in the 1 month data [Fig. 5(A)]. Post hoc comparisons showed that the highest AVS value in the 1 year data, at noise level 4, is slightly but reliably greater than its value at noise level 1 (row 25). AVS begins to fall at noise level 5 in the 1 year data [Fig. 5(B)] and reaches its lowest value at noise level 7, similar to the lowest value reached in the 1 month data (also at noise level 7). In both time samples, AVS is significantly lower at noise level 7 than at noise level 1 (rows 15 and 26).

b. SPL and duration As seen in Fig. 5, standardized durations are clearly shorter at 1 year than at 1 month postimplant, consistent with earlier observations on unstandardized measures (cf. Perkell *et al.* 1992). SPL grows with increasing N/S to a greater extent in the 1 year data than in the 1 month data. One-way repeated measures ANOVAs on the 1 month and 1 year data showed main effects of noise level on SPL and duration (one-month: rows 17, 20; one-year: rows 28, 31).

Turning to the subject variable and its interaction with noise level, there were significant effects of these two variables on SPL in both time samples (1 month: rows 16, 18; 1 year: rows 27, 29). SPL increased with noise level in both time samples, but the shape of the function resembles that of the normal-hearing speakers (Fig. 3) more in the 1 year data than in the 1 month data. There were also significant effects of subject and its interaction with noise level on duration in both time samples (1 month: rows 19, 21; 1 year: rows 30, 32).

B. Sibilants

Figure 6 shows plots of sibilant contrast distance (C), spectral mean for /s/ (S) and spectral mean for /ʃ/ (H) as a function of N/S (standardized) values. The N/S values are taken from the vowel data described above. The data are averaged across the seven control subjects [Fig. 6(A)] and the seven implant users at 1 month [6(B)] and 1 year [6(C)]. For the controls, contrast distance shows an initial nonsignificant rise from the quiet condition and then drops steadily as N/S increases. Observation of spectral means reveals that the decline in contrast distance is due mainly to an increase in the spectral mean for /ʃ/, beginning at the fourth noise level. The spectral mean for /s/ remains relatively flat across all N/S levels.

The data averaged across the implant users [Figs. 6(B) and 6(C)] are much more variable and show only one discernable trend: at 1 month postimplant [Fig. 6(B)], there is an initial increase in contrast distance followed by a decline over the first four N/S levels which appears to be due to a corresponding increase in spectral mean for /ʃ/. This contrast-distance decline in the 1 month data begins at a higher value than that found at any of the N/S levels in the 1 year data; at the end of the decline, it is in the same range as the values shown at 1 year across all N/S levels.

C. Control experiment

For each of the speakers who participated in the control experiment, whatever their hearing status, the sound levels they produced had no reliable effect on their vowel contrast or sibilant contrast. Between groups, implant users did not differ reliably from hearing speakers in vowel contrast ($F[1,88]=0.7, p>0.05$) nor in sibilant contrast ($F[1,83]=0.57, p>0.05$). There was no significant interaction between hearing status and the effects of produced sound level on vowel contrast ($F[3,264]=2.2, p>0.05$).

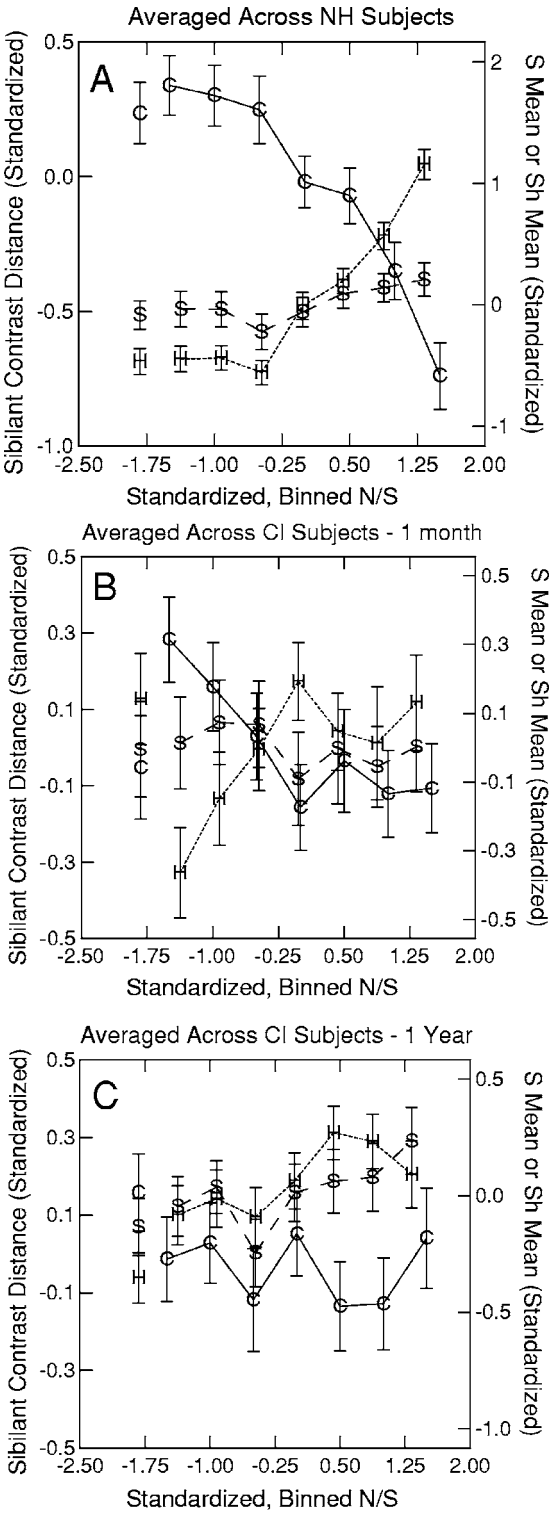


FIG. 6. Sibilant contrast distance (C), spectral mean for /s/ (S) and spectral mean for /ʃ/ (H) vs N/S (standardized values, as for Figs. 3 and 5). Results for speakers with normal hearing are shown in panel A, for implant users at one-month in panel B and for one-year post-implant in panel C. (For further details, see captions of Figs. 3 and 5.)

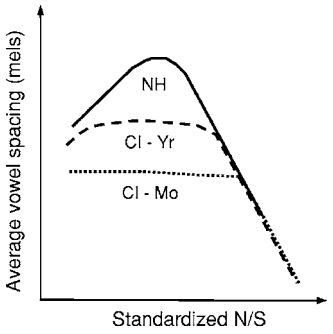


FIG. 7. Schematic diagram illustrating a possible explanation of the results. Average vowel spacing vs. normalized N/S. The solid function (NH) shows the result for the speakers with normal hearing; the dashed function (Cl-Yr), for the implant users at one year post-implant, and the dotted line (CI-Mo), for the implant users at one month.

IV. DISCUSSION

A. Summary and interpretation of results

1. Vowels

Within each of the three experimental conditions (control, and implant users at 1 month, and 1 year post implant) there was considerable variation from one speaker to the next in the relations between the dependent variables and N/S. Among the implant users, possible sources of variance include processor strategy, insertion depth, and demographic factors identified in Table I, such as age at profound hearing loss. Furthermore, speakers with normal hearing differ in their understanding of how extensively they should adapt their speech when they are subjected to increased noise (Lane and Tranel, 1971). In spite of this between-speaker variation, when the results were averaged across subjects, they demonstrated support for the hypotheses. This support is most evident in the functions for AVS, SPL, and duration in the normal-hearing speakers (Fig. 3). Over the entire range of increasing N/S, vowel SPL and durations increase and at the lower N/S levels AVS begins to increase (Hypothesis 1). At intermediate noise levels, AVS begins to fall, while the other two parameters continue increasing (Hypothesis 2). At higher N/S levels, AVS drops to a level below the condition with no added noise. Thus we infer that as long as speakers can perceive the degree of vowel contrast in the presence of masking noise, they will attempt to increase it. The more masking interferes with the perception of vowel contrast, the more contrast drops. The results of the loudness-target control experiment—no systematic relation between AVS and produced sound level—indicate that the observed relations between vowel contrast and N/S are not simply due to speaking louder.

Hypothesis 3 predicted that the function would inflect at a higher N/S level for the speakers with normal hearing than for the implant users and Hypothesis 4 predicted that the inflection point would be at a higher N/S level for the implant users at one year than at one month. However, observation of Figs. 3 and 5 reveals a tendency toward the reverse ordering. The inflection point on the function labeled "A" is at the third or fourth noise level for the normal-hearing speakers, at the fourth level for the implant users at 1 year and at the fifth for the implant users at 1 month. A possible explanation for this preliminary observation, and for the observed differences in overall level of AVS among the three data sets, may be developed from further consideration of principles incorporated into the DIVA model as described in the introduction, along with the schematic diagram shown in Fig. 7.

The figure shows three hypothetical AVS functions: for normal-hearing speakers, (NH, solid line), for implant users at 1 year postimplant (CI-Yr, dashed line), and for implant users at 1 month postimplant (CI-Mo, dotted line). As schematized in the figure and observed above, the NH function rises and then drops. The CI 1 year function begins at a lower level of AVS than the NH function, and it only rises slightly before beginning to turn downward. The CI 1 month function begins at the lowest level and stays flat until it also turns downward. The observed tendency in ordering of the inflection points (third or fourth noise level for the normal-hearing speakers, fourth level for the implant users at 1 year, and fifth for the implant users at 1 month) is also approximated in the schematized functions in Fig. 7.

The introduction suggested that an inflected AVS function as exemplified by the solid (NH) curve in Fig. 7 could be the product of two underlying functions: (1) the speaker's increasing contrast distance as noise level increases and (2) a predominating influence of economy of effort as masking increasingly prevents the speaker from using auditory feedback to help achieve auditory goals. The contrasts produced by implant users at 1 month postimplant are presumably influenced by a long-term and gradual degradation of their feedforward commands and auditory goals while they were deaf. After implantation, a poorly tuned auditory feedback subsystem (which was still adjusting to the novel stimulus from the implant) could have made it impossible to increase vowel contrasts above current working levels. This would result in low levels of contrast despite increasing N/S as shown by the dotted (CI-Mo) function. After a year's experience with the implant, when the auditory feedback subsystem presumably had been tuned up and auditory goals and feedforward commands had been updated, the speakers operated at a higher level of contrast and were able to increase it somewhat under moderately adverse listening conditions, reflected in the dashed function (CI-Yr). In effect, a year's experience of refining auditory goals and retuning feedforward and feedback control subsystems could have raised the implant users' contrast "ceiling" to some extent, but not to the level of speakers with normal hearing.

Thus, all three data samples (normal-hearing speakers, implant users at 1 month and at 1 year) hypothetically may be characterized by a single function that incorporates the effects of clarity, economy of effort and masking, along with a superimposed clarity ceiling that presumably depends on the state of the speakers' feedback and feedforward control subsystems. These four factors taken together, then, would have established the observed ordering of inflection points, viz., at the third or fourth noise level for the normal-hearing speakers, at the fourth level for the implant users at one year, and at the fifth N/S level for the implant users at 1 month.

2. Sibilants

The speakers with normal hearing operated at a higher overall level of sibilant contrast (1134 Hz) than the implant users [669 Hz; ($F[1,198]=82$, $p<0.001$]. The functions in Fig. 6 show that sibilant contrast was more vulnerable to masking noise than the vowel contrast. Unlike AVS, sibilant contrast for the normal-hearing subjects showed no significant increase at lower values of N/S; it dropped continuously with increased levels of masking. Sibilant contrasts for the implant users were even more vulnerable to masking—showing large amounts of intersubject and within-subject variability and only a partially systematic relation to level of masking; a decline over the first four noise levels at 1 month postimplant. Considering that this decline is observed at 1 month, the lack of any systematic relation to masking at 1 year is somewhat puzzling. The sibilants' greater vulnerability to noise compared to vowels is not surprising. The sibilants are differentiated from one another by characteristics of their noise spectra, which are much more likely to be masked by speech-shaped noise than the higher amplitude spectral peaks of vowels that are being excited by a voicing source.

For the most part, the unstandardized contrast values in the quiet condition for the individual implant users (not shown) displayed generally good sibilant contrasts, although this relation in the average plots of standardized values relating sibilant contrast to N/S is inverted at 1 month [Fig. 6(B)] and is not significant at 1 month or at 1 year [Fig. 6(C)].

The finding that the decrease in sibilant contrast distance is due mainly to an increase in the spectral mean for /ʃ/ while that for /s/ remains flat [Figs. 6(A) and 6(B)] is compatible with earlier results. Matthies *et al.* (1996) describe a gradual increase in spectral median for /ʃ/ in an implant user during a period of about 17 min while his implant was turned off. The increase in spectral median was accompanied by a gradual movement of the tongue blade forward, which would diminish the size and raise the resonant frequency of the cavity anterior to the constriction. When the implant was turned on again, making it possible again for the subject to hear the consequences of his articulations, on the next utterance the subject had brought his tongue blade back to its normal place of articulation, returning the spectral median to its normal value.

Perkell *et al.* (2004) investigated the sibilant contrast further with measures of 19 normal-hearing speakers' produced spectral contrast distance, their use of contact of the tongue tip with the lower alveolar ridge (hypothetically a somatosensory goal), and their auditory acuity. The results showed that speakers who had higher auditory acuity for the sibilant contrast, and who used tongue contact for /s/ but not /ʃ/, produced the greatest contrast distances; those who evidenced moderate acuity or s-contact produced intermediate contrast distances; and those who showed neither high acuity nor s-contact produced the smallest contrast distances. These findings were interpreted as support for the hypothesis that articulatory goals are auditory and somatosensory: in this case, the goals would be a higher spectral center of gravity and tongue contact with the lower alveolar ridge for /s/ but not /ʃ/. The use of contact to stabilize the production of /s/ is one of a number of examples of saturation effects that stabilize the production of virtually any consonant (Perkell *et al.*, 2000, 2004). Thus the spectral mean for /s/ is relatively unaffected by increased masking, because speakers are able to rely on the saturation effect for that sound in order to maintain the articulation in a way that is not affected by the loss

of auditory feedback. On the contrary, masking affects /ʃ/ presumably because speakers have to rely more on its auditory goal.

If speakers use an articulatory saturation effect for /s/ as described, the sibilant contrast would not be influenced by a tradeoff between clarity and economy of effort as much as the vowel contrasts would be. Were a speaker to change the amount of pressure between the tongue tip and lower alveolar ridge, that change in "effort" would presumably not be reflected in any change in the sibilant acoustics. Vowels, with their predominantly auditory/acoustic targets, generally are characterized by more continuous relations between changes in articulation and acoustics (except for place of articulation for /i/, /ɑ/ and /u/, Stevens, 1989, 1998; and degree of constriction for /i/, Perkell *et al.*, 2000; Perkell and Nelson, 1985).

B. Relations to other findings and conclusions

These findings are compatible with another experiment (Lane *et al.*, in press) conducted in our laboratory that employed the same implant users (plus an additional subject) and a group of normal-hearing subjects that included some of the same speakers as in this study. In the experiment reported by Lane *et al.*, AVS of eight vowels (as opposed to the current four) and sibilant contrast distance were examined preimplant and postimplant at two time samples—1 month and 1 year—with the implant processor (i.e., auditory feedback) turned off and on. The normal-hearing subjects, for feedback off, were presented with 95 dB of masking noise and for feedback on, no masking. In each of the recorded speech samples under feedback-on and feedback-off conditions, measures of both kinds of contrast, vowel and sibilant, were lower with feedback off than with it on. Parallel results were found in the current study: contrasts produced at the highest noise level (feedback maximally masked) were lower than the contrasts produced with no added noise. Also, contrasts for the larger vowel set were lower for the implant users at each time sample than were the contrasts of the normal-hearing speakers, as in the results of the present study.

In the only prior study that examined several vowel parameters produced under different levels of masking noise, Van Summers *et al.* (1988) measured RMS amplitudes, durations, F0, spectral slope and F1 and F2 from pronunciations of the digits "zero" through "nine" (plus some "control" words) by two male subjects with normal hearing. Five repetitions were pronounced at each of four levels of noise: quiet (no masking noise), 80, 90, and 100 dB. The two subjects differed somewhat from one another, but in general, amplitudes and durations followed the same trends as in the current study—increasing at higher noise levels. There was also a tendency for F0 values and spectral slope to increase with increasing levels of noise. F1 and F2 were analyzed separately and showed different trends for the two speakers; however, when we calculated AVS from the mean values in Figs. 6 and 7 of Van Summers *et al.* (1988, p. 922), both subjects showed lower values of AVS with 100 dB of masking noise than in the quiet.

Many investigators have found (e.g., Lane, 1963; Hawkins and Stevens, 1950; Hirsh *et al.*, 1954) that the intelligibility of speech mixed with noise decreases with increasing N/S. Assuming that speakers have implicit knowledge of noise-induced decrements in intelligibility, such knowledge provides motivation to enhance sound contrasts as much as possible in the face of environmental noise. The current results support the idea that speakers will increase clarity as much as they can, until the level of environmental noise begins to interfere with the use of auditory feedback mechanisms. With increasing noise, speakers are less and less able to perceive their own sound contrasts, and a presumed influence of economy of effort becomes a more predominant factor, causing contrast decrements. Speakers with compromised hearing (implant users in this case) habitually operate at lower levels of contrast than those with normal hearing, and when confronted with environmental noise, are less able to maintain and enhance contrasts.

ACKNOWLEDGMENTS

This research was supported by Grant No. R01-DC003007 from the National Institute on Deafness and Other Communication Disorders, National Institutes of Health. We are grateful to Dr. Donald Eddington of the Massachusetts Eye and Ear Infimary and Dr. Daniel Lee of the University of Massachusetts Medical Center for referring the implant users to the study, and to the implant users for their devotion of considerable amounts of their time. We also thank Advanced Bionics, Inc., and the Nucleus Corporation for their generous donations of research implant processors and three reviewers for their helpful comments on the manuscript.

[1]The study by Van Summers *et al.* (1988) is the closest in design to the current study. However, it was of only two male speakers with normal hearing; it did not employ as many noise levels and did not use levels near a detection threshold; and although F1 and F2 values were reported, contrast distances were not. Their results are compared further with the current study in the Discussion.

[2]In the earliest recordings, at 1 month postimplant for implant users MM, FJ, and MJ, the noise levels were randomized. The paradigm was changed to make the task more straightforward for the subjects.

[3]For all subjects, the noise was presented initially at 70 dB SPL, so that they would know what to listen for. To establish the lower threshold for all the implant users and all but three of the normal-hearing subjects, the noise was then presented at 11 dB SPL and the level was increased in 2 dB increments until the subject first reported hearing the noise. (For the first three normal-hearing controls, FNH3, FNH4, and MNH1, the lower threshold was set arbitrarily at 23 dB SPL.) When the lower threshold had been established for the implant users, the noise was presented again at 70 dB SPL and the level was increased in 2 dB increments until the subject indicated that the noise was as loud as he or she would tolerate. These procedures were carried out once on each subject. The upper limit was set at 95 dB SPL for all the normal-hearing controls. Although this method established only an approximation of each subject's dynamic range, it was considered adequate for the purposes of the experiment.

[4]The control experiment took place at a later date. Two of the seven NH subjects (FNH5, FNH6) and one of the seven implant users (MM) were not available to participate.

[5]Spectral mean has been shown to provide a robust, meaningful acoustic measure of the contrast (Forrest *et al.*, 1988; Jongman, Wayland, and Wong, 2000; Matthies *et al.*, 1994, 1996). Matthies *et al.* (1994) showed that late-deafened adults with reduced sibilant contrasts had improved contrasts after 6 months of implant use; spectral median and spectral skew gave similar results.

[6]The subject's produced vowel sound level varied somewhat from token to token within each of the presented noise levels. For those tokens produced with more extreme sound level values at any given noise level, the calculated value of N/S fell into the bin for the next higher or lower presented noise level; therefore, the number of tokens underlying the averaged values shown in Figs. 3, 5, and 6 vary somewhat from the standardized N/S value (bin) to the next.

[7]Differences in an implant subject's range between the thresholds for detection and discomfort at 1 month and 1 year may be due to the subjects' having used different sensitivity settings (for what felt "comfortable" at the time), which were beyond the experimental control.

Black, J. (**1951**). "The effect of noise-induced temporary deafness upon vocal intensity," Speech Monographs **18**, 74–77.

Blamey, P. J., Dowell, R. C., Brown, A. M., Clark, G. M., and Seligman, P. M. (**1987**). "Vowel and consonant recognition of cochlear implant patients using formant-estimating speech processors," J. Acoust. Soc. Am. **82**, 48–57.

Bond, Z. S., Moore, T. J., and Gable, B. (**1989**). "Acoustic-phonetic characteristics of speech produced in noise and while wearing an oxygen mask," J. Acoust. Soc. Am. **85**, 907–912.

Chen, F. R. (**1980**). "Acoustic characteristics and intelligibility of clear and conversational speech at the segmental level," Masters thesis, Massachusetts Institute of Technology (unpublished).

Chen, F. R., Zue, V. W., Picheny, M. A., Durlach, N. I., and Braida, L. D. (**1983**). "Speaking clearly: Acoustic characteristics and intelligibility of stop consonants," Working Papers, Research Laboratory of Electronics, Massachusetts Institute of Technology, **2**, 1–8.

Draegert, G. L. (**1951**). "Relationships between voice variables and speech intelligibility in high level noise," Speech Monographs **18**, 272–278.

Dreher, J. J., and O'Neill, J. J. (**1958**). "Effects of ambient noise on speaker intelligibility of words and phrases," Laryngoscope **68**, 539–548.

Ferguson, S. H. (**2004**). "Talker differences in clear and conversational speech: Vowel intelligibility for normal-hearing listeners," J. Acoust. Soc. Am. **116**, 2365–2373.

Ferguson, S. H., and Kewley-Port, D. (**2002**). "Vowel intelligibility in clear and conversational speech for normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **112**, 259–271.

Forrest, K., Weismer, G., Milenkovic, P., and Dougall, R. N. (**1988**). "Statistical analysis of word-initial voiceless obstruents: Preliminary data," J. Acoust. Soc. Am. **84**, 115–123.

Gagné, J. P., and Tye-Murray, N. (**1994**). *Research in Audiological Rehabilitation: Current Trends and Future Directions Monograph Supplement to Journal of the Academy of Rehabilitative Audiology* (American Academy of Rehabilitative Audiology, Cedar Falls), Vol. 27.

Guenther, F. H. (**1995**). "Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production," Psychol. Rev. **102**, 594–621.

Guenther, F. H., Ghosh, S. S., and Tourville, J. A. (**2006**). "Neural modeling and imaging of the cortical interactions underlying syllable production," Brain Lang **96**, 280–301.

Guenther, F. H., Hampson, M., and Johnson, D. (**1998**). "A theoretical investigation of reference frames for the planning of speech movements," Psychol. Rev. **105**, 611–633.

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.

Hazan, V., and Markham, D. (**2004**). "Acoustic-phonetic correlates of talker intelligibility for adults and children," J. Acoust. Soc. Am. **116**, 3108–3118.

Hawkins, J. E., Jr., and Stevens, S. S. (**1950**). "The masking of pure tones and speech by white noise," J. Acoust. Soc. Am. **22**, 6–13.

Hirsh, I. J., Reynolds, E. G., and Joseph, M. (**1954**). "Intelligibility of different speech materials," J. Acoust. Soc. Am. **26**, 530–538.

Hong, R. S., Rubinstein, J. T., Wehner, D., and Horn, D. (**2003**). "Dynamic range enhancement for cochlear implants," Otol. Neurotol. **4**, 590–595.

Jongman, A., Wayland, R., and Wong, S. (**2000**). "Acoustic characteristics of English fricatives," J. Acoust. Soc. Am. **108**, 1252–1263.

Kishon-Rabin, L., Taitelbaum, R., Tobin, Y., and Hildesheimer, M. (**1999**). "The effect of partially restored hearing on speech production of postlingually deafened adults with multichannel cochlear implants," J. Acoust. Soc. Am. **106**, 2843–2857.

Krause, J. C., and Braida, L. (**2003**). "Effects of listening environment on intelligibility of clear speech at normal speaking rates," Iranian Journal of Audiology **2**, 39–47.

Lane, H. (**1963**). "Foreign accent and speech distortion," J. Acoust. Soc. Am. **35**, 451–453.

Lane, H., Denny, M., Guenther, F. H., Matthies, M. L., Menard, L., Perkell, J. S., Stockmann, E., Tiede, M., Vick, J., and Zandipour, M. (**2005**). "Effects of bite blocks and hearing status on vowel production," J. Acoust. Soc. Am. **118**, 1636–1646.

Lane, H., Matthies, M. L., Denny, M., Guenther, F. H., Perkell, J. S., Burton, E., Tiede, M., Vick, J., and Zandipour, M. (submitted).

Lane, H., Matthies, M. L., Perkell, J. S., Vick, J., and Zandipour, M. (**2001**). "The effects of changes in hearing status in cochlear implant users on the acoustic vowel space and coarticulation," J. Speech Lang. Hear. Res. **44**, 552–563.

Lane, H., and Webster, J. (**1991**). "Speech deterioration in postlingually deafened adults," J. Acoust. Soc. Am. **89**, 859–866.

Lane, H., Wozniak, J., Matthies, M., Svirsky, M., Perkell, J., O'Connell, M., and Manzella, J. (**1997**). "Changes in sound pressure and fundamental frequency contours following changes in hearing status," J. Acoust. Soc. Am. **101**, 2244–2252.

Lane, H., and Tranel, B. (**1971**). "The Lombard sign and the role of hearing in speech," J. Speech Hear. Res. **14**, 677–709.

Langereis, M. C., Bosman, A. J., van Olphen, A. F., and Smoorenburg, G. F. (**1997**). "Changes in vowel quality in post-lingually deafened cochlear implant users," Audiology **36**, 279–297.

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, Dordrecht), pp. 403–439.

Liu, S., Del, R. E., Bradlow, A. R., and Zeng, F. G. (**2004**). "Clear speech perception in acoustic and electric hearing," J. Acoust. Soc. Am. **116**, 2374–2383.

Matthies, M. L., Svirsky, M. A., Lane, H. L., and Perkell, J. S. (**1994**). "A preliminary study of the effects of cochlear implants on the production of sibilants," J. Acoust. Soc. Am. **96**, 1367–1373.

Matthies, M. L., Svirsky, M. A., Perkell, J. S., and Lane, H. (**1996**). "Acoustic and articulatory measures of sibilant production with and without auditory feedback from a cochlear implant," J. Speech Hear. Res. **39**, pp. 936–946.

McKay, C. M., and McDermott, H. J. (**1993**). "Perceptual performance of subjects with cochlear implants using the Spectral Maxima Sound Processor (SMSP) and the Mini Speech Processor (MSP)," Ear Hear. **14**, 350–367.

Moon, S. J., and Lindblom, B. (**1989**). "Formant undershoot in clear and citation-form speech: A second progress report," Speech Transmission Laboratory, QPSR, 1/1989, 121–123.

Payton, K. L., Uchanski, R. M., and Braida, L. D. (**1994**). "Intelligibility of conversational and clear speech in noise and reverberation for listeners with normal and impaired hearing," J. Acoust. Soc. Am. **95**, 1581–1592.

Perkell, J. S., Matthies, M. L., Tiede, M., Lane, H., Zandipour, M., Marrone, N., Stockmann, E., and Guenther, F. H. (**2004**). "The distinctness of speakers' /s/-/ʃ/ contrast is related to their auditory discrimination and use of an articulatory saturation effect," J. Speech Lang. Hear. Res. **47**, 1259–1269.

Perkell, J. S., and Nelson, W. L. (**1985**). "Variability in production of the vowels /i/ and /a/," J. Acoust. Soc. Am. **77**, 1889–1895.

Perkell, J., Numa, W., Vick, J., Lane, H., Balkany, T., and Gould, J. (**2001**). "Language-specific, hearing-related changes in vowel spaces: A preliminary study of English- and Spanish-speaking cochlear implant users," Ear Hear. **22**, 461–470.

Perkell, J. S., Guenther, F. H., Lane, H., Matthies, M. L., Perrier, P., Vick, J., Wilhelms-Tricarico, R., and Zandipour, M. (**2000**). "A theory of speech motor control and supporting data from speakers with normal hearing and with profound hearing loss," Phonetica **28**, 233–272.

Perkell, J. S., Lane, H., Svirsky, M., and Webster, J. (**1992**). "Speech of cochlear implant patients: A longitudinal study of vowel production," J. Acoust. Soc. Am. **91**, 2961–2979.

Perkell, J. S., Zandipour, M., Matthies, M. L., and Lane, H. (**2002**). "Economy of effort in different speaking conditions. I. A preliminary study of inter-subject differences and modeling issues," J. Acoust. Soc. Am. **112**, 1627–1651.

Peters, R. W. (**1955**). "The effect of filtering of sidetone on speaker intelligibility," J. Speech Hear Disord. **20**, 371–375.

Picheny, M. A., Durlach, N. I., and Braida, L. D. (**1985**). "Speaking clearly for the hard of hearing I: Intelligibility differences between clear and conversational speech," J. Speech Hear. Res. **28**, 96–103.

Picheny, M. A., Durlach, N. I., and Braida, 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.
Pickett, J. (**1956**). "Effects of vocal force on the intelligibility of speech sounds," J. Acoust. Soc. Am. **28**, 902–905.
Plant, G. (**1984**). "The effects of an acquired profound hearing loss on speech production," Br. J. Audiol. **18**, 39–48.
Smyth, V., Murdoch, B., McCormack, P., and Marshall, I. (**1991**). "Objective and subjective evaluation of subjects fitted with the cochlear multichannel prostheses: 3 studies," Australian Journal of Human Communication Disorders **19**, 31–52.
Stevens, K. N. (**1989**). "On the quantal nature of speech," J. Phonetics **17**, 3–46.
Stevens, K. N. (**1998**). *Acoustic Phonetics* (MIT Press, Cambridge, MA).
Tartter, V. C., Gomes, H., and Litwin, E. (**1993**). "Some acoustic effects of listening to noise on speech production," J. Acoust. Soc. Am. **94**, 2437–2440.
Van Summers, W., Pisoni, D. B., Bernacki, R. H., Pedlow, R. I., and Stokes, M. A. (**1988**). "Effect of noise on speech production: Acoustic and perceptual analyses," J. Acoust. Soc. Am. **84**, 917–928.
Waldstein, R. (**1990**). "Effects of postlingual deafness on speech production: Implications for the role of auditory feedback," J. Acoust. Soc. Am. **88**(5), 2099–2114.
Wilson, B., Lawson, D., Zerbi, M., Finley, C., and Wolford, R. (**1995**). "New processing strategies in cochlear implantation," Am. J. Otol. **16**, 669–681.
Young, M. A. (**1993**). "Supplementing tests of statistical significance: Variation accounted for," J. Speech Hear. Res. **36**, 644–656.

Sentence recognition in native- and foreign-language multi-talker background noise[a)]

Kristin J. Van Engen[b)] and Ann R. Bradlow
Department of Linguistics, Northwestern University, 2016 Sheridan Road, Evanston, Illinois 60208

(Received 18 April 2006; revised 25 October 2006; accepted 25 October 2006)

Studies of speech perception in various types of background noise have shown that noise with linguistic content affects listeners differently than nonlinguistic noise [e.g., Simpson, S. A., and Cooke, M. (2005). "Consonant identification in N-talker babble is a nonmonotonic function of N," J. Acoust. Soc. Am. **118**, 2775–2778; Sperry, J. L., Wiley, T. L., and Chial, M. R. (1997). "Word recognition performance in various background competitors," J. Am. Acad. Audiol. **8**, 71–80] but few studies of multi-talker babble have employed background babble in languages other than the target speech language. To determine whether the adverse effect of background speech is due to the linguistic content or to the acoustic characteristics of the speech masker, this study assessed speech-in-noise recognition when the language of the background noise was either the same or different from the language of the target speech. Replicating previous findings, results showed poorer English sentence recognition by native English listeners in six-talker babble than in two-talker babble, regardless of the language of the babble. In addition, our results showed that in two-talker babble, native English listeners were more adversely affected by English babble than by Mandarin Chinese babble. These findings demonstrate informational masking on sentence-in-noise recognition in the form of "linguistic interference." Whether this interference is at the lexical, sublexical, and/or prosodic levels of linguistic structure and whether it is modulated by the phonetic similarity between the target and noise languages remains to be determined. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2400666]

PACS number(s): 43.71.Es, 43.71.Hw, 43.72.Dv [PEI] Pages: 519–526

I. INTRODUCTION

The substantial literature on speech-in-noise perception has been successful in revealing the relative resistance of various speech signal features to degradation from noise, as well as in assessing the relative abilities of various listener populations to recover from the detrimental effects of background noise. A particularly noteworthy finding of several recent linguistic and audiological studies is that the presence of background noise can force a "re-ranking" of acoustic cues to linguistic categories such that "secondary" cues in quiet become the only available, and hence "primary," cues in noise (Parikh and Loizou, 2005; Jiang *et al.*, 2006). Mattys *et al.* (2005) also provide evidence that listeners assign different weights to various cues for word segmentation when the speech signal is fully available versus degraded by noise. Similarly, the presence of noise can "re-rank" listener groups such that groups that perform equivalently in quiet may perform differently in noise (Nábělek and Donohue, 1984; Takata and Nábělek, 1990; Mayo *et al.*, 1997; Van Wijngaarden *et al.*, 2002; but see Cutler *et al.*, 2004 for comparable effects of noise on native and non-native listener phoneme identification). Furthermore, noteworthy discrepancies between quiet and noisy test conditions have been observed for intelligibility of native- versus foreign-accented speech (Rogers *et al.*, 2004). These findings suggest that listeners process speech signals differently when they are embedded in noise as opposed to in quiet, and that a comprehensive understanding of speech perception requires studies of speech perception under various noise conditions.

Accordingly, the present study investigated English sentence perception in the presence of multi-talker babble with varying numbers of talkers in the babble, varying signal-to-noise ratios (SNRs), and varying languages in the background noise. Since speech perception in noise is likely to be affected by a combination of lower-level (peripheral, energetic) masking and higher-level (central cognitive, linguistic, informational) masking, particularly in the case of background speech babble noise, there is likely to be a range of noise characteristics (some SNRs, types of noise) where the linguistic content of the noise has a direct influence on the recognition of target speech. If we can identify this range of noise characteristics, then we can begin to isolate the various linguistic features (fine-grained acoustic phonetic segment-level to lexical and higher-level prosodic) that are involved in speech-in-speech perception. Our overall interest is in developing a deeper understanding of the linguistic factors involved in speech-in-noise recognition.

A key strategy for investigating the effects of noise on speech processing and for ultimately developing a principled account of these effects is to compare different types of noise, which vary with respect to the kind and degree of interference they impose on speech signals. To this end, lin-

[a)]A version of this study was presented at the 150th meeting of the Acoustical Society of America, Minneapolis, MN, October 2005 and at the Mid-Continental Workshop on Phonology 11, University of Michigan, November 2005.

[b)]Author to whom correspondence should be addressed. Electronic mail: k-van@northwestern.edu

guistic and audiological studies have employed a wide variety of noise types, including single-talker maskers, multi-talker babble with various numbers of talkers, speech-shaped noise, and white noise. In general, these studies have shown that, regardless of the type of noise, performance on speech recognition tasks decreases as the level of the noise increases relative to the level of the target speech. With respect to speech noise in particular, they have shown that greater similarity between masker and target voices in terms of characteristics such as vocal tract size and fundamental frequency decreases intelligibility (Brungart *et al.*, 2001). Target intelligibility also generally decreases as additional voices are added to multi-talker babble (Bronkhorst and Plomp, 1992; Bronkhorst, 2000; Brungart *et al.*, 2001; Rhebergen and Versfeld, 2005; Simpson and Cooke, 2005).[1] As for comparisons across speech and nonspeech noise, Simpson and Cooke (2005) found lower speech intelligibility scores in natural babble than in babble-modulated noise when there were more than two talkers in the noise (see Sperry *et al.*, 1997 for a similar result). This difference in the effects of natural babble and babble-modulated noise suggests that linguistic interference plays a role in the effects of natural multi-talker babble on target speech perception.

Although studies such as Simpson and Cooke (2005) have provided evidence for particularly linguistic effects of multi-talker noise, there has been little investigation of the factors involved in such effects. In multi-talker babble studies, for example, the language spoken in the babble has typically matched the language spoken in the target. As a consequence, we have limited information about the precise linguistic features—phonemes, words, prosodic characteristics—that are most responsible for the greater masking effects of speech noise than nonspeech noise.

Two recent studies, however, have used multiple noise languages in order to examine other aspects of noise and perception (Rhebergen *et al.*, 2005; Garcia Lecumberri and Cooke, 2006). Rhebergen *et al.* (2005) used two noise languages to examine the effects of time-reversing interfering speech. In general, target speech intelligibility is known to be better in time-reversed interfering speech than in forward interfering speech—an effect attributed to the removal of any interfering informational content in the noise. However, reversing speech also results in increased forward masking, which increases the energetic masking imposed by the noise on the target speech. Rhebergen *et al.* (2005) assessed the relative effects of these two opposing factors (reduced informational masking but increased energetic masking in time-reversed speech) by comparing the effects of forward and reversed Dutch babble with forward and reversed Swedish babble on the recognition of Dutch speech for Dutch listeners. A comparison of the Dutch-in-Dutch noise versus Dutch-in-Swedish noise (without time reversal) showed better speech reception thresholds in the Swedish noise condition.

Garcia Lecumberri and Cooke (2006) used two noise languages in a study of native and non-native listeners' perception of English consonants in noise. Their primary conclusion, based on a comparison of a variety of noise types, was that non-native listeners were more adversely affected than native listeners by both energetic and informational masking. A secondary finding, of direct relevance to the present study, was that the native English speakers performed slightly better in Spanish noise than in English noise. The present study expands on this finding with a more direct and systematic study of the effects of two noise languages on native listeners. Furthermore, by examining sentence intelligibility rather than consonant identification, this study involves more levels of linguistic knowledge and more closely represents real-world listening situations in which listeners must extract meaningful messages from noisy environments.

TABLE I. Aspects of native language versus foreign language two- and six-talker babble and their predicted effects on target speech intelligibility. "No" represents a feature of the noise that is expected to hinder target speech intelligibility. "Linguistic differentiation" refers broadly to differences in phonetic, phonological, lexical, and prosodic characteristics of the languages. It should be noted that the babble in this study was constructed from semantically anomalous sentences, so word transition probabilities and/or sentential semantics are not taken into account.

	Same language two-talkers	Different language two-talkers	Same language six-talkers	Different language six-talkers
Temporal gaps	Yes	Yes	No	No
Linguistic differentiation	No	Yes	No(?)	Yes

II. METHOD

This study compares the intelligibility of native-accented English sentences for native English listeners in the presence of English two- and six-talker babble versus Mandarin two- and six-talker babble at SNRs of +5, 0 and −5 dB.[2] Mandarin is particularly well suited to this investigation because it differs significantly from English with respect to several levels of linguistic structure—phoneme inventory, syllable structure, rhythmic properties, and prosodic properties. By comparing languages that differ dramatically, such as these, the chance of observing differential speech noise effects is maximized.

Table I lists aspects of two- and six-talker babble in the target speech language and in a different language that can be expected to affect target speech intelligibility for native speakers of the target language: (a) the amount/duration of temporal gaps in the noise and (b) the amount of linguistic differentiation between the target and the noise with respect to phonetic, phonological, lexical, and prosodic characteristics. Based on previous multi-talker babble findings, we expect that six-talker babble will be a more effective masker than two-talker babble.[3] With respect to linguistic characteristics, it is hypothesized that the degree to which linguistic information in the noise—individual phonemes, phonotactics, prosody, lexical items—matches the linguistic information in the target will correlate with the amount of interference caused by that noise type on the signal. For this reason, same-language noise should be more detrimental to target speech intelligibility than different-language noise.[4]

Listeners were presented with target sentences in English mixed with multi-talker babble in either English or Mandarin and were asked to write down what they heard.

TABLE II. Experimental design: conditions, block types, and block ordering.

	No. of Talkers in Noise	Block 1 Mandarin	Block 2 English	Block 3 Mandarin	Block 4 English
Condition 1	6	SNR: +5	SNR: +5	SNR: 0	SNR: 0
Condition 2	6	SNR: 0	SNR: 0	SNR: −5	SNR: −5
Condition 3	2	SNR: +5	SNR: +5	SNR: 0	SNR: 0
Condition 4	2	SNR: 0	SNR: 0	SNR: −5	SNR: −5

Four independent groups of listeners participated in four different experimental conditions, each one containing four blocks of trials, as shown in Table II. These conditions allowed for direct comparison of the effect of two versus six-talker babble (between subjects), the effect of English versus Mandarin babble (within subjects), and the effect of different signal-to-noise ratios (within subjects).

A. Participants

1. Speakers

Six monolingual native speakers of general American English (three males and three females between the ages of 28 and 48 years) and six native speakers of Mandarin Chinese (three males and three females between the ages of 24 and 37) provided recordings in their native languages to be used for the English and Mandarin noise tracks (described below). The English speakers were graduate students and postdoctoral researchers in the Northwestern University Linguistics Department, recorded for a previous experiment (Smiljanic and Bradlow, 2005). The Mandarin speakers were graduate students and family members of graduate students at Northwestern University. A different adult female speaker of general American English produced the target sentences.

2. Listeners

Seventy-three undergraduate participants were recruited from the Northwestern University Linguistics Department subject pool and received course credit for their participation in the study. Seven participants were omitted from the final analysis—two reported a hearing loss, two were non-native speakers of English, two were English-Mandarin bilinguals, and one was omitted due to computer error during data collection. The remaining 66 participants were native English speakers between the ages of 18 and 23, 18 of which were bilingual speakers of English and a language other than Mandarin. All reported having normal speech and hearing. The distribution of subjects across conditions was as follows: condition 1: $n=16$; condition 2: $n=17$; condition 3: $n=17$; condition 4: $n=16$.

B. Stimuli

1. Generating multi-talker babble

For the "noise" sentences, each speaker produced a set of 20 semantically anomalous sentences in either English (e.g., *Your tedious beacon lifted our cab; My puppy may stress their fundamental gallon*) or Mandarin. These English sentences were developed for unrelated research (Smiljanic and Bradlow, 2005), and were used in this study to eliminate the possibility that participants might extract an entire meaningful sentence from a speaker other than the target. The Mandarin sentences were direct translations of the English sentences (translated by one native Mandarin speaker and checked by another).

Participants were instructed to speak in a natural, conversational style, and to repeat any sentences in which they produced disfluencies. Recordings took place in a sound-attenuated booth in the phonetics laboratory of the Department of Linguistics at Northwestern University. Participants read the sentences from index cards and spoke into a microphone, recording directly to disk using an Apogee PSX-11 analog/digital and digital/analog converter. Recordings were digitized at a sampling rate of 16 kHz with 24 bit accuracy. Sentences were then separated into individual files and equated for rms amplitude so that they would all contribute equally to the babble.

English and Mandarin six-talker babble was created from these recordings as follows: for each talker, two sentences (a different pair of sentences for each talker) were concatenated to ensure the duration of the noise tracks would exceed the durations of all target sentences. A multiple of 100 ms of silence was added to each talker's file (0–500 ms) in order to stagger the talkers once they were mixed together. All six talkers were then mixed, and the initial 500 ms of the mixed file was removed to eliminate noise that did not contain all six talkers. The first 100 ms of the completed noise file was faded in, and the final noise file was leveled in rms amplitude to produce SNRs of +5, 0, and −5 dB when mixed with the stimulus sentences. The stimulus sentences were each leveled to the same rms amplitude (60 dB), and the relevant SNRs were produced by leveling each noise file relative to the level at which the sentence files had been leveled (55, 60, 65 dB).

For two-talker babble, two female voices were used for both English and Mandarin. This was done primarily to match the gender of the target speaker and thus eliminate the variable of gender differences in speech-in-speech intelligibility (see Brungart *et al.*, 2001). Furthermore, it was hoped that using the same gender for the two talkers would lead to better perceptual fusion of the pair so that it would be treated by listeners as two-talker babble. Four different two-talker noise tracks were generated for each language. Again, two sentences by each speaker were concatenated to ensure adequate duration of the noise file (four sentence pairs total per speaker). For one of the two speakers, 500 ms of silence was added to the beginning of the files. The two talkers were mixed as above and the first 500 ms were removed. Finally, the first 100 ms were faded in and the completed noise tracks were each leveled to the three rms amplitudes necessary to produce SNRs of −5, 5, and 0 dB when mixed with the leveled target sentences.

2. Target sentences

Target sentences for the present study were taken from a set of recordings originally made for an unrelated study (Bent and Bradlow, 2003). The sentences were taken from

the Revised Bamford-Kowal-Bench Standard Sentence Test, lists 7–10. Each list contains 16 simple, meaningful sentences (e.g., *The children dropped the bag; five men are working*) and 50 keywords (three or four per sentence) for a total of 64 sentences and 200 keywords. Lists 7, 8, 9, and 10 were selected based on their equivalent intelligibility scores for normal children as reported in Bamford and Wilson (1979). For additional details, see Bent and Bradlow (2003).

These recordings were mixed with the noise files, a 400 ms silent leader was inserted, followed by 500 ms of the noise alone, then the target signal mixed with the noise, and finally 500 ms of noise at the end of each trial. Each target sentence was mixed with each type of noise file, yielding 12 sets of target stimuli: two noise languages (English, Mandarin) X 2 talker numbers (two talker, six talker) X 3 SNRs (+5, 0, −5 dB). For six-talker noise, the same noise file was mixed with every target sentence at each SNR. For the two-talker noise, each of the four different tracks was used for 25% of the target sentences in each language and at each SNR.

C. Procedure

Listeners were seated in a sound-attenuated booth facing a computer monitor. Stimuli were presented diotically over headphones (Sennheiser HD 580) at a comfortable level. Participants were presented with a total of 68 trials—four practice sentences followed by four experimental blocks of 16 sentences each. They were instructed that they would be listening to sentences mixed with noise, and were asked to write down what they heard. They were asked to guess if they were unsure, and also to report individual words if that was all they could identify. The task was self-paced; participants pressed the space bar on the keyboard to advance from trial to trial. Participants could listen to each sentence no more than once. After the practice block, the experimenter verified that the equipment was functioning properly and checked the readability of the participant's handwriting.

Practice items (two sentences in English noise and two in Mandarin noise) were presented at the same SNR as block 1 (+5 dB for conditions 1 and 3, 0 dB for conditions 2 and 4). The sentences in block 1 were mixed with Mandarin noise; in block 2 the sentences were presented at the same SNR as block 1 but with English noise; in block 3 the sentences were presented at the more difficult SNR with Mandarin noise; and in the final block the sentences were presented at the more difficult SNR but with English noise (as shown in Table II).

English noise at the difficult SNR was predicted to be the most difficult block in all conditions due to the higher noise level and the greater linguistic overlap between the noise and the target. Therefore, this block was presented last, giving participants maximal opportunity to adjust to the task and to the target talker, thereby "stacking the cards" against our predicted result of better English sentence recognition with Mandarin noise than with English noise. Because of the possibility that some of the target sentences may be more or less easy to perceive than others, the four target sentence lists were counterbalanced across the four possible orderings. The ordering of blocks with respect to the type and level of noise was consistent for all participants (as described above).

D. Data analysis

Perception scores were determined by a strict keyword-correct count. Each set of 16 sentences contained 50 keywords, and listeners received credit for each keyword transcribed perfectly. Words with added or deleted morphemes were considered incorrect, but obvious spelling errors or homophones were counted as correct. Raw scores were converted to percent correct and then to rationalized arcsine units (RAU). This transformation "stretches" out the upper and lower ends of the scale, thereby allowing for valid comparisons of differences across the entire range of the scale (Studebaker, 1985). Scores on this scale range from −23 RAU (corresponding to 0% correct) to +123 RAU (corresponding to 100% correct).

III. RESULTS

As expected, higher SNRs yielded better target sentence perception in all conditions. Comparison across conditions also shows that sentence perception was better in two-talker noise than in six-talker noise as predicted by previous research (Brungart *et al.*, 2001; Rhebergen and Versfeld, 2005). With respect to the language of the noise, perception was significantly better in Mandarin than in English noise in two-talker babble at SNRs of 0 and −5 dB where those SNRs comprised the second half of the condition. The results for all experimental blocks and conditions are presented in Fig. 1.

Three-way repeated measures analysis of variance (ANOVAs) were performed separately for each condition, with language background (monolingual versus bilingual) as a between-subjects factor, and noise level (easy versus hard SNR) and noise language (English versus Mandarin) as within-subjects factors. There was no main effect of language background, nor any two- or three-way interactions with language background in any of the conditions; therefore language background was removed from all future analyses. This finding established that the mono- and bilingual participants performed equivalently in this study.

Two-way ANOVAs with noise level (easy versus hard) and noise language (English versus Mandarin) as within-subjects factors showed a significant main effect of level in all conditions (condition 1 $[F(1,15)=156.81,\ p<0.0001]$; condition 2 $[F(1,16)=976.09,\ p<0.0001]$; condition 3 $[F(1,16)=80.27,\ p<0.0001]$; condition 4 $[F(1,15)=103.02,\ p<0.0001]$). Condition 4 also showed a significant main effect of noise language $[F(1,15)=7.74,\ p=0.0140]$. Finally, both conditions 3 and 4 (the conditions that used two-talker babble) showed two-way interactions between noise level and noise language (condition 3 $[F(1,15)=6.151,\ p=0.0246]$; condition 4 $[F(1,15)=24.532,\ p=0.0002]$).

Post hoc pairwise comparisons (paired t tests) of the two-talker babble conditions showed a significant difference between English and Mandarin noise at the "hard" levels for

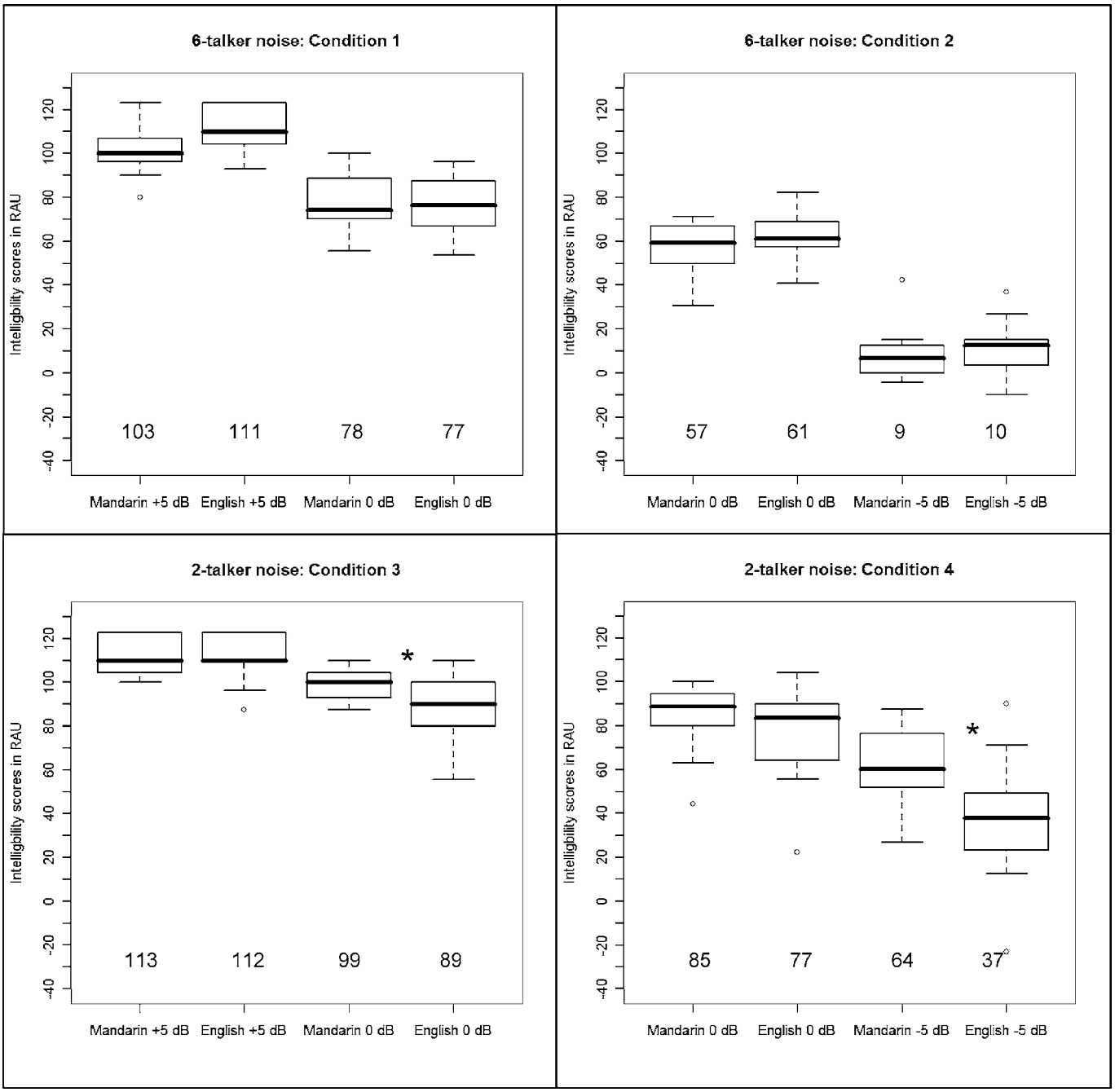


FIG. 1. Boxplots showing the interquartile ranges of intelligibility scores (in RAU) for conditions 1–4. Whiskers extend to the most extreme data point that is no more than 1.5 times the interquartile range of the box. Stars indicate significant differences between the two noise languages within a given noise condition. The mean is given at the bottom of each plot.

both condition 3 [$t(16)=2.438$, $p=0.0268$] and condition 4 [$t(15)=4.053$, $p=0.0010$], but no significant difference between the languages at the "easy" level in either condition.

Finally, three-way ANOVAs with noise level (easy vs hard) and noise language (English vs Mandarin) as within-subjects factors and number of noise talkers (two vs six) as a between-subjects factor were performed across conditions 1 and 3 (+5/0 SNRs) and conditions 2 and 4 (0/−5 SNRs). This analysis was included in order to compare the effects of two- vs six-talker noise. Both ANOVAs showed that intelligibility was significantly better in two-talker versus six-talker noise and at higher versus lower SNRs—findings which replicate patterns observed in previous studies (Bronkhorst and Plomp, 1992; Bronkhorst, 2000; Brungart *et al.*, 2001; Rhebergen and Versfeld, 2005).

For conditions 1 and 3, there was a main effect of number of talkers [$F(1,31)=24.65$, $p<0.0001$] and noise level [$F(1,31)=236.43$, $p<0.0001$]. As expected from the analyses on individual conditions, which showed the language effect only for "hard" noise levels and only in two-talker noise, all two-way interactions were significant: noise level and number of talkers [$F(1,31)=12.64$, $p=0.0012$], noise language and number of talkers [$F(1,31)=5.04$, $p=0.0320$] and noise level and noise language [$F(1,31)=6.05$, $p=0.0197$].

For conditions 2 and 4, main effects also emerged for number of talkers [$F(1,31)=61.43$, $p<0.0001$] and for noise level [$F(1,31)=579.40$, $p<0.0001$]. In addition, the effect of language approached significance [$F(1,31)=4.075$, $p=0.0522$]. All two-way interactions were also significant: noise level and number of talkers [$F(1,31)=34.37$, $p<0.0001$], noise language and number of talkers [$F(1,31)=10.23$, $p=0.0032$], and noise level and noise language [$F(1,31)=14.39$, $p=0.0006$]. Finally, there was a three-way interaction between noise level, noise language, and number of talkers [$F(1,31)=9.32$, $p=0.0046$]. Again, these interactions are predicted by the results of the analyses of individual conditions, which showed the effect of language to be significant only in two-talker noise at difficult SNRs. In summary, statistical analysis across two- and six-talker noise conditions indicates that sentence intelligibility is better with fewer talkers in the noise and with lower noise levels.

It should be noted that experience with the task and/or the target talker had an effect on participants' intelligibility scores. This is best illustrated by observing the results from the 0 SNR blocks. Where the first pair of blocks presented to participants was at an SNR of 0 (the "easy" SNR for the condition), they performed worse than when the same SNR was presented in the second half of the experiment (as the "hard" SNR). This difference due to ordering is observed for six-talker noise (compare conditions 1 and 2), as well as for two-talker noise (compare conditions 3 and 4). It is also true for both English and Mandarin noise, and in fact, the language effect only emerged as significant where 0 SNR was in the second half of the two-talker experiment. It is assumed that adjustment to the task and/or increased familiarity with the target voice accounts for such differences. Furthermore, these practice effects outweigh any effects of fatigue, which would cause decline in participant performance over the course of the experimental blocks.

IV. DISCUSSION

This study has shown that the language of interfering noise *can* affect the intelligibility of the target speech (for a similar result, see Rhebergen *et al.*, 2005; Garcia Lecumberri and Cooke, 2006). Specifically, where the noise contains two talkers and is presented at levels equal to or greater than the target speech, English noise is more detrimental than Mandarin noise to native English speakers who are listening to a native English target. This effect of noise language provides evidence that, under certain conditions, linguistic interference plays a role in the perception of speech in noise. The results of this study also replicate previous findings that the perceptibility of speech in multi-talker babble decreases as the number of talkers in the noise increases, as well as by an increase in the level of the noise with respect to the target (Bronkhorst and Plomp, 1992; Bronkhorst, 2000; Brungart *et al.*, 2001; Rhebergen and Versfeld, 2005; Simpson and Cooke, 2005). The overall results are summarized in Table III in the terms laid out in the introduction in Table I.

Before discussing the potential sources of the language effect in the two-talker condition, the lack of this effect in six-talker babble must be addressed. First, it should be noted that in post-experiment interviews, most participants in the six-talker babble conditions did not report noticing a shift in background noise language, and those that were aware of the presence of two noise languages did not report greater distraction from either one. In addition, whereas some subjects transcribed words from the babble in two-talker English conditions, this did not occur in six-talker conditions. The higher intelligibility scores in two- versus six-talker babble overall shows that the greater spectral and temporal density of six-talker noise, which yields greater energetic masking of the signal, made target intelligibility worse. At the same time, these characteristics of six-talker noise eliminated any informational masking differences between the two noise languages. Whatever "benefit" linguistic differentiation between the target and the noise may have provided in two-talker noise conditions was eliminated in six-talker noise. In sum, differential linguistic effects emerged only where the linguistic content of the noise was relatively available to the listener.

TABLE III. Aspects of native versus foreign-language two- and six-talker babble and their predicted effects on target speech intelligibility, presented with intelligibility scores averaged across participants and (matched) conditions.

	Same language two-talkers	Different language two-talkers	Same language six-talkers	Different language six-talkers
Temporal gaps	Yes	Yes	No	No
Linguistic differentiation	No	Yes	No (?)	Yes
Average words correct (RAU)	79	90	65	62

There are several possible sources of the greater masking by English noise than Mandarin noise for English target speech in two-talker babble: first, it is possible that differences in the long-term average spectra of the two languages contributed to the different effects, such that language-specific spectral similarities between the English noise and the English target increased the amount of energetic masking due to greater spectral overlap.

Running *t*-test analysis of the long-term average spectra of the English and Mandarin two-talker noise (averaged across the four tracks in each language) reveals statistically significant differences at several, but not all, frequencies. While these differences may contribute to the intelligibility asymmetry, the differences between the spectra are small and not consistent over the entire spectrum. The overall similarity in the long-term average spectra of the noise in the two languages suggests that spectral differences are not the sole source of the language effect. Furthermore, findings by Byrne *et al.* (1994) showed that, when averaged across talkers, languages do not differ significantly with respect to long-term average spectrum. A more likely explanation for the language effect observed here is that it is indeed an effect of different amounts of informational masking, i.e., a linguistic effect.

The precise aspect(s) of linguistic content that contribute to the language effect remain to be determined. The effect

may be primarily a whole-word lexical effect such that hearing and activating English words in the babble is what makes English noise more difficult to tune out than Mandarin noise. Participants frequently transcribed entire words from the English noise in their responses, indicating that interference occurred at this level. However, there may also be different amounts of interference from noise in different languages at sublexical and/or prosodic processing levels. Differences in the phoneme inventories and syllable structures of the languages, for example, may contribute to differential interference effects, with English phonemes and phonotactic patterns creating greater interference for the English listeners. Differences in rhythmic properties which correlate with syllable structure (Ramus *et al.*, 1999; Grabe and Low, 2002) and prosodic patterns may also contribute to the effect. It is likely that a combination of some or all of these whole-word, sublexical, and prosodic factors contribute to the different amounts of linguistic masking at play when the language of the noise matches or mismatches the language of the target speech. Future research must address their precise contributions to the noise language effect.

One method for conducting a fine-grained analysis of the locus of the linguistic masking reported in this study is to compare English noise to English nonword noise (composed of words that are phonologically legal in English but are not real words). This comparison should allow us to determine whether whole words from the background noise intrude on the target speech or whether sublexical properties can cause as much intrusion as lexical items. Preliminary findings from a study using nonword noise constructed by altering onsets, codas, or vowels in the content words of the original noise sentences showed no significant difference in intelligibility scores. This finding provides evidence against a strict lexical explanation for the language effect. However, it is likely that the high degree of similarity between the nonwords and real English words still causes listeners to activate items in the lexicon. The observation that participants in this preliminary experiment frequently transcribed real words that sounded like the nonwords present in the babble supports this interpretation. In future research, additional manipulations of noise content will be required to provide further insight into the question of the source of the language effect. These may include other types of nonwords, various accents of the target language, babble constructed from nonsentential materials (e.g., syllable strings, word lists), and noise from other languages.

V. SUMMARY/CONCLUSION

This study has shown that higher noise levels and greater numbers of talkers in multi-talker babble decrease target speech intelligibility, no matter what language is being spoken in the noise. However, in certain conditions (few talkers, difficult SNRs), linguistic effects appear to come into play, as shown by the differences observed between the effects of English and Mandarin two-talker babble: native English listeners performed better on a sentence intelligibility task in the presence of Mandarin two-talker babble than in English two-talker babble. We conclude that greater similarity between target and masker in the linguistic domain creates greater interference in target intelligibility, and must be taken into consideration in a principled account of speech perception in noise.

ACKNOWLEDGMENTS

This research was supported by Grant No. NIH-R01-DC005794 from NIH-NIDCD. The authors gratefully acknowledge the help of Ethan Cox.

[1]Simpson and Cooke (2005) showed that this relationship is nonmonotonic. This study varied the number of talkers (N) in both natural talker babble and babble-modulated noise (speech-shaped noise modulated by the envelope of N-talker babble) for a consonant identification task. In natural babble, intelligibility scores decreased with increasing numbers of talkers from N=1 to N=6, but were constant for N=6 to N=128. In babble-modulated noise, by contrast, intelligibility decreased gradually with increasing N values.

[2]The numbers of talkers and SNRs were chosen based on the authors' intuitions, the parameters used in previous speech-in-noise studies, and pilot data. It is recognized, however, that these are, to some extent, arbitrary decisions necessitated by limitations of the experimental design. It remains for future research to investigate a fuller range of talker numbers and SNRs.

[3]It is typically assumed that listeners are able to take advantage of temporal gaps in noise energy to hear relatively unobstructed "glimpses" of the target signal when there are fewer talkers in the noise (Bronkhorst, 2000; Brungart *et al.*, 2001; Assman and Summerfield, 2004; Rhebergen and Versfeld, 2005; Rhebergen *et al.*, 2005; Simpson and Cooke, 2005).

[4]The "no" under six-talker, same language noise is marked with a "?" in recognition of the fact that the presence of a greater number of talkers interferes with the availability of linguistic information in the babble due to temporal and spectral overlap of the various speakers in the babble. This babble-internal masking may decrease the linguistic interference that the matched-language noise would otherwise be predicted to impose on the target. However, it is possible that some isolated words and/or other linguistic characteristics will still be identifiable from the six-talker babble, and therefore we mark this cell with "no?" in the table.

Assman, P. F., and Summerfield, Q. (**2004**). "The perception of speech under adverse acoustic conditions," *Speech Processing and the Auditory System*, edited by S. Greenberg, W. A. Ainsworth, A. N. Popper, and R. R. Fay (Springer, Berlin), Vol. 18.

Bent, T., and Bradlow, A. R. (**2003**). "The interlanguage speech intelligibility benefit," J. Acoust. Soc. Am. **114**, 1600–l610.

Bronkhorst, A. W. (**2000**). "The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions," Acust. Acta Acust. 86, 117–128.

Bronkhorst, A. W., and Plomp, R. (**1992**). "Effect of multiple speechlike maskers on binaural speech recognition in normal and impaired hearing," J. Acoust. Soc. Am. **92**, 3132–3138.

Brungart, D. S., Simpson, B. D., Ericson, M. A., and Scott, K. R. (**2001**). "Informational and energetic masking effects in the perception of multiple simultaneous talkers," J. Acoust. Soc. Am. **110**, 2527–2538.

Byrne, D., Dillon, H., Tran, K., Arlinger, S., and Wilbraham, K. (**1994**). "An international comparison of long-term average speech spectra," J. Acoust. Soc. Am. **96**, 2108–2120.

Cutler, A., Webber, A., Smits, R., and Cooper, N. (**2004**). "Patterns of English phoneme confusions by native and non-native listeners," J. Acoust. Soc. Am. **116**, 3668–3678.

Garcia Lecumberri, M. L., and Cooke, M. (**2006**). "Effect of masker type on native and non-native consonant perception in noise," J. Acoust. Soc. Am. **119**, 2445–2454.

Grabe, E., and Low, E. L. (**2002**). "Durational variability in speech and the Rhythm Class Hypothesis," *Papers in Laboratory Phonology 7* (Moutons-Gravenhage).

Jiang, J., Chen, M., and Alwan, A. (**2006**). "On the perception of voicing in syllable-initial plosives in noise," J. Acoust. Soc. Am. **119**, 1092–1105.

Mattys, S. L., White, L., and Melhorn, J. F. (**2005**). "Integration of multiple speech segmentation cues: A hierarchical framework," J. Exp. Psychol. **134**, 477–500.

Mayo, L. H., Florentine, M., and Buus, S. (**1997**). "Age of second-language

acquisition and perception of speech in noise," J. Speech Lang. Hear. Res. **40**, 686–693.
Nábělek, A. K., and Donohue, A. M. (**1984**). "Perception of consonants in reverberation by native and non-native listeners," J. Acoust. Soc. Am. **75**, 632–634.
Parikh, G., and Loizou, P. (**2005**). "The influence of noise on vowel and consonant cues," J. Acoust. Soc. Am. **118**, 3874–3888.
Ramus, F., Nespor, M., and Mehler, J. (**1999**). "Correlates of linguistic rhythm in the speech signal," Cognition **7**, 265–292.
Rhebergen, K. S., and Versfeld, N. J. (**2005**). "A Speech Intelligibility Index-based approach to predict the speech reception threshold for sentences in fluctuating noise for normal-hearing listeners," J. Acoust. Soc. Am. **114**, 2181–2192.
Rhebergen, K. S., Versfeld, N. J., and Dreschler, W. A. (**2005**). "Release from informational masking by time reversal of native and non-native interfering speech (L)," J. Acoust. Soc. Am. **118**, 1274–l277.
Rogers, C. L., Dalby, J., and Nishi, K. (**2004**). "Effects of noise and proficiency on intelligibility of Chinese-accented English," Lang Speech **47**, 139–154.
Simpson, S. A., and Cooke, M. (**2005**). "Consonant identification in N-talker babble is a nonmonotonic function of N," J. Acoust. Soc. Am. **118**, 2775–2778.
Smiljanic, R., and Bradlow, A. R. (**2005**). "Production and perception of clear speech in Croatian and English," J. Acoust. Soc. Am. **118**, 1677–1688.
Sperry, J. L., Wiley, T. L., and Chial, M. R. (**1997**). "Word recognition performance in various background competitors," J. Am. Acad. Audiol **8**, 71–80.
Studebaker, G. A. (**1985**). "A "Rationalized" Arcsine Transform," J. Speech Hear. Res. **28**, 455–462.
Takata, Y., and Nábělek, A. K. (**1990**). "English consonant recognition in noise and in reverberation by Japanese and American listeners," J. Acoust. Soc. Am. **88**, 663–666.
Van Wijngaarden, S., Steeneken, H., and Houtgast, T. (**2002**). "Quantifying the intelligibility of speech in noise for non-native listeners," J. Acoust. Soc. Am. **111**, 1906–1916.

Automatic acoustic synthesis of human-like laughter[a)]

Shiva Sundaram[b)] and Shrikanth Narayanan[c)]
Speech Analysis and Interpretation Lab (SAIL), Department of Electrical Engineering-Systems, 3740 McClintock Ave, EEB400, University of Southern California, Los Angeles, California 90089[d)]

(Received 3 February 2006; revised 18 October 2006; accepted 18 October 2006)

A technique to synthesize laughter based on time-domain behavior of real instances of human laughter is presented. In the speech synthesis community, interest in improving the expressive quality of synthetic speech has grown considerably. While the focus has been on the linguistic aspects, such as precise control of speech intonation to achieve desired expressiveness, inclusion of nonlinguistic cues could further enhance the expressive quality of synthetic speech. Laughter is one such cue used for communicating, say, a happy or amusing context. It can be generated in many varieties and qualities: from a short exhalation to a long full-blown episode. Laughter is modeled at two levels, the overall episode level and at the local call level. The first attempts to capture the overall temporal behavior in a parametric model based on the equations that govern the simple harmonic motion of a mass-spring system is presented. By changing a set of easily available parameters, the authors are able to synthesize a variety of laughter. At the call level, the authors relied on a standard linear prediction based analysis-synthesis model. Results of subjective tests to assess the acceptability and naturalness of the synthetic laughter relative to real human laughter samples are presented. [DOI: 10.1121/1.2390679]

PACS number(s): 43.72.Ja [DOS] Pages: 527–535

I. INTRODUCTION

Expressiveness is a unique quality of natural human speech. The ability to adequately convey and control this key aspect of human speech is a crucial challenge faced in rendering machine-generated speech more generalizable and acceptable. While the primary focus of past efforts in speech synthesis has been on improving intelligibility, and to some extent naturalness, recent trends are increasingly targeting on improving the expressive quality of synthetic speech (Hamza *et al.*, 2004; Junichi Yamagishi and Kobayashi, 2003, 2004; Narayanan and Alwan, 2004). For instance, natural expressive speech quality is essential for synthesizing long exchanges of human-machine dialogs and for information relaying monologs. There are many ingredients that play a role in imparting expressive quality to speech. These include variations in speech intonation and timing (Dutoit, 1997), modifications of spectral properties, appropriate choice of words, use of other nonlexical (expressions such as throat clearing, tongue clicks, lip smacks, laughter, etc.) and nonverbal (physical gestures, facial expression, etc.) cues. Emotion is an important underlying expressive quality of natural speech that is communicated by a combination of the aforementioned variations. Inclusion of nonlexical and/or nonverbal cues in emotional speech can also regulate the type and degree of emotion being expressed and also improve the clarity of emotions in speech. For example, in Robson and MackenzieBeck, 1999 (and references therein) it has been determined that speech with labial spreading (a nonverbal cue) is aurally interpreted as "smiled" or happy sounding speech. Another avenue being explored involves addition of nonlexical cues in machine synthesized speech (Sundaram and Narayanan, 2003) that can better express the desired emotion or the state of a human-machine dialog. Nonlexical cues for expressing happy sounding speech is important in this respect. Prior work has shown (Bulut *et al.*, 2002; Trouvain and Schröder, 2004) that synthesizing happy sounding speech is one of the most challenging problems and that one has to look beyond just intonation variation, especially, if speech accommodates laughter. This is because the implicit nature of laughter causes variations in the supporting speech and vice versa (Nwokah *et al.*, 1999). Laughter may have different functions in interpersonal speech communication, expressing an amusing or happy context is key among them. Hence it is an important attribute in this context of expressive, synthesized speech. The focus of the present work is restricted to automatic acoustic synthesis of laughter by machines. It can be used to enhance the expressive quality of the accommodating synthesized speech and/or aid in communicating a happy or amusing context.

Speech, in humans, is a more controlled and a better understood process that is governed by the rules of a language's grammar. Therefore, for machine synthesis of speech, for example, the phrase "*How are you*?," the required sequence of sounds (as phonemes) and the expected intonational variation are fairly well prescribed, even for different situations and context. Text analysis on the given phrase, and existing intonation models are used in the generation of the final waveform. Also, natural speech audio examples are abundantly available for aiding analysis and modeling. Thus the inputs required to generate any word in

[a)]Abstract previously appeared in the J. Acoust. Soc. Am. 116, 2481 (2004). Part of this paper has been published previously as an invited lay language paper for the 148th ASA Meeting, San Diego, California.
[b)]Electronic mail: shiva.sundaram@usc.edu
[c)]Electronic mail: shri@sipi.usc.edu
[d)]URL: http://sail.usc.edu

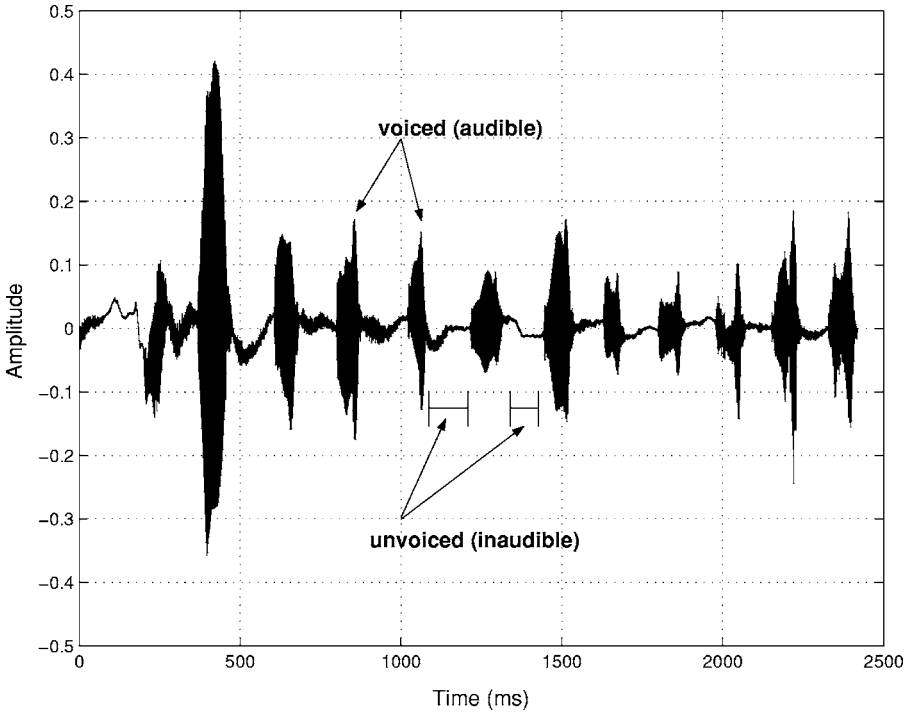


FIG. 1. The alternating phenomenon of a laughter: A laugh cycle with intermittent laugh pulses.

synthesized speech are relatively well defined. Likewise, to synthesize laughter, we require appropriate signal models to generate a particular type of laughter. Unlike spoken language, while there is no guiding grammar for synthesizing laughter, there is a characteristic *texture* to it. Using the simple model proposed in this paper, it is possible to generate laughter using a set of input control parameters, and different types of laughter can be generated by varying these parameters. In this work, we attempt to answer the basic question of *how to synthesize laughter* at the acoustic level; the cognitive/semantic aspects of laughter generation or issues related to including synthesized laughter in conjunction with synthesized speech are beyond the scope of this paper. In the following sections we introduce some common terminologies used to describe laughter and discuss some issues of interest in the synthesis of laughter.

A. Background

We present a brief description of the terms used in this paper to describe the various segments of laughter that have been adopted from acoustic primatology (Bachorowski *et al.*, 2001). In Fig. 1, the waveform of an actual laughter episode is shown. A single instance/episode of laughter, beginning with an inhalation to its end, is known as a laughter *bout*. Each bout comprises alternating voiced (audible) and unvoiced (relatively inaudible) sections. This alternating phenomenon is also termed as a *laugh cycle* with intermittent *laugh pulses* with aspiration sounds in between them (Ruch and Ekman, 2001). The voiced section, illustrated by sections of large amplitude in the figure, is known as a laughter *call* or *laugh pulse* (also referred to as voiced call in this paper). The time interval between two laughter calls is the *inter-call* interval. A laughter call can be a vowel-like sound, for example, in calls such as "ha," or have a grunt-like or snort-like quality (Bachorowski *et al.*, 2001). Such qualities are more evident in a spectrogram of the complete laughter bout. Related details about segmentation of laughter based on its acoustic analysis can be found in Bachorowski *et al.*, 2001; Provine, 2000; and Ruch and Ekman, 2001. Other segmentation schemes are also possible. In Trouvain, 2003 the author discusses syllable and phrase level segments for laughter and its relationship to the terms introduced previously. While it can be useful for studying or categorizing laughter types, these concepts are not directly relevant for the laughter generation model presented here.

B. Variation in laughter and its synthesis

The production of laughter is a highly variable physiological process. Provine (Provine, 2000), describes it to be a very strange expression whose peculiarity is masked by its familiarity. Laughter is an expression with a very distinct pattern. The texture of laughter has variations across gender, and across individuals (Bachorowski *et al.*, 2001; Provine, 2000). Every situation has its own appropriate and inappropriate types of laughter, and even for the same context, an individual can choose to laugh differently at different times. It is used as a vocalized punctuation in a question (Provine, 2000), and it also occurs along with speech (termed as "speech laughs") (Nwokah *et al.*, 1999; Trouvain, 2001). Overall, it can be interpreted as a vocalized expression that bridges the gap between an emotional state of excitement and a neutral emotional state. Laughter differs from smiling because the later is essentially a nonverbal facial expression which, under certain circumstances, may lead to a distinct, audible laughter episode. However, laughter and smiling may share the same facial expression.

While the qualities of the vocalization and issues of duration take their own course during an episode, a limited control by the individual determines the overall duration and number of laugh pulses in a bout or laugh cycle. Thus large

variations in the number of calls per bout and duration of each call is observed in real laughter (Bachorowski *et al.*, 2001). It is known axiomatically that no two instances of laughter are exactly the same, yet they have implicit characteristics that bring out individual traits. Some specific attributes that cause these variations include pitch changes during a bout, pitch changes within a call, duration of the complete bout, duration of a voiced call, the loudness of the calls, and the type of call (vowel-like or grunt-like, etc). Thus, specifications of these components are required to generate an episode of laughter. While the vowel-like sound within a call is a matter of choice, the duration of a bout, duration of each call, and periodicity of the laughter calls are a part of the pattern of the laughter bout. The specifications for the latter are the input control parameters obtained from an appropriately defined generative model for laughter production. From an engineering perspective, a generative model for laughter is challenging because it should meet the following constraints:

- The model should be able to handle a wide range of the variability seen in the physiological process of laughter.
- It should have the provision to generate different types of laughter, e.g., short bursts or a long train of laughter depending on the immediate context. A parametric control over the generated laughter is preferred from an automatic synthesis point of view.
- The model should be convenient to use. It should be able to generate laughter based on simple, easily available information.

The model described in this paper has two major components. The first component focuses on modeling the behavior of the overall episode (or bout), and is based on a simple second-order mass-spring dynamical systems model, akin to one that describes the simple harmonic motion of an oscillating pendulum (refer to Fig. 2). The second component uses a linear prediction (LP) based analysis-synthesis model, which is widely used in speech processing. Note that for this second component, any other speech synthesis/modification technique such as the time domain pitch synchronous overlap add (TD-PSOLA) (Moulines and Charpentier, 1990) may be used. In this work, we restrict our study to the varieties of laughter that the *spontaneous*, i.e., those that are produced without any restraint. The laughter is assumed to always contain vowel-like voiced calls. The rest of this paper provides details of the model of a mass-spring system and how its equations are used to synthesize a bout of laughter. We also present results of subjective tests performed to assess the perceived naturalness of synthesized laughter against real human laughter. It should be noted, however, that assessment of synthetic speech and laughter is a highly challenging task. It is well known that the rich diversity and variability that make up natural speech also make evaluation of machine-generated speech difficult. The study of perception of everyday natural speech spans a very large domain of problems in speech synthesis and other related sciences. The techniques that are available for speech analysis/synthesis tackle only a subset of the rich possibilities in problems of speech generation (Dutoit, 1994; McAulay and Quatieri, 1986; Moulines and Charpentier, 1990). Many of the challenges to achieve even near natural speech for synthesis are yet to be solved for a fair evaluation of natural versus synthesized speech. For example, in Syrdal *et al.*, 1998 the authors subjectively compared two diphone based speech synthesis techniques with natural speech in terms of intelligibility, naturalness, and pleasantness. It was found that natural speech was consistently perceived to be better than synthetic speech. Still, comparison of real, natural oral gestures to machine synthesized ones provides an assessment of the variables that are useful or lacking in mimicking the gesture under study. We follow a similar approach in evaluating the synthetic laughter samples created in this work.

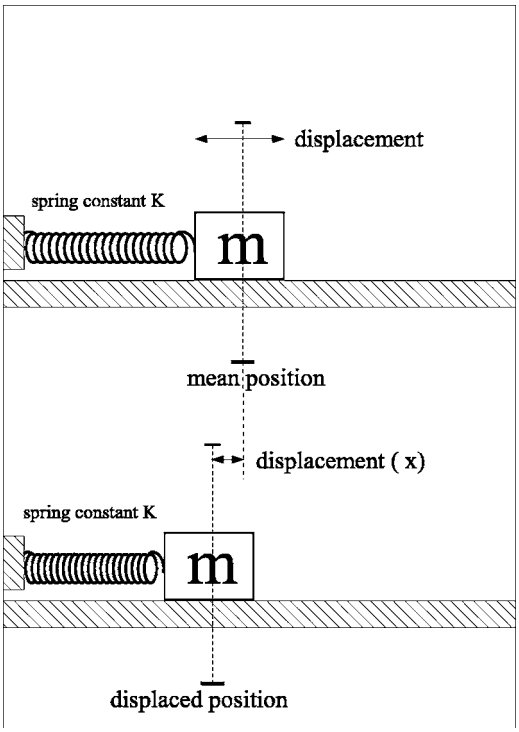


FIG. 2. A mass-spring model.

II. ACOUSTIC MODEL FOR LAUGHTER

An engineering solution to describe an unknown system is to propose a mathematical model based on a set of observations of the system behavior. Figure 1 exemplifies two striking features of a typical laughter bout: alternating segments of audible, voiced section and inaudible unvoiced parts with the envelope of the peaks of the voiced calls falling across the duration of the laughter bout. A laughter starts with a contextual or semantic impulse, that puts the speaker in a *laughing state*. While laughing, there are bursts of air exhalation (along with audible voicing) and aspiration (unvoiced segment) that each last for a short period. This intermittent voicing pattern can be seen as an *oscillatory behavior* that can be observed in most laughter bouts. This pattern has been noted by other researchers as well (Bachorowski *et al.*, 2001; Provine, 2000; Ruch and Ekman, 2001).

We model this oscillatory behavior of alternate voiced and unvoiced segments with equations that describe the simple harmonic motion of a mass attached to the end of a

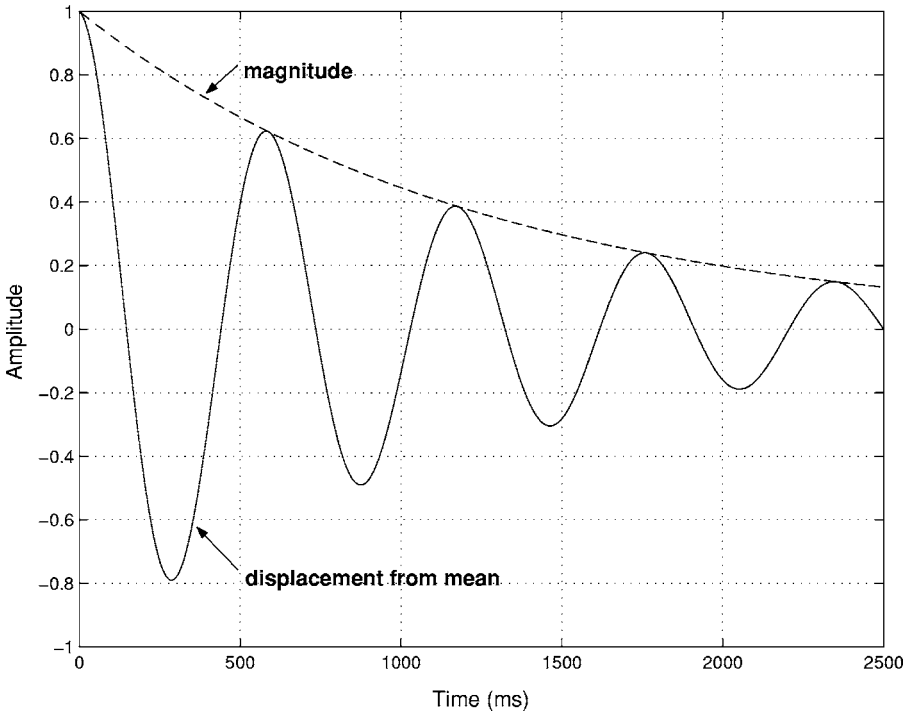


FIG. 3. Damped simple harmonic motion. Amplitude variation over time.

spring (illustrated in Fig. 2). In this simple mass-spring system, the stiffness of the spring and the weight of the mass determine the frequency of oscillation of the mass. The initial displacement and the damping factor determine how long the mass would continue to oscillate, and the rate of envelope decay. We next briefly explain the steps in building a mathematical model for an oscillating mass attached to a spring, and also motivate the idea that this model can be used to explain the oscillatory behavior of laughter for its automatic synthesis.

A. Oscillatory behavior of laughter

Let a mass m be displaced from its initial rest position by x (refer to Fig. 2). This will cause the spring to compress in length by an amount x. When the mass is released at some time t, the compressed spring will act on the mass and accelerate the mass in a direction opposite to the initial displacement. This force that the spring exerts on the mass (denoted by F_{spring}) is directly proportional to the compression x, i.e.,

$$F_{\text{spring}} \propto x.$$

By Newton's First Law, the mass m, its acceleration $a = d^2x/dt^2$ and the force F_{spring} are related by the equation

$$m\frac{d^2x}{dt^2} = -kx, \tag{1}$$

where k is the constant of proportionality, also known as the spring constant. The negative sign on the right hand side of Eq. (1) arises because the direction of force generated by the compressed spring is opposite to the direction of the displacement causing the compression.

A solution to this second-order system is given by the expression

$$x = e^{-j(\sqrt{k/m})t}. \tag{2}$$

This is a sinusoid with $(1/2\pi)\sqrt{k/m}$ as its frequency of oscillation. If this system experiences a damping force proportional to its velocity, then Eq. (1) becomes

$$m\frac{d^2x}{dt^2} = -kx - b\frac{dx}{dt}, \tag{3}$$

where b is the damping constant (this case arises with a simplified damping due to a fluid external to the mass m) and the corresponding general solution for damped simple harmonic motion becomes

$$x(t) = Ae^{-Bt}e^{-j(\sqrt{k/m})t}, \tag{4}$$

where $B = b/2m$. The result obtained in Eq. (4) is that of a damped sinusoid, that parametrically describes the motion of a damped simple harmonic motion system.

Figure 3 illustrates the plot of time t versus amplitude x (solid line) of such a damped sinusoid with Ae^{-Bt} (dotted line). By studying the figure it becomes evident that the peak-amplitude envelope decay of the voiced calls in a laughter bout is similar to a damped sinusoid, but where the parameters A, k, m, B are actually $A(t)$, $k(t)$, $m(t)$, $B(t)$, i.e., they are allowed to vary over time. This is illustrated in Fig. 4 where the plot of a damped sinusoid model is superimposed on a real human laughter sample. Here, the parameters $A(t)$, $k(t)$, $m(t)$, $B(t)$ are allowed to vary as a piecewise linear function of time. On visual inspection, even the duration of the positive cycle of the oscillator (which is directly related to the frequency of oscillation) matches with the duration of the intermittent laugh pulses of the laugh cycle. This is also true for the unvoiced segments of the bouts that match with the duration of the negative cycle of the oscillation. This

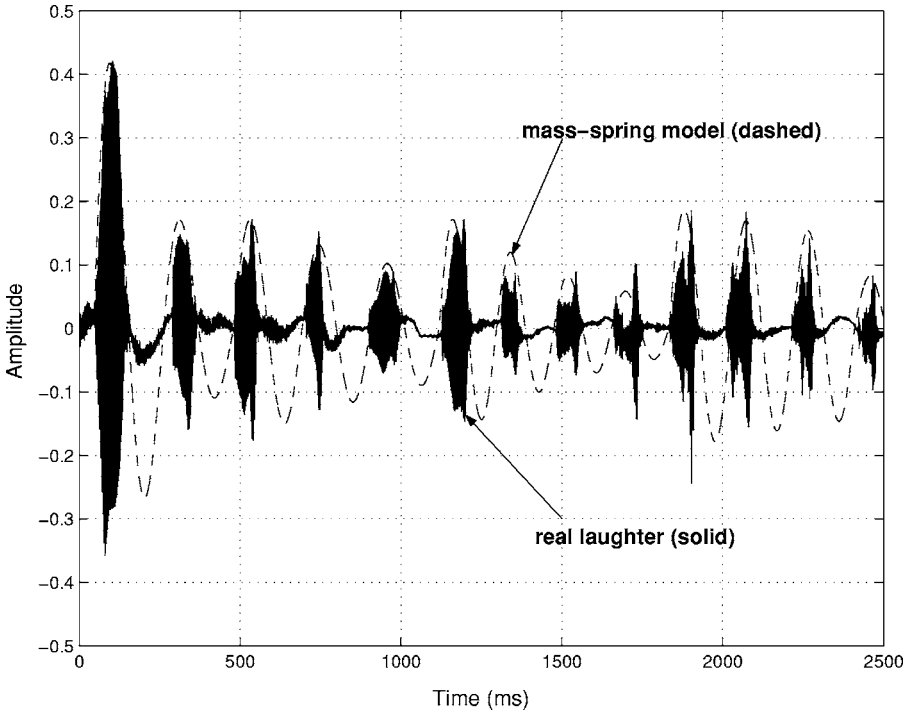


FIG. 4. A mass-spring model trajectory superimposed on a real laughter bout.

aspect is further substantiated by the finding that the call duration and the inter-call intervals are comparable in laughter bouts (Bachorowski *et al.*, 2001).

Other possible variations of the damped oscillator described earlier include forced and damped oscillation where the system is either forced periodically or at random instances during the oscillation of the body. Variations in the nature of the damping force can also cause different oscillatory behavior. One such example is when the damping force is a constant frictional force (Marchewka *et al.*, 2004). In essence, any arbitrarily complex oscillatory behavior can be generated using this basic model, and virtually any pattern of oscillation can be generated by controlling the basic parameters such as $A(t)$, $k(t)$, $m(t)$, $B(t)$.

The other model component relates to the voiced-call units. Since these are vowel-like vocalizations, analysis-synthesis techniques used in conventional speech processing can be directly adopted. Different vowel-like laughter calls can be synthesized by changing the user-defined linear prediction (LP) coefficients or the speech data associated with the synthesizer. The procedure is explained briefly. LP based analysis-synthesis assumes a source-filter model of speech production. The LP coefficients can be extracted (the analysis part) from a sample waveform of speech using standard, well known procedures such as the Levinson-Durbin algorithm (many existing speech analysis software tools have inbuilt LP analysis functions). The estimated LP coefficients define an all-pole filter; and when excited with an appropriate input (such as a pulse train), it can generate (the synthesis) a speech sound at the output (for example, a vowel). Since the set of LP coefficients is primarily dependent on the sample waveform at the time of analysis, different vowel-like sounds for laughter calls can be synthesized by changing the speech data during analysis (essentially using a different set of LP coefficients). Further details about the LP analysis-synthesis techniques can be found in Rabiner and Schafer, 1978.

Thus, one could synthesize segments of voiced calls by using the above duration, time-position and peak-amplitude information, and thereby synthesize a complete laughter bout. By changing the input parameters, laughter bouts with different patterns can be generated. For example, if the damping factor of the previously described system is reduced, then the oscillation will last for a longer duration and thus a longer laughter bout can be synthesized. Similarly, if the values of mass or spring constant are changed, then the frequency of the laughter calls in a bout can be changed. Also, by using different waveform synthesis schemes, other snorting or grunt-like qualities can be imparted to the laughter calls.

The main advantages of this model are summarized below:

- For a given set of parameters, the same model directly presents the duration, timing, and peak-amplitude decay of the laughter calls in an episode of laughter simultaneously.
- By an appropriate choice of parameters such as pitch variation, and choice of $A(t)$, $k(t)$, $m(t)$, $B(t)$ functions, any real human laughter can be accurately represented.
- There is a clear, direct, and predictable relation between the control parameters and the generated pattern of laughter.
- There is no restriction on the speech synthesis technique used for synthesizing the calls in the laughter. Linear prediction (LP) analysis-synthesis method has been used in this work due to its ease of implementation. Other speech modification techniques such as the TD-PSOLA can also be used.

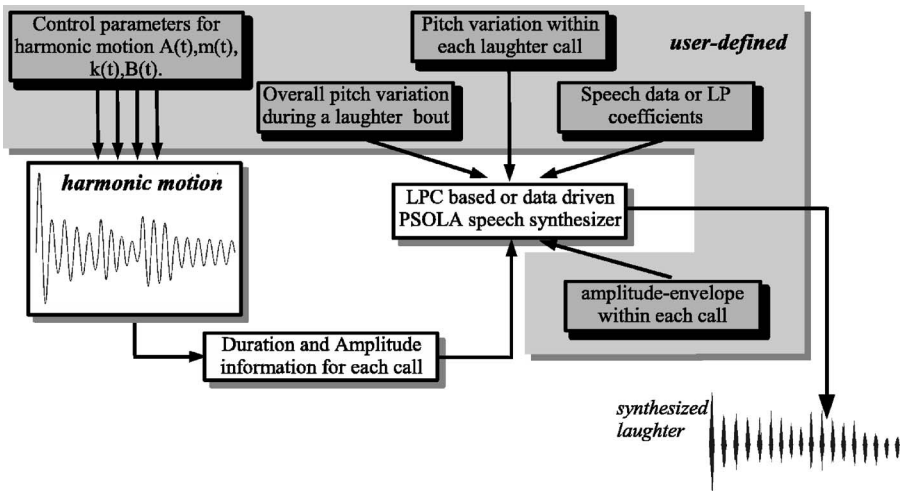


FIG. 5. Steps involved in the synthesis of laughter described in this paper. The shaded boxes depict the inputs required from a user and the unshaded boxes combine to form the laughter synthesizer.

Thus we have a simple *generative* model with user-defined input control parameters such as $A(t)$, $k(t)$, $B(t)$, $m(t)$ which would give us the ability to model and vary the duration of the laughter calls, time position of the calls, and the peak amplitude variation of the calls over the course of a laughter bout. Depending on the synthesizer used, there is also no restriction on the type of laughter call that would be used to generate the complete laughter bout. Figure 5 illustrates the complete laughter synthesis methodology adopted in this paper.

B. Laughter synthesis procedure

Referring to Fig. 5, the procedure followed to synthesize laughter is summarized below:

1. For a set of given (user-defined) time varying functions $A(t)$, $k(t)$, $b(t)$, $m(t)$ calculate the duration and peak amplitude and onset time of each positive cycle in the resulting harmonic motion $x(t)$. Let this harmonic motion comprise N_{pos} positive cycles. Thus we will have N_{pos} laughter calls in the synthesized laughter bout.
2. Let $P_{\text{mean}}(i)$, $p_{\text{var}}(t,i)\,\forall\, i \in \{1,2,\ldots,N_{\text{pos}}\}$ and $t \in [0, T_d(i)]$ be the mean pitch and pitch variation within each call, respectively, where $T_d(i)$ is the duration of the ith laughter call. The values for $P_{\text{mean}}(i)$ and $p_{\text{var}}(t,i)$ are defined by the user. Alternatively, these parameters can also be obtained from acoustic analysis of real laughter clips. Note that $P_{\text{mean}}(i)$ is a discrete positive value of a laughter call pitch and the set $p_{\text{var}}(t,i)$ is a set of functions continuous in time that have positive real values $\forall i$. To have meaningful outputs, the order of $P_{\text{mean}}(i)$ and $p_{\text{var}}(t,i)$ is equal to that of measured F0 values of normal speech. Usually, the exact target values are obtained by analysis of clips of real human laughter.
3. Using the peak-amplitude and duration information obtained in Step 1, in addition to the $P_{\text{mean}}(i)$ and $p_{\text{var}}(t,i)$, synthesize each laughter call $\forall i \in \{1,2,\ldots,N_{\text{pos}}\}$.
4. Similar to Step 1, the duration and peak-amplitude information can also be extracted for the negative cycles of the harmonic motion. This can be used to include audible aspiration noise.
5. Finally, arrange the N_{pos} laughter calls in series in time according to the onset time instances obtained in Step 1 and thus construct the overall laughter bout.

The complete laughter synthesis system described in Fig. 5 was implemented in MATLAB (*http://www.mathworks.com*). The user-defined inputs included the overall variation of pitch in a laughter bout, the pitch variation within each laughter, amplitude envelope within each call, and parameters for the call level synthesis. These were provided to the system by the authors using a graphical user interface (GUI) at runtime. The $A(t)$, $k(t)$, $b(t)$, $m(t)$ values were also provided at runtime. The GUI inputs for the amplitude envelope within each call was low-pass filtered with a third-order finite impulse response low-pass filter to smooth the envelope.

It is important to point out that extracting timing and peak-amplitude information for voiced and unvoiced elements of laughter from the positive and negative cycles, respectively, is a matter of practical convenience. It does not bear any direct relevance to the actual physiology of laughter. However, the work we present here alludes to the fact that real human laughter can be interpreted as a form of oscillation. The voiced calls of a real laughter episode are not truly vowel-like sounds. In a real laughter episode, the vocal tract configuration can change rapidly and/or other noise (such as aspiration noise) is always present. Therefore, to get satisfactory synthesis quality, the data for the LP parameters for the call level waveform synthesis were extracted from voiced segments of normal speech. Samples of synthetic laughter can be found at http://sail.usc.edu/emotion.

The next section describes the subjective experiment to assess the perceived naturalness of the synthesized laughter.

III. EXPERIMENT

Subjective evaluation tests were performed on 28 naive volunteers at the Speech Analysis and Interpretation Laboratory (SAIL) at USC. The volunteers were presented with 25 laughter-only clips of which 17 clips were synthesized offline using the technique presented here. The number of calls in the synthesized laughter, its duration, the F0 changes in

TABLE I. A summary of the properties of the 17 synthesized and eight real samples of laughter used in the listening experiments. The eight real laughter samples are marked with an (*).

Sample	mean F0 (Hz)	min F0 (Hz)	max F0 (Hz)	Std F0 (Hz)	No. calls	Duration (s)	Gender
01	203.30	178.0	235.0	17.00	12	2.30	M
02	246.75	175.0	305.0	45.15	9	1.60	M
03	224.60	202.0	259.0	21.05	5	0.84	M
04	276.50	223.0	372.0	53.64	10	1.60	F
05	244.50	165.0	299.0	32.55	18	2.90	F
06	274.00	211.0	350.0	55.94	7	1.00	F
07	190.00	142.0	231.0	26.81	15	2.50	M
08	354.00	286.0	417.0	39.00	8	1.32	M
09	218.66	202.0	235.0	16.50	3	0.43	M
10	365.00	293.0	425.0	15.00	6	0.97	M
11	198.00	171.0	221.0	17.75	8	1.18	M
12	333.40	302.0	362.0	20.48	12	2.45	F
13	267.20	211.0	317.0	46.85	5	0.82	F
14	250.00	192.0	312.0	41.87	6	0.93	M
15	279.35	223.0	419.0	49.92	18	3.02	F
16	225.00	139.0	319.0	51.70	16	2.39	F
17	340.50	304.0	393.0	29.75	11	1.90	F
01*	204.17	85.5	331.5	74.15	11	2.06	M
02*	199.60	111.7	315.5	52.64	17	3.15	M
03*	309.72	234.2	364.5	26.70	10	2.20	F
04*	174.14	81.1	250.4	46.93	10	2.00	F
05*	243.85	142.3	314.4	39.21	17	2.81	F
06*	257.50	161.6	332.7	56.57	18	3.67	F
07*	180.81	111.4	294.2	30.08	12	2.02	M
08*	255.90	192.6	358.9	44.36	14	3.02	M

each sample are given in Table I. The remaining eight were clips of real human laughter. The laughter type in these eight clips matched the 17 clips of synthesized laughter. The 25 clips were randomly played and not grouped in any particular order. The tests were performed in a typical quiet office environment on a computer terminal. Each volunteer had to listen and score each sample for *naturalness* and *acceptability* according to their preference on a scale of 1–5: 1-Very Poor, 2-Poor, 3-Average, 4-Good, 5-Excellent. The samples were presented on an interactive webpage-like GUI. The subject could click on a sample, listen to it and click on the appropriate *naturalness* and *acceptance* score. The samples were played at 22,050 Hz sample rate over a pair of commercially available Sony MDR-XD100 headphones that could be adjusted to snugly fit the listener. The complete evaluation took about 11 min for each subject.

The eight clips of real isolated laughter were collected from two sources: four were extracted from a compact disk and downsampled to 22,050 Hz (Junkins, 2005). These were from tracks of recorded laughter intended for laughter therapy and the remaining four were obtained from a database of laughter episodes that we recorded independently. This database was created by recording volunteer subjects who were simply asked to laugh impromptu for a laughter synthesis project. For each of the subjects the first few laughter instances seemed to be forced and were rejected and the later episodes that seemed more natural to us were kept in the database. The eight clips were selected based on how well they aurally matched the synthesized laughter used in listening tests. Many candidate clips were rejected because of speaker's movement while laughing, unidentified noises picked up by the microphone, and change of laughter call type during the bout. To make all the tracks similar, and reduce any extraneous bias in listener assessment, noise extracted from the silent parts of the compact disc tracks were extracted, downsampled to 22,050 Hz sample rate, and added to the other synthesized and recorded clips.

IV. RESULTS

At the time of analysis of the results the evaluations of the volunteers were grouped into Group I and Group II according to their language background. Group I comprised four female and five male subjects whose first language was American English and Group II comprised seven female and twelve male subjects whose second or third language was English. For the analysis of the evaluations, we make the assumption that each laughter clip is an independent encounter by an individual subject. Thus, for $N=28$ subjects and 17 synthesized samples, we have a total of $28\times17=476$ samples and for the eight real laughter clips, we have $28\times8=224$ samples.

The mean and variance of the evaluation scores are listed in Table II. The evaluation results are summarized below:

Mean evaluation scores: A t Test (with unequal variance, and degrees of freedom (df)$=440$) was performed to compare the evaluation scores of real and synthesized laughter clips. For the given experiment it was found that at $\alpha=10^{-4}$, there is a significant difference in the mean natural-

TABLE II. Mean evaluation scores for natural and synthesized clips

Group	Synthesized clips *mean, variance*	Real clips *mean, variance*
Groups I evaluations	naturalness: 1.49, 0.044	naturalness: 4.36, 0.014
	acceptability: 1.66, 0.053	acceptability: 4.34, 0.079
Groups II evaluations	naturalness: 1.81, 0.084	naturalness: 4.38, 0.034
	acceptability: 2.21, 0.077	acceptability: 4.36, 0.014
Groups I & II evaluations	naturalness: 1.71, 0.600	naturalness: 4.28, 0.59
(overall)	acceptability: 2.03, 1.020	acceptability: 4.35, 0.73

ness scores between the real and synthesized laughter clips. It was also found that at $\alpha=10^{-4}$, there is a significant difference in the mean acceptability scores of real and synthesized laughter clips.

Sample-wise test: A parametric single-factor analysis of variance (ANOVA) was performed to determine differences in the evaluation among the synthesized clips. For a total of $N=28$ evaluations for each clip, it was found that at ($\alpha=0.05$; $df=16/459$) there was a significant difference in the mean naturalness scores among the synthesized laughter clips. However, at $\alpha=10^{-4}$, there was no significant difference in the mean naturalness score. A single-factor ANOVA performed on the acceptability scores among the synthesized clips indicated that for ($\alpha=10^{-4}$; $df=16/459$) there was no significant difference in the mean scores among the synthesized laughter clips. A single-factor ANOVA of the mean naturalness score ($\alpha=10^{-4}df=7/216$) for the real laughter clips showed no significant differences among the evaluations of the real laughter clips. The single-factor ANOVA of the mean acceptability scores ($\alpha=10^{-4}$, $df=7/216$) also indicated no significant differences among the evaluations of the real-laughter clips.

Group-wise test: A parametric single-factor ANOVA test of the naturalness scores of real laughter clips between Group I and II indicated a significant difference in the scores at ($\alpha=0.05$; $df=1/223$). However, at ($\alpha=10^{-4}$; $df=1/223$), the test indicated no significant differences in the mean evaluation scores between the groups. This same trend was observed when the mean evaluation acceptability scores for synthesized clips were compared for Group I and II. The same ANOVA test ($\alpha=10^{-4}$; $df=1/223$) performed on the mean acceptability scores indicated no significant difference in the mean scores between the two groups for the synthesized clips. However, for synthesized laughter clips it indicated significant difference between Group I and II at ($\alpha=0.05$; $df=1/223$) and at ($\alpha=10^{-4}$; $df=1/223$).

V. DISCUSSION, CONCLUSION AND FUTURE WORK

In this paper we have presented a two-level parametric model for human laughter. The first level of the model captures the overall temporal behavior of a laughter episode. At the next level, we model the audible calls with conventional LP coefficients based analysis-synthesis and/or TD-PSOLA speech modification technique that are widely used in speech processing. The model presented is based on the idea that laughter in human beings can be interpreted as an oscillation, where exhalation alternates with inaudible segments. We also presented properties of laughter episodes that can be captured by the model parameters. Motivated by the need for computer synthesis of laughter for emotional speech synthesis, we applied this idea to synthesize different varieties of laughter and evaluate them in terms of two subjective measures: *naturalness* and *acceptability*. We also compared this evaluation with evaluation of real laughter clips.

The results obtained are similar to the results obtained by Syrdal *et al.*, 1998 where different speech synthesis techniques were evaluated against natural speech. Subjective assessment of real, human, natural expressions is consistently better than synthesized ones. Two main factors that can be attributed to this dichotomous result are the limitations in the variety of features that are included in synthesized laughter and the inherent artifacts present during the final waveform synthesis. For example, unlike natural laughter bouts, we synthesize bouts with relatively simple vowel-like voiced calls. Also, the perceivable artifacts during waveform synthesis are caused due to issues with precise generation of natural sounding pitch contours, and obtaining smooth frame to frame spectral variations. These artifacts give negative cues to the listener that result in unnatural perception of synthesized clips. Another issue deals with the underlying quality that is being evaluated: perceived naturalness. While naturalness is a loose term, a very stringent set of standards is followed to label perceived speech as natural. What is truly regarded as natural and/or acceptable is already encoded in the listener. This is because everyday human speech communication is perceived as highly natural speech and it is abundant with a wide range of qualities that are not imparted in synthesized speech. For the particular case of laughter, due to its high degree of variability, the evaluation in terms of perceived naturalness becomes a bigger issue. It is also difficult to define a quantitative measure or a quantitative set of parameters to define an *acceptable* form of laughter. This is because, such a measure covers a gamut of social, acoustical, and perceptual metrics. Even in the case of real human laughter, for example, if the bout is spontaneous and placed appropriately in a dialog, it is *more* natural and acceptable than when its forced and/or inappropriate. Thus it is difficult to make raw comparisons. The results of the experiments also indicate that synthesized laughter is interpreted differently by different individuals. The evaluation experiments presented here are very limited in scope: they evaluate results of isolated laughter episodes without context or accompanying speech. In attempting to answer the question "*What makes laughter laughter*?" the research presented in this pa-

per sheds a different light on this question by generating laughter than the ones addressed by other researchers through acoustic analysis. The experiments have been designed to evaluate only the synthesis aspects of laughter and its perception.

Computer synthesis of laughter is primarily for expressive speech synthesis, a challenge currently being addressed in the speech synthesis our simple approach appears promising, much remains to be done in integrating laughter within an overall synthesis system. We would like to extend this work to include laughter in synthesized happy speech. To merge laughter and speech requires appropriate prosodic and intonational modifications to the accompanying speech and appropriate choice of words and context tracking. This is a harder problem and part of our future goals. The proposed model can also be incorporated with audio-visual synthesis such as with computer generated *avatars* and other virtual agent technologies. Such an effort would entail combining the acoustic aspects of synthesis with visual gestures such as movement of the lips, face and head. These efforts are topics of our ongoing and future work.

ACKNOWLEDGMENT

The work reported in this paper was supported in part by grants from the NSF and the U.S. Army.

Bachorowski, J.-A., Smoski, M. J., and Owren, M. J. (**2001**). "The acoustic features of human laughter," J. Acoust. Soc. Am. **110**, 1581–1597.

Bulut, M., Narayanan, S., and Syrdal, A. (**2002**). "Expressive speech synthesis using a concatenative synthesizer," in *Proceedings of the Seventh International Conference on Speech and Language Processing (ICSLP)*, Denver, pp. 1265–1268.

Dutoit, T. (**1994**). "High quality text-to-speech synthesis: A comparison of four candidate algorithms," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Vol. 1, pp. 565–568.

Dutoit, T. (**1997**). "An Introduction to text-to-speech Synthesis" (Kluwer, Dordrecht).

Hamza, W., Bakis, R., Eide, E. M., Picheny, M. A., and Pitrelli, J. F. (**2004**). "The IBM expressive speech synthesis system," in *Proceedings of the International Conference on Spoken Language Processing (ICSLP)*, Jeju, South Korea, pp. 2577–2580.

Junichi Yamagishi, T. M., and Kobayashi, T. (**2003**). "Modeling of various speaking styles and emotions for HMM-based speech synthesis," in *Proceedings of Eurospeech, Geneva, Switzerland*, pp. 2461–2464.

Junichi Yamagishi, T. M., and Kobayashi, T. (**2004**). "HMM-based expressive speech synthesis-towards TTS with arbitrary speaking styles and emotions," Special workshop in Maui (SWIM).

Junkins, E. (**2005**). "Lots of laughter," http://www.laughtertherapy.com, 3200 N. MacArthur Blvd., Ste. 106, Irving, TX 75062. Last accessed 12/6/06.

Marchewka, A., Abbot, D. S., and Beichner, R. J. (**2004**). "Oscillator damped by a constant-magnitude friction force," Am. J. Phys. **74**(4), 477–483.

McAulay, R. J., and Quatieri, T. F. (**1986**). "Speech analysis synthesis based on sinusoidal representation," IEEE Trans. Acoust., Speech, Signal Process. **34**, 744–754.

Moulines, E., and Charpentier, F. (**1990**). "Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones," Speech Commun. **9**, 453–467.

Narayanan, S., and Alwan, A. (**2004**). *Text-To-Speech Synthesis: New Paradigms and Advances* (Prentice–Hall, Englewood Cliffs, NJ).

Nwokah, E. E., Hsu, H.-C., Davies, P., and Fogel, A. (**1999**). "The integration of laughter and speech in vocal communication: A dynamic systems perspective," J. Speech Lang. Hear. Res. **42**, 880–894.

Provine, R. R. (**2000**). *Laughter: A Scientific Investigation* (Viking, New York).

Rabiner, L. R., and Schafer, R. W. (**1978**). *Digital processing of speech signals*, Prentice-Hall Signal Processing Series (Prentice–Hall, Englewood Cliffs, NJ).

Robson, J., and MackenzieBeck, J. (**1999**). "Hearing smiles-perceptual, acoustic and production aspects of labial spreading," in *Proceedings of the International Conference of the Phonetic Sciences (ICPhS)*, San Francisco, pp. 219–222.

Ruch, W., and Ekman, P. (**2001**). "The expressive pattern of laughter," in *Emotions, Qualia and Consiousness, World Scientific, Series on Biophysics and Biocybernetics*, edited by Alfred Kaszniak (World Scientific, Singapore), Vol. 10, pp. 426–443.

Sundaram, S., and Narayanan, S. (**2003**). "An empirical text transformation method for spontaneous speech synthesizers," in *Proceedings of EUROSPEECH*, Geneva, Switzerland, pp. 1221–1224.

Syrdal, A., Stylianou, Y., Garrison, L., Conkie, A., and Schroeter, J. (**1998**). "TD-PSOLA versus harmonic plus noise model (HNM) in diphone based speech synthesis," *IEEE International Conference on Acoustics, Speech and Signal Processing*, Seattle, WA, pp. 273–276.

Trouvain, J. (**2001**). "Phonetic Aspects of 'Speech-Laughs'," in *Proceedings of Conference on Orality and Gestuality (ORAGE)*, Aix-en-Provence, France, pp. 634–639.

Trouvain, J. (**2003**). "Segmenting phonetic units in laughter," in *Proceedings of the 15th International Conference of the Phonetic Sciences (ICPhS)*, Barcelona, Spain, pp. 2793–2796.

Trouvain, J., and Schröder, M. (**2004**). "How (not) to add laughter to synthetic speech," in *Proceedings of the Workshop on Affective Dialogue Systems*, Kloster Irsee, Germany, pp. 229–232.

Quasistatic nonlinear characteristics of double-reed instruments

André Almeida[a)]
IRCAM–Centre Georges Pompidou–CNRS UMR9912, 1 Place Igor Stravinsky, 75004 Paris, France

Christophe Vergez[b)]
Laboratoire de Mécanique et Acoustique–CNRS UPR7051, 31 Ch. Joseph Aiguier, 13402 Marseille Cedex 20, France

René Caussé[c)]
IRCAM–Centre Georges Pompidou–CNRS UMR9912, 1 Place Igor Stravinsky, 75004 Paris, France

(Received 15 June 2006; revised 12 October 2006; accepted 15 October 2006)

This article proposes a characterization of the double reed in quasistatic regimes. The nonlinear relation between the pressure drop, Δp, in the double reed and the volume flow crossing it, q, is measured for slow variations of these variables. The volume flow is determined from the pressure drop in a diaphragm replacing the instrument's bore. Measurements are compared to other experimental results on reed instrument exciters and to physical models, revealing that clarinet, oboe, and bassoon quasistatic behavior relies on similar working principles. Differences in the experimental results are interpreted in terms of pressure recovery due to the conical diffuser role of the downstream part of double-reed mouthpieces (the staple). © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2390668]

PACS number(s): 43.75.Ef [NHF] Pages: 536–546

I. INTRODUCTION

A. Context

The usual method for studying and simulating the behavior of self-sustained instruments is to separate them into two functional parts that interact through a set of linked variables: the resonator, typically described by linear acoustics, and the exciter, a nonlinear element. Although this separation may be artificial because of the difficulty in establishing a precise boundary between the two systems, it is usually a simplified view that allows one to describe the basic functioning principles of the instrument. In reed instruments, for instance, the resonator is assimilated to an air column inside the bore, and the exciter to the reed, which acts as a valve.

In the resonator of reed instruments, the relation between the acoustic variables, pressure (p) and volume flow (q), can be described by a linear approximation to the acoustic propagation which has no perceptive consequences in sound simulations [Gilbert *et al.* (2005)]. On the other hand, the exciter is necessarily a nonlinear component, so that the continuous source of energy supplied by the pressure inside the musician's mouth can be transformed into an oscillating one [Helmholtz (1954); Fletcher and Rossing (1998)]. The characterization of the exciter thus requires the knowledge of the relation between variables p and q at the reed output (the coupling region). In principle this relation is noninstantaneous, because of inertial effects in the reed oscillation and the fluid dynamics. Nevertheless, a first insight (and comparison to theoretical models) can be achieved by restricting the measurement of the characteristics to a case where delayed dependencies (or, equivalently, time derivatives in the mathematical description of the exciter) can be neglected.

This paper aims at measuring the relation between the pressure drop across the reed and volume flow at the double-reed output in a quasistatic case, that is, when the time variations of p and q are sufficiently small so that all time derivatives can be neglected in the nonlinear characteristic relation, and proposing a model to explain the measured relation.

B. Elementary reed model

In quasistatic conditions, a simple model can be used to describe the reed behavior [Wilson and Beavers (1974); Backus (1963)]. The reed opening area (S) is controlled by the difference between the pressure inside the reed (p_r) and the pressure inside the mouth (p_m). In the simplest model, the relation between pressure and reed opening area is considered to be linear and related through a stiffness constant (k_s),

$$(\Delta p)_r = p_m - p_r = k_S(S_0 - S). \tag{1}$$

In this formula, S_0 is the reed opening area at rest, when the pressure is the same on both sides of the reed. In most instruments (such as clarinets, oboes, or bassoons) the reed is said to be *blown-closed* (or *inward-striking*) [Helmholtz (1954)], because when the mouth pressure (p_m) is increased, the reed opening area decreases.

The role of the reed is to control and modulate the volume flow (q) entering the instrument. The Bernoulli theorem applied between the mouth and the reed duct determines the velocity of the flow inside the reed (u_r) independently of the reed opening area,

[a)]Electronic mail: andre.almeida@ircam.fr
[b)]Electronic mail: vergez@lma.cnrs-mrs.fr
[c)]Electronic mail: rene.causse@ircam.fr

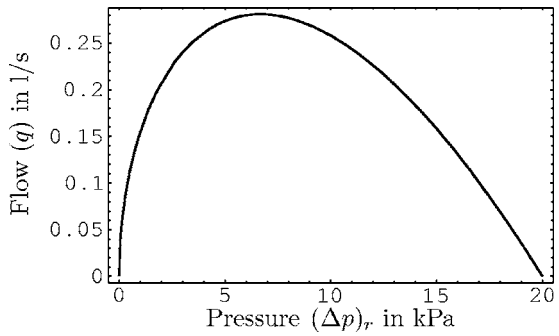


FIG. 1. A theoretical nonlinear characteristic curve for a reed of dimensions similar to an oboe reed, given by Eq. (4) using $p_M=20$ kPa and $k_s=5\times10^9$ kg m^{-3} s^{-2}.

$$p_m+\frac{1}{2}\rho u_m^2=p_r+\frac{1}{2}\rho u_r^2. \tag{2}$$

In this equation, ρ is the air density. Usually, the flow velocity u_m is neglected inside the mouth, because of volume flow conservation: inside the mouth the flow is distributed along a much wider cross section than inside the reed duct.

The volume flow (q) is the integrated flow velocity (u_r) over a cross section of the reed duct. For the sake of simplicity, the flow velocity is considered to be constant over the whole opening area, so that $q=Su_r$. Using Eq. (2), the flow is given by

$$q=S\sqrt{\frac{2(p_m-p_r)}{\rho}}. \tag{3}$$

Combining Eq. (3) and Eq. (1), it is possible to find the relation between the variables that establish the coupling with the resonator (p_r and q),

$$q=\frac{p_M-(\Delta p)_r}{k_s}\rho\sqrt{\frac{2(\Delta p)_r}{\rho}}. \tag{4}$$

The relation defined by Eq. (4) is plotted in Fig. 1, constituting what will be called in this article the elementary model for the reed.

The static reed beating pressure [Dalmont *et al.* (2005)] $p_M=k_sS_0$ (minimum pressure for which the reed channel is closed) is an alternative parameter to S_0, and can be used as a magnitude for proposing a dimensionless pressure,

$$\tilde{p}=(\Delta p)_r/p_M. \tag{5}$$

Similarly a magnitude can be found for q, leading to the definition of the dimensionless volume flow,

$$\tilde{q}=\frac{k_s}{p_M^{3/2}}\sqrt{\frac{\rho}{2}}q. \tag{6}$$

Equation (4) can then be rewritten in terms of these dimensionless quantities,

$$\tilde{q}=(1-\tilde{p})\tilde{p}^{1/2}. \tag{7}$$

This formula shows that the shape of the nonlinear characteristic curve of the elementary model is independent of the reed and blowing parameters, although the curve is scaled along the pressure p and volume flow q axis both by the stiffness k_s and the beating pressure $p_M=k_sS_0$.

C. Generalization to double reeds

For reed instruments, the quasistatic nonlinear characteristic curve has been measured in a clarinet mouthpiece [Backus (1963); Dalmont *et al.* (2003)], and the elementary mathematical model described above can explain the obtained curve remarkably well almost until the reed beating pressure (p_M).

For double-reed instruments it was not verified that the same model can be applied. In fact, there are some geometrical differences in the flow path that can considerably change the theoretical relation of Eq. (7). Local minima of the reed duct cross section may cause the separation of the flow from the walls and an additional loss of head of the flow [Wijnands and Hirschberg (1995)], and in that case the characteristics curve could change from single-valued to multivalued in a limited pressure range. This kind of change could have significant consequences on the reed oscillations.

However, the nonlinear characteristic relation was never measured before for double reeds, justifying the work that is presented below.

II. PRINCIPLES OF MEASUREMENT AND PRACTICAL ISSUES

The characteristic curve requires the synchronized measurement of two quantities: the pressure drop across the reed $(\Delta p)_r$ and the induced volume flow q.

A. Volume flow measurements

One of the main difficulties in the measurement of the reed characteristics lies in the measurement of the volume flow. There are instruments which can accurately measure the flow velocity in an isolated point (LDA, hot-wire probes) or in a region of a plane (PIV), but it can be difficult to calculate the corresponding flow by integrating the velocity field. In fact, it is difficult to do a sampling of a complete cross section of the reed because a large number of points would have to be registered. Supposing that the flow is axisymmetric at the reed output [which is confirmed by experimental results in Almeida (2006)], the measurement along a diameter of the reed would be sufficient, but regions close to the wall are inaccessible.

On the other hand, commercial flow meters usually have the disadvantage of requiring a direct reading, which would have been impractical for a complete characteristic measurement (large number of readings in a short time interval).

An indirect way of measuring the flow was then preferred to the above-mentioned methods. It consists of introducing a flow resistance in series with the reed, for which the pressure can be accurately related to the flow running through it (see Fig. 2).

The diaphragm method, used successfully by Ollivier (2002) to measure the nonlinear characteristic of single reeds, is based on this principle. The resistance is simply a perforated metal disk which covers the reed output.

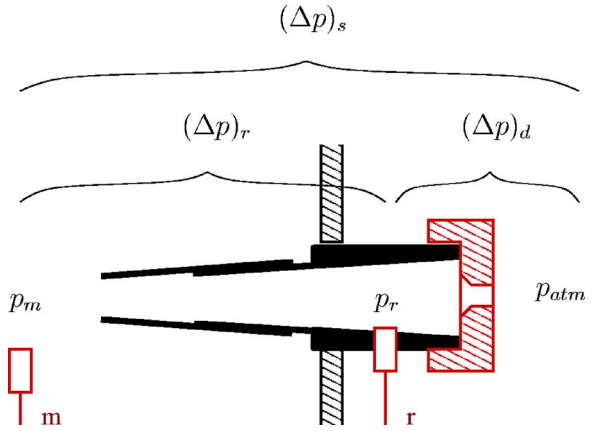


FIG. 2. (Color online) Use of a diaphragm to measure flow and pressure difference in the reed. Labeled rectangles correspond to the pressure probes used in the measurement.

For such a resistance, and assuming laminar, inviscid flow, the pressure drop $(\Delta p)_d = p_r - p_{\text{atm}}$ across the diaphragm can be approximated by the Bernoulli law, because the flow velocity at the reed output is neglected when compared to the velocity inside the diaphragm ($S_d \ll S_{\text{output}}$),

$$(\Delta p)_d = p_r - p_{\text{atm}} = \frac{1}{2}\rho\left(\frac{q}{S_d}\right)^2, \tag{8}$$

where q is the flow crossing the diaphragm, S_d the cross section of the hole, and ρ the density of air. In our experiment, pressure p_{atm} is the pressure downstream of the diaphragm (usually the atmospheric pressure, because the flow opens directly into free air). The volume flow q is then determined using a single pressure measurement p_r.

B. Practical issues and solutions

1. Issues

The realization of the characteristic measurement experiments encountered two main problems.

a. Diaphragm reduces the range of $(\Delta p)_r$ for which the measurement is possible. The addition of a resistance to the air flow circuit of the reed changes the overall nonlinear characteristic of the reed plus diaphragm system [corresponding to $(\Delta p)_s$ in Fig. 2 and to the dashed line in Fig. 3]. The solid line plots the flow against $(\Delta p)_r$, the pressure drop

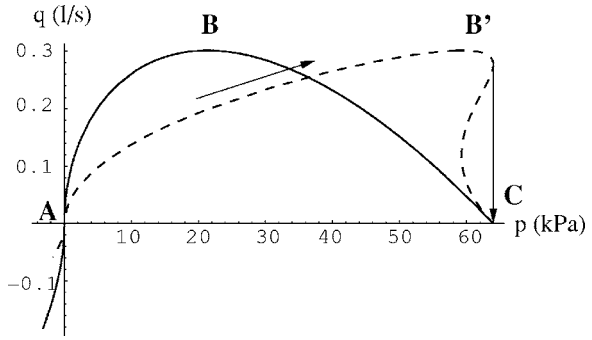


FIG. 3. Comparison of the theoretical reed characteristics (solid line) with the model of the overall characteristics of the reed associated with a diaphragm (dashed)—mathematical models, based on the Bernoulli theorem: Based on Wijnands and Hirschberg (1995).

needed to plot the nonlinear characteristics. When the resistance is increased, the maximum value of the system's characteristic is displaced towards higher pressures [Wijnands and Hirschberg (1995)], whereas the static beating pressure (p_M) value does not change [because when the reed closes there is no flow and the pressure drop in the diaphragm ($(\Delta p)_d$) is zero].

Therefore, if the diaphragm is too small (i.e., the resistance is too high), part of the decreasing region (**B'C**) of the system's characteristics becomes vertical, or even multivalued, so that there is a quick transition between two distant flow values, preventing the measurement of this part of the characteristic curve [Dalmont *et al.* (2003)] as illustrated in Fig. 3. A critical diaphragm size ($S_{d,\text{crit}} = 0.58 S_0$) can be found below which the characteristic curve becomes multivalued (see the Appendix).

b. Reed auto-oscillations. Auto-oscillations have to be prevented here to stay consistent with the quasistatic measurement (slow variations of pressure and flow). This proved to be difficult to achieve in practice. In fact, auto-oscillations become possible when the reed ceases to act as a passive resistance (a positive $\partial q/\partial p$, which absorbs energy from the standing wave inside the reed channel) to become an active supply of energy ($\partial q/\partial p < 0$). All real acoustic resonators are slightly resistive (the input admittance Y_{in} has a positive real part). This can compensate in part the negative resistance of the reed in its active region, but only below a threshold pressure, where the slope of the characteristic curve is smaller than the real part of Y_{in} for the resonator as shown by Debut and Kergomard (2004).

One way to avoid auto-oscillations is thus to increase the real part of Y_{in}, which is the acoustic resistance of the resonator. It is known that an orifice in an acoustical duct with a steady flow works as an acoustic resistance [Durrieu *et al.* (2001)], so that if the diaphragm used to measure the flow (see Sec. II A) is correctly dimensioned, the acoustic admittance seen by the reed Y_{in} can become sufficiently resistive to avoid oscillations.

2. Solutions proposed to address these issues

a. Size of the diaphragm. The volume flow is determined from the pressure drop across the diaphragm placed downstream of the reed. In practice, there is a trade-off that determines the ideal size of the diaphragm. If it is too wide, the pressure drop is too small to be measured accurately, and reed oscillations are likely to occur. If the diaphragm is too small, the system-wide characteristic can become too steep, making part of the $((\Delta p)_r)$ range inaccessible.

The ideal diaphragm cross section is then found empirically, by trying out several resistance values until one complete measurement can be done without oscillations or sudden closings of the reed. The optimal diaphragm diameter is sought using a medical flow regulator with continuously adjustable cross section as a replacement for the diaphragm.

b. Finer control of the mouth pressure p_m. During the attempts to find an optimal diaphragm, it was found that sudden closures were correlated to sudden increases in the mouth pressure. A part of the problem is that the mouth pressure depends both on the reducer setting and on the

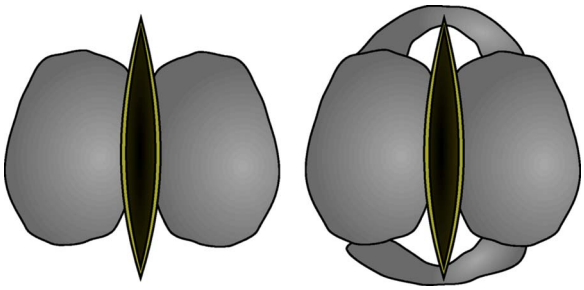


FIG. 4. (Color online) Front view of the reed (sketch) with attached masses, at left in dry conditions, at right in soaked conditions (to prevent the masses from slipping).

downstream resistance. By introducing a leak upstream of the experimental apparatus (thus not altering the experiment), it is possible to improve measurements in the decreasing region of the characteristic (**BC**), at least when the system-wide characteristic is not multivalued (see Sec. II B 1 a).

c. Increase the reed mass. One other way to reduce the oscillations is thus to prevent the appearance of instabilities, or to reduce their effects. An increase in the reed damping would certainly be a good method to avoid oscillations, because it cancels out the active role of the reed (which can be seen as a negative damping), Debut (2004).

It is difficult to increase the damping of the reed without altering its opening or stiffness properties. The simplest way found to prevent reed oscillations was thus an increase in the reed mass.

This mass increase was implemented by attaching small masses of *Blu-Tack* (a plastic sticking material usually used to stick paper to a wall) to one or both blades of the reed (Fig. 4). During measurements on previously soaked reeds it was difficult to keep the masses attached to the reed, so that an additional portion of *Blu-Tack* is used to connect the two masses together, wrapping around the reed. This wrapping is not expected to have a great effect on the measured elastic properties, because it does not pull the masses together; its objective is to avoid the main masses from falling due to the effect of gravity. A comparison of the results using different masses showed that their effect on the quasistatic characteristics can be neglected (effects are weaker than variations for experiments in the same reed) [Almeida (2006)].

C. Experimental setup and calibrations

The experimental device is shown in Fig. 5. An artificial mouth [Almeida *et al.* (2004)] was used as a blowing mechanism and support for the reed. The window in front of the reed allows the capture of frontal pictures of the reed opening. Artificial lips, allowing adjustment of the initial opening area of the reed, were not used here, to avoid modifications in some of the elastic properties of the reed, possibly in a different way from what happens with real lips.

As stated before, the plot of the characteristic curve requires two coordinated measurements: the pressure difference $(\Delta p)_r$ across the reed and the induced volume flow q, determined from the presure drop $(\Delta p)_d$ across a calibrated diaphragm (Sec. II A).

In practice thus, the experiment requires two pressure measurements p_m and p_r, as shown in Fig. 5.

1. Pressure measurements

The pressure is measured in the mouth and in the reed using Honeywell SCX series, silicon-membrane differential pressure sensors whose range is from −50 to 50 kPa.

These sensors are not mounted directly on the measurement points, but one of the terminals in each sensor is connected to the measurement point using a short flexible tube (about 20 cm in length). Therefore, one tube opens in the inside wall of the artificial mouth, 4 cm upstream from the reed, and the other tube crosses the rubber socket attaching the diaphragm to the reed output. The use of these tubes does not influence the measured pressures as long as their variations are slow.

The signal from these sensors is amplified before entering the digital acquisition card. The gain is adjusted for each type of reed. The system consisting of the sensor connected to the amplifier is calibrated as a whole in order to find the voltage at the amplifier output corresponding to each pressure difference in the probe terminals: the stable pressure drop applied to the probe is also measured using a digital manometer connected to the same volumes, and compared to the probe tension read using a digital voltmeter. Voltage is found to vary linearly with the applied pressure within the measuring range of the sensor.

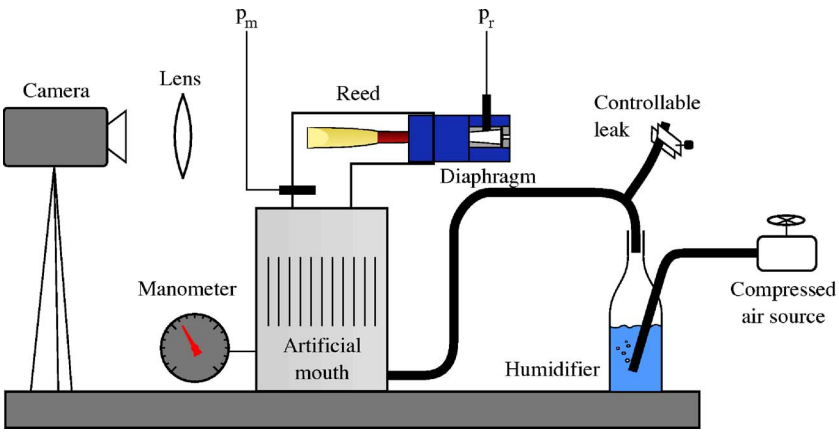


FIG. 5. (Color online) Device used for characteristics measurements.

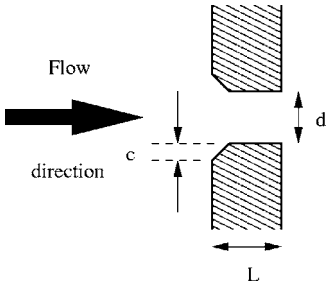


FIG. 6. Detail of the diaphragm dimensions.

2. Diaphragm calibration

The curve relating volume flow q to the pressure difference through diaphragms $(\Delta_p)_d$ can be approximated by the Bernoulli theorem. In fact, diaphragms are constructed so as to minimize friction effects (by reducing the length of the diaphragm channel) and jet contraction—the upstream edges are smoothed by chamfering at 45° (Fig. 6). The chamfer height (c) is approximately 0.5 mm. The diaphragm channel is 3 mm long (L).

Nevertheless, this ideal characteristics was checked in stationary conditions for each diaphragm (see Fig. 7) using a gas volume meter, which would not be usable for variable volume flows. It was found that the effective cross section is slightly smaller than the actual cross section (about 10%), which is probably due to some *Vena Contracta* effect in the entrance of the diaphragm. Moreover, above a given pressure drop the volume flow increase is lower than what is predicted by Bernoulli's theorem (corresponding to a lower exponent than 1/2 predicted by Bernoulli). This difference is probably due to turbulence generated for high Reynolds numbers. In Fig. 7 the dashed line corresponding to the critical value of the Reynolds number ($\mathrm{Re}_c = ud/\nu = 2000$) is shown. It is calculated using the following formulas for u (the average flow velocity in the diaphragm) and d (the diaphragm diameter):

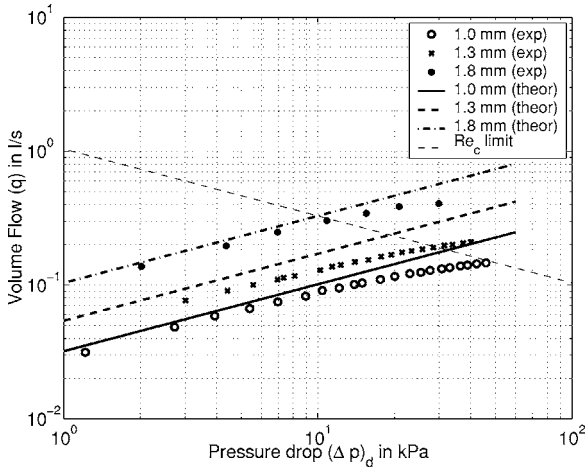


FIG. 7. Calibration of diaphragms used in characteristic measurements (dots are experimental data and lines are Bernoulli predictions using the measured diaphragm diameters). The dashed black line represents the pressure/flow relation corresponding to the expected transition between laminar and turbulent flows ($\mathrm{Re}_c = 2000$), parametrized by the diaphragm diameter d.

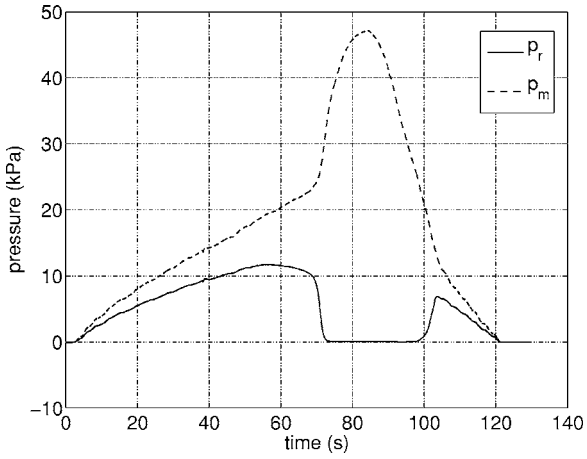


FIG. 8. Time variation of the mouth pressure (p_m) and the pressure inside the reed (p_r) during a successful characteristics measurement.

$$d = \left(\frac{2}{\pi}\frac{q}{u}\right)^{1/2} \tag{9}$$

$$u = \left(\frac{2(\Delta p)_d}{\rho}\right)^{1/2}, \tag{10}$$

so that the constant Reynolds relation is given by

$$Q^{1/2}\Delta p^{1/4} = \mathrm{Re}_c \nu \left(\frac{\pi}{2}\right)^{1/2}\left(\frac{\rho}{2}\right)^{1/4}, \tag{11}$$

where the right-hand side should be a constant based on the diaphragm geometry.

Since a suitable model was not found for the data displayed in Fig. 7, we chose to interpolate the experimental calibrations in order to find the flow corresponding to each pressure drop in the diaphragm. Linear interpolation was used in the (p, q^2) space.

3. Typical run

In a typical run, the mouth pressure p_m is balanced with the atmospheric pressure in the room at the beginning of the experiment. Both p_m and p_r are recorded in the computer through a digital acquisition device at a sampling rate of 4000 Hz. Pressure p_m is increased until slightly above the pressure at which the reed closes, left for some seconds above this value, and then decreased back to the atmospheric pressure. The whole procedure lasts for about 3 min, and is depicted in Fig. 8.

D. Double reeds used in this study and operating conditions

Among the great variety of double reeds that are used in musical instruments, we chose as a first target for these measurements a natural cane oboe reed fabricated using standard procedures (by *Glotin*), and sold to the oboist (usually a beginner oboist) as a final product (i.e., ready to be played).

The choice of a ready-to-use cane reed was mainly retained because it can be considered as an average reed. This avoids considering a particular scraping technique among many used by musicians and reed makers. Of course, this

does not greatly facilitate the task of the reed measurement, because natural reeds are very sensitive to environment conditions, age, or time of usage.

Other reeds were also tested, as a term of comparison with the natural reeds used in most of the experiments. However, none of these reeds was produced by a professional oboist or reed maker, although it would be an interesting project to investigate the variations in reeds produced by different professionals.

To conclude, the results presented in the next section may depend to a certain extent on the reed chosen for the experiments, and a larger sample of reeds embracing the big diversity of scraping techniques needs to be tested before claiming for the generality of the results that will be presented.

Another remark has to be made on the conditions during the experiments. The kind of reeds used in most experiments are always blown with highly moisturized air. In fact, in real life, reeds are often soaked before they are used, and constantly maintained wet by saliva and water vapor condensation. These conditions were sought throughout most of the experiments, although the sensitivity of the reed to environmental conditions was also investigated. For instance, the added masses were found to have no practical influence on the nonlinear characteristics, whereas the humidity increases the hysteresis in the complete measurement cycle (increasing followed by decreasing pressures), while reducing the reed opening at rest [Almeida (2006)].

In our measurements, humidification is achieved by letting the air flow through a plastic bottle half-filled with hot water at 40° (see Fig. 5), recovering it from the top. Air arriving in the artificial mouth has a lower temperature, because its temperature is approximately 10° when entering the bottle. This causes the temperature and humidity to decrease gradually along the experiments. Future measurements should include a thermostat for the water temperature in order to ensure stable humidification.

III. RESULTS AND DISCUSSION

A. Typical pressure vs flow characteristics

Using the formula of Eq. (8), and the calibrations carried out for the diaphragm used in the measurement, the volume flow (q) is determined from the pressure inside the reed (p_r). The pressure drop in the reed corresponds to the difference between the mouth and reed pressures ($(\Delta p)_r = p_m - p_r$). Volume flow is then plotted against the pressure difference ($(\Delta p)_r$), yielding a curve shown in Fig. 9.

In this figure, the flow is seen to increase until a certain maximum value ($(\Delta p)_r \simeq 6$ kPa). When the pressure is increased further, flow decreases due to the closing of the reed. Instead of completely vanishing for $(\Delta p)_r = p_M$, as predicted by the elementary model shown in Sec. I B, the volume flow first stabilizes at a certain minimum value and then slightly increases when the pressure is increased further, indicating that it is very hard to completely close the reed.

The flow remaining after the two blades are in contact suggests that, despite the closed appearance of the double reed, some narrow channels remaining between the two

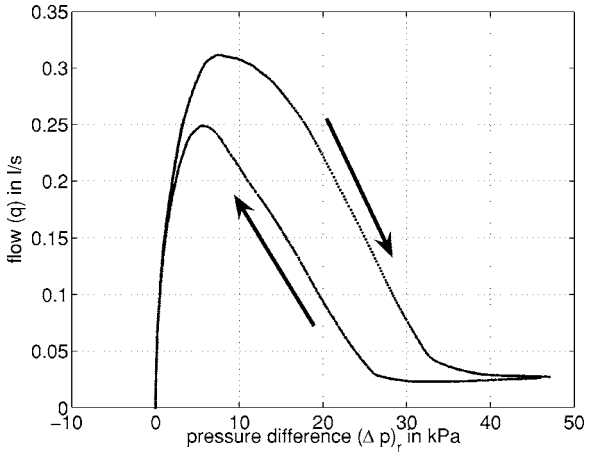


FIG. 9. A typical result for the measurement of the volume flow vs pressure characteristic of a natural cane oboe reed.

blades are impossible to close, behaving like rigid capillary ducts, which is corroborated by the slight increase in the residual flow for high pressures. Since the logarithmic plot of the nonlinear characteristics (Fig. 10) shows a 1/2 power dependence on the residual volume flow, this suggests that the residual flow is controlled by inertia rather than viscosity.

When reducing the pressure back to zero, the reed follows a different path in the p/q space than the path for increasing pressures. This hysteresis is due to memory effects of the reed material which have been investigated experimentally for single-reed [Dalmont *et al.* (2003)] and double-reed instruments [Almeida *et al.* (2006)].

B. Comparison with other instruments

1. Bassoon

Since oboes are not the only double-reed instruments, it is interesting to compare the nonlinear characteristic curves from different instruments. The bassoon is also played using a double reed, but its dimensions are different: its opening

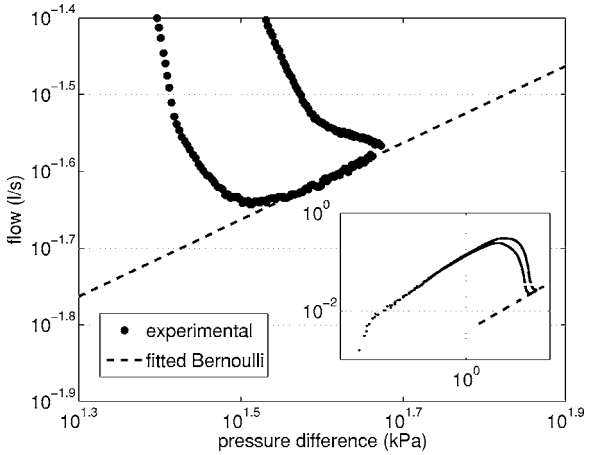


FIG. 10. Double-logarithmic plot of the characteristic curve of Fig. 9 to show the 1/2 power dependence when the reed is almost shut. Inset shows the whole range of data, from which the part corresponding to the closed reed is magnified in the main graph.

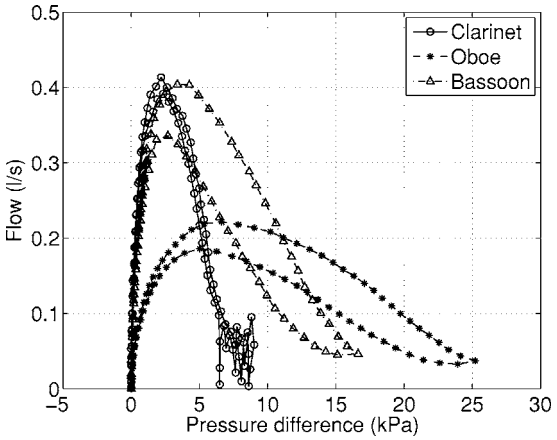


FIG. 11. Comparison of the characteristic curves of different reed exciters for different instruments. Clarinet data were obtained by Dalmont *et al.* (2003) for a PlastiCover® reed. Oboe and bassoon reeds are blown using moisturized air.

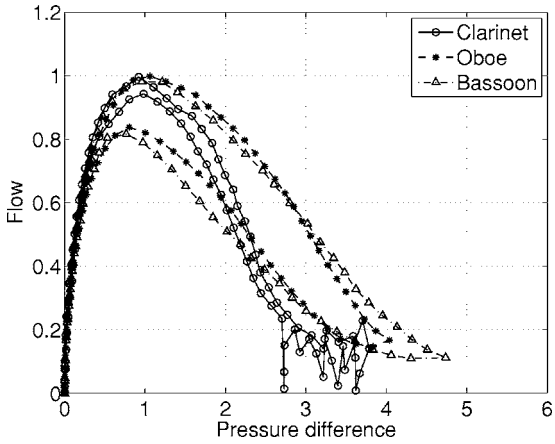


FIG. 12. Data from Fig. 11, normalized along q by the maximum flow for increasing pressures, and along p by the corresponding pressure.

area at rest is typically 7 mm^2 (against around 2 mm^2 for the oboe) and the cross-section profile varies slightly from oboe reeds.

Figure 11 compares two characteristic curves for natural cane oboe and bassoon reeds. Both were measured under similar experimental conditions, as far as possible. The reed was introduced dry in the artificial mouth, but the supplied air is moisturized at nearly 100% humidity, and masses were added to both reeds to prevent auto-oscillations. The diaphragms used in each measurement are different, however, and this is because the opening area of the reed at rest is much larger in the case of the bassoon, so that a smaller resistance (a larger diaphragm) is needed to avoid the reed closing suddenly in the decreasing side of the characteristic curve [see Sec. II C and Eq. (A10)]. This should not have any consequences in the measured characteristic curve.

In the q axis, the bassoon reed reaches higher values, and this is probably a consequence of its larger opening area at rest, although the surface stiffness is likely to change as well from the oboe to the bassoon reed. In the p axis, the bassoon reed extends over a smaller range of pressures so that the reed beating pressure is about 17 kPa in the case of the bassoon reed, whereas it is near 33 kPa for the oboe reed.

Apart from these scaling considerations, the shapes of the curves are similar and this can be better observed if flow and pressure are normalized using the maximum flow point of each curve (Fig. 12).

2. Clarinet

The excitation mechanism of clarinets and saxophones share the same principle of functioning with double reeds. However, there are several geometric and mechanical differences between single reeds and double reeds. For instance, flow in a clarinet mouthpiece encounters an abrupt expansion after the first 2 or 3 mm of the channel between the reed and the rigid mouthpiece, and the single reed is subject to fewer mechanical constraints than any double reed. These differences suggest that the characteristic curve of single-reed instruments might present some qualitative differences with respect to the double reed [Vergez *et al.* (2003).]

The nonlinear characteristic curve of clarinet mouthpieces displayed in Figs. 11 and 12 was measured by Dalmont *et al.* (2003) using similar methods as the ones we used for the double reed. A comparison between the curves for both kinds of exciters (in Fig. 11) shows that the overall behavior of the excitation mechanism is similar in both cases. Similarly to when comparing oboe to bassoon reeds, the scalings of the characteristic curves of single reeds are different from those of oboe reeds, although closer to those of the bassoon. This is probably a question of the dimensions of the opening area.

A different issue is the relation between reference pressure values in the curve (shown in the adimensionalized representation of Fig. 12). As predicted by the elementary model described in Sec. I B, in the single reed the pressure at maximum flow is about 1/3 of the beating pressure of the reed, whereas in double-reed measurements, the relation seems to be closer to 1/4. This deviation from the model is shown in Sec. IV to be linked with the diffuser effect of the conical staple in double reeds.

Figure 12 also shows that in the clarinet mouthpiece used by Dalmont *et al.* (2003) the hysteresis is relatively less important than in both kinds of double reeds. In fact, whereas the measurements for double reeds were performed in wet conditions, the PlastiCover® reed used for the clarinet was especially chosen because of its smaller sensitivity to environment conditions.

IV. ANALYSIS

A. Comparison with the elementary model

The measured nonlinear characteristic curve of Fig. 9 can be compared to the model described in Sec. I B. In this model, two parameters (k_s and S_0) control the scaling of the curve along the p and q axis. They are used to adjust two key points in the theoretical curve to the experimental one: the reed beating pressure p_M and the maximum volume flow $q_{\max}$.

Once $q_{\max}$ is determined through a direct reading, the stiffness k_s is calculated using the following relation:

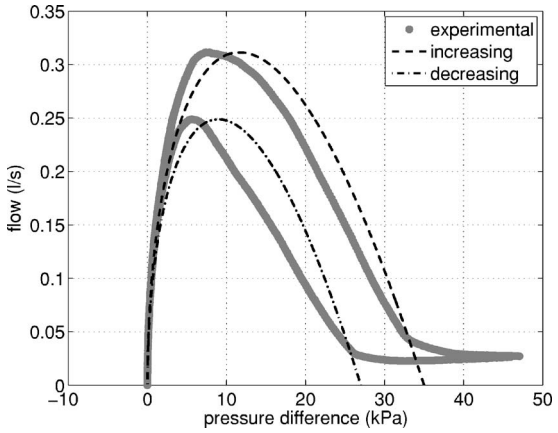


FIG. 13. Comparison of the experimental nonlinear characteristics curve with the elementary model shown in Fig. 1. Two models are fitted, for increasing (p_M=35 kPa, k_s=1.04×10^{10} kg m^{-3} s^{-2}) and decreasing (p_M=27 kPa, ks=8.86×10^9 kg m^{-3} s^{-2}) mouth pressures.

$$k_s = q_{\max}^{-1}\left(\frac{2}{3}p_M\right)^{3/2}\rho^{-1/2}. \tag{12}$$

This allows adjustment of a theoretical characteristic curve [corresponding to the elementary model of Eq. (4)] to each of the branches of the measured characteristic curve, for increasing and decreasing pressures (Fig. 13).

When compared to the elementary model of Sec. I B, the characteristic curve associated with double reeds shows a deviation of the pressure at which the flow reaches its maximum value. In fact, it can be easily shown that for the elementary model this value is 1/3 of the reed beating pressure p_M, which is also verified in the clarinet (Sec. III B 2). In the measured curves however, this value is usually situated between $1/4p_M$ and $1/5p_M$. Nevertheless, the shapes of the curves are qualitatively similar to the theoretical ones.

B. Conical diffuser

The former observations about the displacement of the maximum value can be analyzed in terms of the pressure recoveries due to flow decelerations inside the reed duct. Variations in the flow velocity are induced by the increasing cross section of the reed towards the reed output (Fig. 14). This can be understood simply by considering energy and mass conservation between two different sections of the reed,

$$p_{\text{in}} + \frac{1}{2}\rho\left(\frac{q}{S_{\text{in}}}\right)^2 = p_{\text{out}} + \frac{1}{2}\rho\left(\frac{q}{S_{\text{out}}}\right)^2, \tag{13}$$

where q is the total volume flow that can be calculated either at the input or the output of the conical diffuser by integrating the flow velocity over the cross section S_{in} or S_{out}, respectively.

In practice, however, energy is not expected to be completely conserved along the flow because of its turbulent nature. In fact, for instance at the reed output (diameter d), the Reynolds number of the flow [Re=ud/ν=4($q/\pi d\nu$)] can be estimated using data from Fig. 9 to reach a maximum value of 5000. Given that this number is inversely proportional to

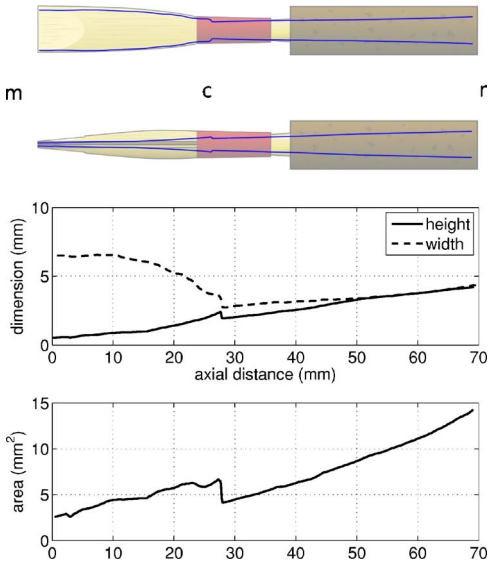


FIG. 14. (Color online) Cross-section profiles (axis and area) of an oboe reed, measured on a mold of the reed channel, and indexes used in Sec. IV B: **m**outh, **c**onstriction, and **r**eed output.

the diameter of the duct d, the Reynolds number increases upstream, inside the reed duct, so that the flow is expected to be turbulent also for lower volume flows.

For turbulent flows, no theoretical model can be applied to calculate the pressure recovery due to the tapering of the reed duct. However, phenomenological models are available in engineering literature, where similar duct geometries are known as "conical diffusers." Unlike in clarinet mouthpieces, where the sudden expansion of the profile is likely to cause a turbulent mixing without pressure recovery [Hirschberg (1995)], this effect must be considered in conical diffusers. The pressure recovery is usually quantified in terms of a *recovery coefficient* C_p stating the relation between the pressure difference between both ends of the diffuser and the ideal pressure recovery which would be achieved if the flow was stopped without losses,

$$C_P = \frac{p_{\text{out}} - p_{\text{in}}}{\frac{1}{2}\rho u_{\text{in}}^2}. \tag{14}$$

C_P values range from 0 (no recovery) to 1 (complete recovery, never achieved in practice).

According to Eq. (14), pressure recovery is proportional to the square of the flow velocity at the entrance of the conical diffuser, and consequently to the squared volume flow inside the reed. The overall pressure difference across the reed ($p_m - p_{\text{out}}$) is deduced from the corresponding pressure difference without pressure recovery ($p_m - p_{\text{in}}$) according to the formula

$$(p_m - p_{\text{out}}) = (p_m - p_{\text{in}}) - \alpha q^2, \tag{15}$$

where $\alpha = \frac{1}{2}\rho(C_p/S_{\text{in}}^2)$ is a constant. This explains why the curve $q=f(p_m - p_{\text{out}})$ in Fig. 13 is more shifted to the left compared to the curve $q=f(p_m - p_{\text{in}})$ at the top, where q is higher.

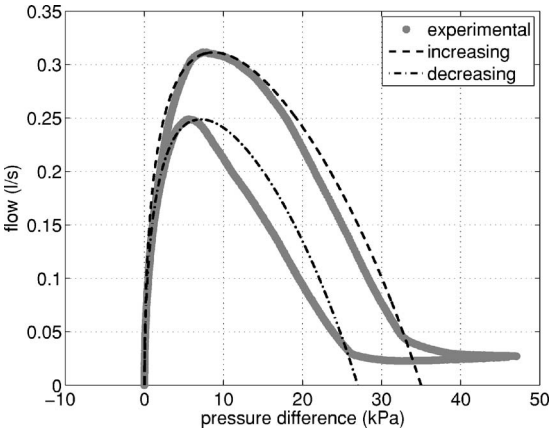


FIG. 15. Comparison of the experimental nonlinear characteristics curve with a reed model with pressure recovery in the final part of the duct. Fitted k_s and S_0 are the same as in Fig. 13 and different for increasing and decreasing mouth pressures. In both cases the value $C_p=0.8$ was used.

1. Reed model with pressure recovery

In order to take into account the pressure recovery before the reed output, the flow is divided into two sections, the upstream, until the constriction at 28 mm (index c in Fig. 14) and the conical diffuser part from the constriction until the reed output. In the upstream section, no pressure recovery is considered, so that the flow velocity can be calculated using the pressure difference between the mouth and this point using a Bernoulli model, as in Eq. (3), but replacing p_r with p_c,

$$q = S\sqrt{\frac{2(p_m - p_c)}{\rho}}. \tag{16}$$

Similarly, the reed opening is calculated using the same pressure difference,

$$(\Delta p)_c = p_m - p_c = k_S(S_0 - S). \tag{17}$$

The total pressure difference used to plot the characteristic curve, however, is different, because the recovered pressure has to be added to $(\Delta p)_c$,

$$p_m - p_r = (\Delta p)_c - C_p\frac{1}{2}\rho\left(\frac{q}{S_c}\right)^2, \tag{18}$$

where S_c is the reed duct cross section at the diffuser input, i.e., at the constriction, which is found from Fig. 14, $S_c=4\times10^{-6}$ m^2.

Using these equations, the modified model can be fitted to the experimental data. In Fig. 15, the same parameters k_s and S_0 were used as in Fig. 13, leaving only C_p as a free parameter for the fitting. Figure 15 was obtained for a value of $C_p=0.8$. This value can be compared to typical values of pressure recovery coefficients found in industrial machines [Azad (1996)].

In engineering literature, C_P is found to depend mostly on the ratio between output and input cross sections ($AR=S_{\text{out}}/S_{\text{in}}$) and the diffuser length to initial diameter ratio (L/d_{in}) [White (2001)]. The tapering angle θ influences the growth of the boundary layers, so that above a critical angle ($\theta=8°$) the flow is known to detach from the diffuser walls, considerably lowering the recovered pressure. An in-depth study of turbulent flow in conical diffusers can be found in the literature [Azad (1996)], usually for diffusers with much larger dimensions than the ones found in the double reed.

The geometry of the conical diffuser studied in Azad (1996) can be compared to the one studied in our work: the cross section is circular and the tapering angle $\theta=3.94°$ is not very far from the tapering angle of the reed staple $\theta=5.2°$ [in particular, both are situated in regions of similar flow regimes, Kilne and Abbott (1962), as a function of the already mentioned AR and L/d_{in}]. Reynolds numbers of his flows ($\text{Re}=6.9\times10^4$) are also close to the maximum ones found at the staple input ($\text{Re}\simeq10^4$). The length to input diameter ratio of the reed staple $L/d=20$ is bigger than that found in Azad (1996); however, the pressure recovery coefficient can be extrapolated from his data to find the value $C_P\simeq0.8$ [Fig. 2 in Azad (1996)], or a slightly smaller value of $C_P\simeq0.7$ based on Fig. 6.28b in White (2001).

V. CONCLUSION

The quasistatic nonlinear characteristics were measured for double reeds using a similar device as the one used for single-reed mouthpieces by Dalmont *et al.* (2003). The obtained curves are close to the ones found for single reeds, and in particular no evidence of multivalued flows for a same pressure was found, as was suggested by theoretical considerations made by Wijnands and Hirschberg (1995) or Vergez *et al.* (2003).

However, double-reed characteristic curves present substantial quantitative differences for high volume flows when compared to elementary models for the reed. These differences can be explained using a model of pressure recovery in the conical staple, proportional to the square of the input flow velocity.

In Vergez *et al.* (2003) a similar model had already been considered with the pressure difference between the jet and the output of the reed depending on the square of the volume flow q [Eq. 15, p. 969 in Vergez *et al.* (2003)],

$$p_m = p_j + \frac{1}{2}\rho v_j^2 \tag{19}$$

$$p_j = p_r + \frac{1}{2}\rho\frac{\Psi^2}{S_{ra}}q^2, \tag{20}$$

with S_{ra} the cross section of the double reed where the jet reattaches. However, through theoretical considerations, the typical values Ψ were estimated to be positive (would correspond to a negative C_p). The result was a nonlinear characteristic $q=f(p_m-p_{\text{out}})$ increasingly shifted to the right compared to the curve $q=f(p_m-p_{\text{in}})$ as q increases.

Based on the experimental results presented in the present paper, it is now possible to explain why the underlying conjectures were wrong. In fact, our estimation was based on a jet contraction factor $\alpha=0.8$, whereas our recent experiments have revealed that no jet contraction occurs [Almeida (2006)]. Consequently, in Eq. 16 of Vergez *et al.* (2003) head losses were the most important terms leading to

a positive Ψ. On the other hand imposing $\alpha=1$ would mean $S_j=S_{ra}$ in Eq. 16 of Vergez *et al.* (2003), leading to $\Psi=-1+(S_{ra}/S_r)^2+\Psi_{\text{losses}}$. Given that $(S_{ra}/S_r)^2\simeq 4\times 10^{-2}$, a negative Ψ, i.e., a positive C_p, is expected.

It is worth noting that direct application of the measured characteristic curves to the modeling of the complete oboe is not that obvious. Indeed, the assumption that the mouthpiece as a whole (reed plus staple) can be modeled as a nonlinear element with characteristics given by the above experiment would be valid if the size of the mouthpiece was negligible with respect to a typical wavelength. Given the 7 cm of the mouthpiece and the several tens of centimeters of a typical wavelength, this is questionable. Should the staple be considered as part of the resonator? In that case, the separation between the exciter and the resonator would reveal a longer resonator and an exciter without pressure recovery. This could be investigated through numerical simulations by introducing the pressure recovery coefficient (C_P) as a free parameter.

Moreover, the underlying assumption of the models is that nonstationary effects are negligible (all flow models are quasistatic). Some clues indicate that this could also be put into question.

(i) First of all, experimental observations of the flow at the output of the staple [through hot-wire measurements, Almeida (2006)] revealed significant differences in the flow patterns when considering static and auto-oscillating reeds.

(ii) Moreover, nondimensional analysis revealed Strouhal numbers much larger than for simple reed instruments [Vergez *et al.* (2003), Almeida (2006)].

ACKNOWLEDGMENTS

The authors would like to thank J.-P. Dalmont and J. Gilbert for experimental data on the clarinet reed characteristics and fruitful discussions and suggestions about the experiments and analysis of data, and A. Terrier and G. Bertrand for technical support.

APPENDIX: CALCULATION OF THE MINIMUM DIAPHRAGM CROSS SECTION

The total pressure drop in the reed-diaphragm system (Fig. 2) is

$$(\Delta p)_s=(\Delta p)_r+(\Delta p)_d. \tag{A1}$$

The system's characteristics become multivalued when there is at least one point on the curve where the slope is infinite,

$$\frac{\partial}{\partial q}(\Delta p)_s=\frac{\partial}{\partial q}(\Delta p)_r+\frac{\partial}{\partial q}(\Delta p)_d=0. \tag{A2}$$

Because of simplicity, the derivatives in Eq. (A2) are replaced by their inverse,

$$\frac{\partial}{\partial q}(\Delta p)_s=\left(\frac{\partial q}{\partial(\Delta p)_r}\right)^{-1}+\left(\frac{\partial q}{\partial(\Delta p)_d}\right)^{-1}=0, \tag{A3}$$

yielding

$$\left(\frac{S}{\rho}\left(\frac{2(\Delta p)_r}{\rho}\right)^{-1/2}-\frac{1}{k_s}\left(\frac{2(\Delta p)_r}{\rho}\right)^{1/2}\right)^{-1}+\frac{\rho}{S_d}\left(\frac{2(\Delta p)_d}{\rho}\right)^{1/2}=0. \tag{A4}$$

Solving for S_d,

$$S_d=-\left(\frac{2(\Delta p)_d}{\rho}\right)^{1/2}\times\left(\frac{S}{\rho}\left(\frac{2(\Delta p)_r}{\rho}\right)^{-1/2}-\frac{1}{k_s}\left(\frac{2(\Delta p)_r}{\rho}\right)^{1/2}\right). \tag{A5}$$

Simplifying,

$$S_d=-S\left(\frac{(\Delta p)_d}{(\Delta p)_r}\right)^{1/2}+\frac{2}{k_s}((\Delta p)_d(\Delta p)_r)^{1/2}. \tag{A6}$$

From Eqs. (3) and (8), we can find

$$(\Delta p)_r=\left(\frac{S_d}{S}\right)^2(\Delta p)_d, \tag{A7}$$

and Eq. (A6) can be written

$$S_d=-S\frac{S}{S_d}+\frac{2}{k_s}\frac{S}{S_d}(\Delta p)_r. \tag{A8}$$

Now we can replace $S=S_0-[(\Delta p)_r/k_s]$ to find

$$S_d^2=\left(-S_0+3\frac{(\Delta p)_r}{k_s}\right)\left(S_0-\frac{(\Delta p)_r}{k_s}\right). \tag{A9}$$

It is clear that the right-hand side of this equation must be positive. Moreover, it is a parabolic function of $(\Delta S)=(\Delta p)_r/k_s$, with its concavity facing downwards.

The maximum value of $S_d^2(S)$,

$$\max(S_d^2(S))=\frac{S_0^2}{3}, \tag{A10}$$

is thus the value for which there is only a single point where the characteristic curve has an infinite slope.

We thus conclude that $S_d=S_0/\sqrt{3}=0.58S_0$ is the minimum value of the diaphragm cross section that should be used for flow measurements.

Almeida, A. (**2006**). "Physics of double-reeds and applications to sound synthesis," Ph.D. thesis, Univ. Paris VI.

Almeida, A., Vergez, C., and Caussé, R. (**2004**). "Experimental investigations on double reed quasi-static behavior," in *Proceedings of ICA 2004*, Vol. **II**, pp. 1229–1232.

Almeida, A., Vergez, C., and Caussé, R. (**2006**). "Experimental investigation of reed instrument functioning through image analysis of reed opening," Acustica (accepted).

Azad, R. S. (**1996**). "Turbulent flow in a conical diffuser: a review," Exp. Therm. Fluid Sci. **13**, 318–337.

Backus, J. (**1963**). "Small-vibration theory of the clarinet," J. Acoust. Soc. Am. **35**(3), 305–313.

Dalmont, J. P., Gilbert, J., and Ollivier, S. (**2003**). "Nonlinear characteristics of single-reed instruments: Quasi-static volume flow and reed opening measurements," J. Acoust. Soc. Am. **114**(4), 2253–2262.

Dalmont, J.-P., Gilbert, J., Kergomard, J., and Ollivier, S. (**2005**). "An analytical prediction of the oscillation and extinction thresholds of a clarinet," J. Acoust. Soc. Am. **118**(5), 3294–3305.

Debut, V. (**2004**). *Deux études d'un instrument de musique de type clarinette: Analyse des fréquences propres du résonateur et calcul des auto-oscillations par décomposition modale* (*Two studies of a clarinet-like musical instrument: Analysis of the eigen frequencies and calculation of the self-sustained oscillations by modal decomposition*). Ph.D. thesis, Université de la Mediterranée Aix Marseille II.

Debut, V., and Kergomard, J. (**2004**). "Analysis of the self-sustained oscillations of a clarinet as a Van-der-pol oscillator," in *Proceedings of ICA 2004*, Vol. II, pp. 1425–1428.
Durrieu, P., Hofmans, G., Ajello, G., Boot, R., Aurégan, Y., Hirschberg, A., and Peters, M. C. A. M. (**2001**). "Quasisteady aero-acoustic response of orifices," J. Acoust. Soc. Am. **110**(4), 1859–1872.
Fletcher, N. H., and Rossing, T. D. (**1998**). *The Physics of Musical Instruments*, 2nd ed. (Springer, Berlin).
Gilbert, J., Dalmont, J.-P., and Guimezanes, T. (**2005**). "Nonlinear propagation in wood-winds," in *Forum Acusticum 2005*, pp. 1369–1372.
Helmholtz, H. (**1954**). *On the Sensations of Tone as a Physiological Basis for the Theory of Music* (Dover, New York) (English translation by Alexander Ellis).
Hirschberg, A. (**1995**). *Mechanics of Musical Instruments* (Springer, Berlin), Chap. 7, pp. 229–290.
Kilne, S. J., and Abbott, D. E. (**1962**). "Flow regimes in curved subsonic diffusers," J. Basic Eng. **84**, 303–312.
Ollivier, S. (**2002**). *Contribution á l'étude des Oscillations des Instruments à Vent à Anche Simple* ("Contribution to the study of oscillations in single-reed wind instruments)." Ph.D. thesis, Université du Maine, Laboratoire d'Acoustique de l'Université du Maine–UMR CNRS 6613.
Vergez, C., Almeida, A., Causse, R., and Rodet, X. (**2003**). "Toward a simple physical model of double-reed musical instruments: Influence of aero-dynamical losses in the embouchure on the coupling between the reed and the bore of the resonator," Acta. Acust. Acust. **89**, 964–973.
White, F. M. (**2001**). *Fluid Mechanics*, 4th ed. (McGraw-Hill, New York).
Wijnands, A. P. J., and Hirschberg, A. (**1995**). "Effect of a pipe neck downstream of a double reed," in *Proceedings of the International Symposium on Musical Acoustics*, pp. 149–152. Societe Française d'Acoustique.
Wilson, T. A., and Beavers, G. S. (**1974**). "Operating modes of the clarinet," J. Acoust. Soc. Am. **56**(2), 653–658.

Vocal tract resonances and the sound of the Australian didjeridu (yidaki). III. Determinants of playing quality

John Smith,[a] Guillaume Rey, and Paul Dickens
School of Physics, University of New South Wales, Sydney NSW 2052, Australia

Neville Fletcher
School of Physics, University of New South Wales, Sydney NSW 2052, Australia
and R.S.Phys.S.E., Australian National University, Canberra, ACT 0200, Australia

Lloyd Hollenberg
School of Physics, Melbourne University, Melbourne, Vic 3010, Australia

Joe Wolfe
School of Physics, University of New South Wales, Sydney NSW 2052, Australia

(Received 24 July 2006; revised 4 October 2006; accepted 9 October 2006)

Traditional didjeridus have a broad range of bore geometries with many details not immediately apparent to a player, and are therefore suitable for examining the relationship between perceived quality and physical properties. Seven experienced players assessed the overall playing quality of 38 didjeridus that spanned a wide range of quality, pitch, and geometry, as well as 11 plastic cylindrical pipes. The ranking of these instruments was correlated with detailed measurements of their acoustic input impedance spectra. Most significantly, the ranked quality of a didjeridu was found to be negatively correlated with the magnitude of its acoustic input impedance, particularly in the frequency range from 1 to 2 kHz. This is in accord with the fact that maxima in the impedance of the player's vocal tract can inhibit acoustic flow, and consequently sound production, once the magnitude of these impedance maxima becomes comparable with or greater than those of the instrument. This produces the varying spectral peaks or formants in the sound envelope that characterize this instrument. Thus an instrument with low impedance and relatively weak impedance maxima in this frequency range would allow players greater control of the formants in the output sound and thus lead to a higher perceived playing quality.
[DOI: 10.1121/1.2384849]

PACS number(s): 43.75.Fg [DD] Pages: 547–558

I. INTRODUCTION

The didjeridu (or didgeridoo) is an unusual and ancient musical instrument originally played in parts of Northern Australia, where the Yolngu people call it the yidaki or yiraki. In didjeridu playing, the configuration of the player's vocal tract has a spectacular effect on the timbre, for reasons that have been explained elsewhere (Tarnopolsky *et al.*, 2005, 2006; Fletcher *et al.*, 2006). In performance, the player (usually a man) places his lips at the narrower end of the instrument and uses the technique of "circular breathing" that allows him to play without stopping. This involves filling his cheeks with air, then inhaling quickly through the nose (and bypassing the mouth at the soft palate) to refill the lungs with air. The contrast between the sounds produced during inhalatory and exhalatory play is often used to establish a rhythm, which is believed to give the instrument its onomatopoeic Western name. Unusual among wind instruments, it usually only plays a single note at the frequency of its lowest resonance, although overblowing to produce notes at the second or very occasionally higher resonances is sometimes used to produce musical accents. Apart from the interesting rhythm, the chief musical interest in performance thus comes not from pitch variation, but from the contrasting and varying timbres that can be produced by either deliberate movement of the player's tongue and/or by movements associated with "circular breathing." There are also vocalizations—sounds that are introduced when the player deliberately sings simultaneously at a frequency different from that of the lips.

Acoustical aspects of the didjeridu previously studied include observations of the lip motion (Wiggins, 1988; Tarnopolsky *et al.*, 2006), numerical modelling of the lip motion (Hollenberg, 2000), the linear acoustics of the instrument (Amir and Alon, 2001; Amir, 2003, 2004; Caussé *et al.*, 2004), the lip-air column interaction (Fletcher, 1983; 1996) and the interactions among the vocal tract, the vocal folds, the lip motion, and the didjeridu air column (Tarnopolsky *et al.*, 2005, 2006; Fletcher *et al.*, 2006).

Another unusual feature of the instrument is that its bore is largely constructed by termites that eat the interiors of small eucalypt trees. This produces an irregular and somewhat flared bore. Suitable tree trunks are selected by listening to the sound made by tapping them, then cut to yield a length of typically 1.2–1.5 m (different cultural groups have different styles). The central bore is then cleared of debris

[a]Electronic mail: john.smith@unsw.edu.au

and a ring of beeswax placed on the smaller end of the instrument to allow a better and more comfortable seal around the player's mouth.

In the first two papers in this series (Tarnopolsky *et al.*, 2006; Fletcher *et al.*, 2006), it was shown how and why changes in the geometry of a player's vocal tract may have a large effect on the timbre of the sound produced by the instrument. But what is the effect of the geometry of the bore? More particularly, what makes a didjeridu a good or a bad instrument in the judgment of performers?

One possibility is to examine the sound produced by a didjeridu when played (e.g. Amir 2003, 2004; Caussé *et al.* 2004): a good instrument should make a good sound. However, this approach has the disadvantage that the quality of the didjeridu/player combination is being studied rather than the intrinsic quality of the didjeridu. (A superior player might be capable of producing a superior sound from an inferior instrument, whereas an inferior player might only produce an inferior sound from a superior instrument.) Further, the very large variation in timbre that an experienced player may produce on one instrument, using his vocal tract, could easily mask differences among instruments.

The above problems can be avoided, in principle, by measuring the intrinsic acoustical properties of instruments, without players, and comparing these with subjective rankings of quality provided by players. However, studies aimed at correlating the intrinsic acoustical properties of orchestral instruments with their quality as judged by players are known to be difficult for a variety of reasons (e.g., Pratt and Bowsher, 1978, 1979). The difficulties may be less with the didjeridu for the following reasons:

(i) Orchestral instruments have usually undergone a long evolutionary period resulting in designs and manufacturing processes that can produce instruments that are close to optimal. Consequently the differences among orchestral instruments of the same type are usually relatively subtle (e.g., Widholm *et al.*, 2001). Didjeridus are quite different. Although considerable skill is involved in selecting a trunk that might produce a good didjeridu, and in reaming out the bore appropriately, the quality is virtually unknown until the didjeridu is actually played. Because the shape of the bore is largely determined by termites and the shape of the tree trunk, the variation in bore, and consequently in quality, is very great among didjeridus; far more than would normally be encountered with orchestral instruments. These extreme variations could help in deciding which acoustical properties are important to players (e.g., Amir, 2004; Caussé *et al.*, 2004).

(ii) Much can be learned about an orchestral instrument from its appearance. For example, a simple visual inspection of a brass instrument can provide information on crucial parameters such as the bore dimensions and profile. Furthermore, the name of the instrument maker, the materials used, the degree of finish, and evidence of previous use can also influence a player's assessment. The avoidance of factors that might affect a player's assessment thus introduces constraints upon performance, e.g., not being allowed to hold the instrument normally, or playing in very dark rooms, or even being blindfolded (e.g. Bertsch, 2003; Inta *et al.*, 2005). The situation is quite different for the didjeridu; there is no simple, obvious correlation between its appearance and its quality (see Fig. 1) and consequently players can handle and play the instrument normally. Furthermore, there are no manufacturers' insignia on a new instrument and the external shape of the instrument provides little information on the critical bore profile and finish. Indeed, the detailed bore profiles are not known even to the experimenters. Perhaps the only information that can be gleaned from a casual inspection is the size and degree of flaring in the bore at the end of the instrument. However one possible source of information might be the tribal markings on some authentic instruments because some cultural groups and/or makers are known to prefer different bore profiles (Knopoff, 1997).

(iii) Orchestral wind instruments usually play many different notes at frequencies that are determined by the extrema in the acoustic input impedance of the instrument. (The acoustic impedance is the complex, frequency-dependent ratio of acoustic pressure to volume flow.) The multiple pitches introduce a wide range of desirable features that should be assessed for each note and combinations of notes including the intonation, stability, timbre, dynamic range, ease of transition, etc. (e.g., Plitnik and Lawson, 1999). In

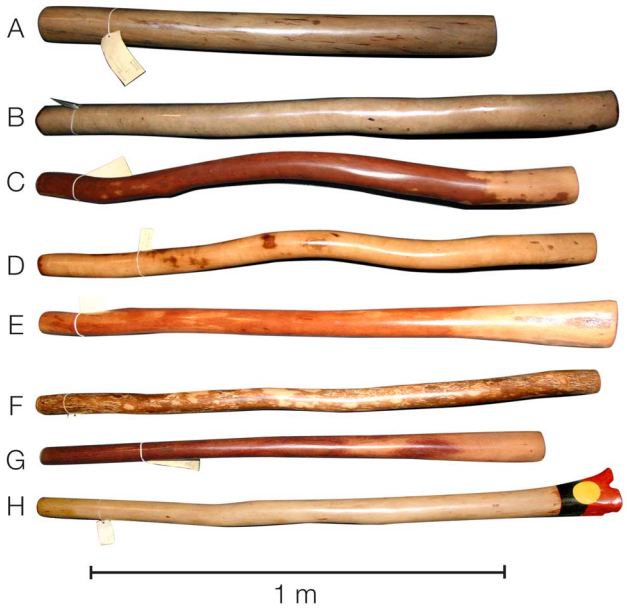


FIG. 1. (Color online) Photographs (to scale) illustrating the difficulty of identifying the quality of a didjeridu from its external appearance and geometry. The instruments labelled A–H were classified by the Didjshop as being *high concert, medium-high concert, medium concert, medium-low concert, low concert, first class, first-second class, and second class*, respectively. Their ranked qualities determined in this project were 5, 6, 7, 13, 21, 35, 37, and 26 respectively. As players know, it is difficult to judge a good instrument by inspection. This composite image was composed from images of the individual instruments that were adjusted to correct relative scale.

contrast, the didjeridu is usually played at only one pitch and at a limited range of volumes. Although the didjeridu introduces a new set of qualities to be assessed that are associated with the influence of the vocal tract and additional use of the vocal folds, the qualities that need to be assessed are still rather fewer than for orchestral instruments. This greatly reduces the testing required for each instrument.

(iv) The effectiveness and ease of operation of the keys or valves on a traditional wind instrument, and even their "feel" under the player's fingers, may influence judgments about quality. Didjeridus have no keys or mechanical parts.

Any decision on the quality of a musical instrument should be made by experienced players. In this paper the quality of 38 naturally made didjeridus, spanning a wide range of quality and geometry, and a typical range of pitch, was assessed by two very experienced players involved in selling didjeridus (from a wholesale and mail order enterprise called "the Didjshop") and also a panel of five experienced players with a range of styles from traditional to contemporary. To determine the possible influence of pitch on assessed quality, a control study was also made on nine didjeridus made from plastic pipe, which differed only in their length. The subjective assessments of players were then correlated with detailed measurements of the acoustic input impedance spectra and geometrical parameters of the 38 didjeridus to determine which acoustical properties are important to players.

II. MATERIALS AND METHODS

A. The didjeridus

The 38 traditional didjeridus studied were all from the collection of the "Didjshop" (www.didjshop.com). This is a commercial enterprise located near Kuranda in North Queensland, approximately 2000 km to the north of Brisbane, the state capital. It handles thousands of didjeridus each year; each one has been harvested and crafted by a local indigenous craftsperson from a termite-eaten *Eucalypt*. This results in instruments with a very wide range of geometries and consequently there is a large variation in quality among these instruments. During the last 13 years the Didjshop has developed a standardised procedure to classify these instruments and maintains a large database including these classifications and other information.

Each didjeridu used in the current study has undergone the standard preparation of instruments handled by the enterprise. Each is subjected to a proprietary treatment to stabilize the wood against cracking, and then varnished or painted if required. (For painting, the instruments are returned to local indigenous artists.) Each is then played and assigned a nominal playing pitch (henceforth denoted as the *pitch*). Two experienced musicians then play the instrument and assess its sound quality in one of eight categories. They also compare it with a permanent set of reference instruments consisting, for each note over the normal pitch range, of several instruments with different degrees of quality. The didjeridu to be sold is then assigned to one of five major categories according to its overall sound and performance quality (see below).

The 38 traditional instruments whose acoustical properties were measured in the current study include 33 instruments from a collection that is not for sale, and therefore in principle could be used in future studies. They were selected to cover a wide range of assessed qualities and nominal pitches ranging from A1 (55 Hz) to G2 (98 Hz). There were five instruments at each of F2, E2, and C2; 4 instruments at D#2 and D2; 3 instruments at G2, C#2, B1, and A#1; 2 instruments at A1.

The internal diameter at the mouth end of the instrument, d_{in}, had been measured previously for each instrument prior to the application of the beeswax. The maximum and minimum internal diameters of the often asymmetric bore at the terminal end of the instrument were measured using calipers, and the effective output diameter, d_{out}, taken as their geometric mean.

The perceived quality of a didjeridu might depend upon its playing pitch. This would be difficult to investigate using traditional instruments because of the wide variations in bore profile between instruments. Consequently a study was also made on 9 PVC tubes with the same diameter (38 mm), but different lengths, that covered the pitch range from A1 to F2. A study was also made on a set of 3 PVC tubes with different diameters (25, 38, and 50 mm), but the same pitch—C2 (nominally 65.4 Hz).

B. The players

In addition to the two experienced members at the Didjshop, five other experienced players volunteered for this study to assess the sound quality of each didjeridu on each of seven criteria. They then also ranked each didjeridu according to its overall quality. These five players had an average of 11 years playing experience, in styles ranging from traditional to contemporary.

C. The assessed quality of each didjeridu

The classification procedure used by the Didjshop involves the staff assessing the sound quality of each instrument by awarding a score out of five in each of the following seven categories; *backpressure, clarity, resonance, loudness, overtones, vocals*, and *speed*. Each instrument is then assessed for its *overall quality* by awarding a score out of ten. This classification system has been used by the staff for 13 years to describe thousands of instruments. For each instrument, these and other data, a brief sound file and a photograph are recorded. Because this database was available, and because the classifications were familiar to the standard panel players, we retain this classification system for the current study. However, in this paper we concentrate on the *overall quality* awarded.

The five major categories used by the Didjshop are given names chosen for commercial reasons. They are (with the associated scores):

High concert class. These didjeridus represent the best 1%–2% of the instruments sold/handled by the Didjshop. (They are judged to score 10 out of 10 for overall quality).

Medium concert class. These didjeridus will have rated 4 out of 5 in most sound attributes and should not score less than 4 in either clarity or resonance (overall quality score 8 of 10).

Low concert class. These didjeridus are guaranteed to have a better sound quality than that of a PVC pipe (overall quality score 6 out of 10).

First class. These are of sound quality similar to or better than that of a PVC pipe (overall quality score 4 out of 10).

Second class. These instruments display a major shortcoming in one or more categories, usually clarity. It is suggested that these are more suited to being used as artifacts for display rather than as musical instruments (overall quality score 2 out of 10).

Instruments are also often assigned to intermediate classes resulting in nine possible classes. Examples of the sound produced by a representative instrument belonging to each of the above five main classes when played by the same musician are available on the web: see http://www.didjshop.com/shop1/soundscapescart.html; The instruments studied spanned the whole range from "*second*" to "*high concert*."

D. Measurements of the input impedance spectra of each didjeridu

An impedance spectrometer described previously (Smith *et al.*, 1997; Epps *et al.*, 1997) was adapted for this study, using some of the techniques of Gibiat and Laloe (1990) and of Jang and Ih (1998), and used to measure the input impedance spectrum of each instrument, i.e., the impedance at the end that is blown by the player. Briefly, a waveform is synthesized from harmonic components, amplified, and input via a loudspeaker and impedance matching horn to a waveguide of known geometry with three microphones. The spectrometer is calibrated by connecting this waveguide to two reference impedances. One reference is an acoustic open circuit—a rigid seal located at the measurement plane. The other reference is an acoustically quasi-infinite cylindrical pipe whose impedance is assumed to be real and equal to its calculated characteristic impedance. The spectrometer is then connected to the impedance to be measured, and the unknown impedance spectrum calculated from the pressure components measured in the measurement and two calibration stages.

The bore diameter of the didjeridu is larger than that of the instruments we have studied previously (Smith *et al.*, 1997; Wolfe *et al.*, 2001) and consequently a lower impedance reference was required. The acoustically semi-infinite cylindrical pipe used for calibration in this study had an internal diameter of 26.2 mm and a length of 194 m. Because the first curve in the pipe occurs at 40 m from the impedance spectrometer and because any curves have a radius of 5 m or greater, the effects of reflections from these curves are expected to be negligible and this reference impedance should be purely resistive.

Measurements were made at a sample rate of 44.1 kHz using a MOTU 828 audio converter (nominal 24-bit) connected via a Firewire serial connection to a MacIntosh G4 iBook. Small electret microphones (Tandy #33-1052) were used in conjunction with a deltatron preamplifier (Bruel & Kjaer 2693 OS2). The stimulus waveform contained 2194 harmonics with a fundamental frequency of 1.346 Hz spanning the frequency range from 50 to 3000 Hz.

III. RESULTS AND DISCUSSION

A. Consistency of the subjective assessments of overall quality

Musicians have different musical backgrounds, preferences and individual styles. This observation is at least as true of the didjeridu players as of most other instrumentalists; indeed the variation may be greater for the didjeridu because of the absence of a formal academic tuition system with examinations. Consequently, a player's assessment of the quality of an instrument could be influenced by its suitability to his particular style of playing. It is thus important to examine how consistently the instruments are judged by the players. Because it is difficult to maintain an absolute scale for these subjective parameters, the relative rankings will be used for analysis rather than the individual scores.

1. Ranking of traditional instruments by the players

The Kendall coefficient of concordance, W, calculated from the individual ranking of overall quality by the 5 players for the 38 traditional didjeridus was calculated to be $W=0.335$. [W is a parameter used to rank the concordance of judgements; $W=1$ if all players ranked the instruments in exactly the same order, $W=0$ if the rankings were essentially at random, see Kendall and Babington Smith (1939).] This value corresponds to a $\chi^2=62$ and a probability $p=0.006$ that such a result might arise by chance. (The "probability" p is used hereafter in this sense, subject to the usual assumptions.)

2. Ranking of PVC pipes by the players

Nine of the 11 PVC pipes studied had the same internal diameter and were identical except for their length; this should only substantially affect the sounding pitch of each instrument. Providing the assessed overall quality is independent of *pitch*, the scores assigned by the five players to these nine instruments would be expected to be very similar, and ranked in no particular order. Indeed the Kendall coefficient of concordance, W, showed no significant correlation in the ranking of overall quality for these instruments. There was also no significant correlation between overall quality and *pitch*. The average score for overall quality for the PVC pipes was (3.0/10) which would have placed them in a traditional didjeridu category midway between *first class* (4/10) and *second class* (2/10). This is consistent with the definitions given in Sec. II C.

Three of the PVC pipes were of the same length, but had different internal diameters (25, 38, and 50 mm). There was a significant difference in their scores and in their rankings. The value of $W=0.88$ indicated a significant correlation between the rankings for overall quality, with a probability $p=0.012$ that this correlation might arise by chance. Players showed a distinct preference for a large internal diameter

with the widest tube ranked as the best and the narrowest tube ranked as the worst of the 11 PVC pipes.

3. Comparison with the Didjshop rankings

The assessments made by the Didjshop involve musicians with experience of thousands of instruments. It is thus quite likely that their ranking might differ from those made by the five volunteer players. However the Spearman rank-order correlation coefficient, r_S (see Kendall and Babington Smith, 1939) calculated between the average ranking awarded by the Didjshop and the average ranking of the five players for the traditional instruments was $r_S=0.51$ with a probability $p=5\times10^{-4}$ that such a non-null result might arise by chance. There was thus good agreement between the rankings of the Didjshop and the players.

The "ranked quality" of each instrument used for the remainder of this paper was derived by averaging the Didjshop ranking for overall quality and the mean ranking of the panel of five players for overall quality. There were only two ties in the ranked quality of the traditional instruments (one pair with rank 17 and one pair with rank 30).

B. Does assessed quality depend upon *pitch*?

A most obvious difference among the 38 traditional didjeridus studied is that they play at different pitches. It is thus important to check first whether the *pitch* influences the ranked quality. For example, it is possible that players might prefer didjeridus in a certain pitch range, perhaps the range they have most commonly used or heard. However the following results indicate that, over the range studied (A1–G2), *pitch* does not influence the ranked quality:

(i) the absence of any significant correlation between *pitch* and ranked quality for the 38 traditional didjeridus;
(ii) the absence of any significant correlation between *pitch* and ranked quality for the 9 PVC pipes;
(iii) the absence of any significant concordance in the order in which players ranked the quality of the 9 PVC pipes of the same diameter with different *pitch* ($W=0.11$, $p=0.84$).

Consequently, we did not divide the didjeridus into groups with limited pitch range, and instead used all of the 38 traditional didjeridus studied for the correlation studies

C. The acoustic input impedance

Figure 2 shows the input impedance spectrum $Z(f)$ measured on three of the instruments. Of the 38 studied, they were the instruments with the best, an intermediate, and the worst overall ranking for quality. All the impedance spectra showed the expected strong resonance at the fundamental frequency that is required for maintenance of the vibration regime. However the amplitude of the higher resonance peaks decreased rapidly with increasing frequency in the $Z(f)$ spectra of didjeridus of higher ranked quality, leaving spectra with little structure and only maxima of low magnitude in the region of 1–2 kHz, which is where formants are

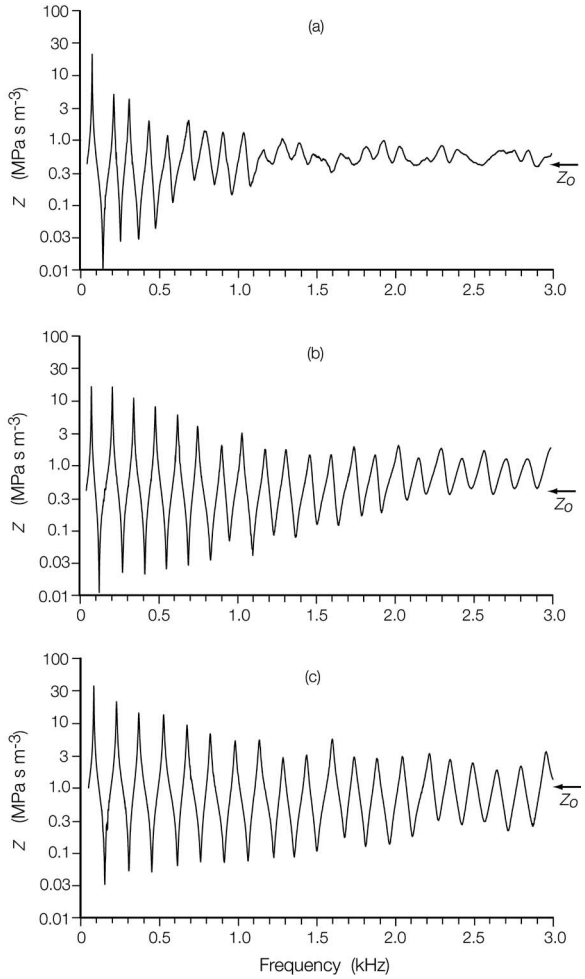


FIG. 2. Semilogarithmic plots of the measured acoustic impedance as a function of frequency for three didjeridus with the highest ranked quality (top), an intermediate ranked quality (middle), and worst ranked quality (bottom) from the 38 traditional instruments. The Didjshop classifications of these instruments were *high concert, medium-low concert, and first-second class*, respectively. The arrow on the right of each figure indicates the estimated value of the characteristic impedance, Z_0, for that instrument.

emphasized in performance. (Throughout this paper, we use "resonance" to refer to a passive acoustic property of a system and "formant" to refer to a consequent broad spectral peak in the envelope of the sound spectrum. Although some writers use "formant" in both senses, in this paper it is necessary to distinguish carefully the two different phenomena.)

In general, the magnitudes of the first few peaks in the spectrum decrease monotonically with increasing frequency. For brevity, we write Z_n to denote the magnitude of the nth maximum in $Z(f)$ with frequency f_n. In only one of the 38 instruments studied was Z_2 significantly higher than Z_1 (instrument #5). The 38 traditional didjeridus studied here are thus quite different from the instrument E1 studied by Amir (2004) for which Z_3 was noticeably higher than Z_1 and Z_2.

D. The relationship between quality and the acoustic impedance of the fundamental resonance

The low frequency impedance of the instrument is mainly responsible for controlling the vibration regime. Although a strong low frequency resonance is obviously essential, the strength of the resonance (however defined) might

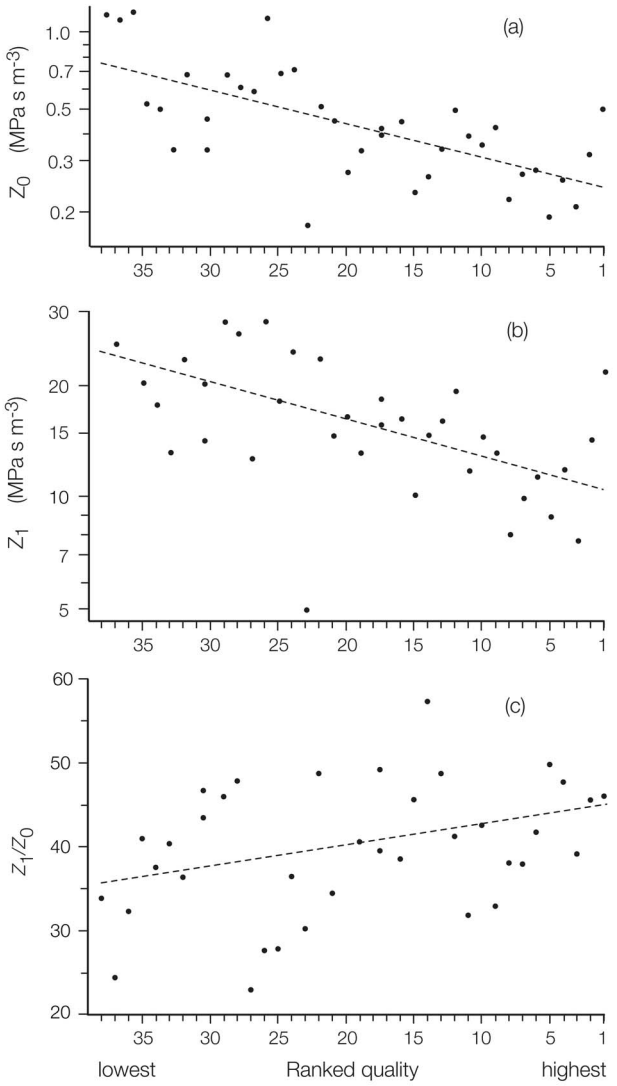


FIG. 3. The relationships between measured features in the low frequency acoustic impedance and the ranked quality of the 38 traditional instruments. (a) displays Z_0, the characteristic impedance, on a logarithmic scale. (b) displays Z_1, the value of the first maximum, on a logarithmic scale. (c) displays the ratio Z_1/Z_0. The datum with an anomalously low value in (a) and (b) was measured on didjeridu #5. This instrument was ranked as #23 and was the only didjeridu where Z_1 was significantly less than Z_2. The dashed lines indicate the correlations. The probabilities were $p=5\times10^{-6}$, $p=1\times10^{-4}$, and $p=0.04$, respectively, that the correlation might occur by chance.

be less important for the didjeridu than for other instruments where precise and rapid changes in pitch are important.

1. Z_0: The characteristic impedance

An effective value for Z_0, deliberately biased by frequencies near that of the playing pitch, was defined as the geometric mean of the magnitudes of the first maximum (Z_1) and first minimum ($Z_{\min}$) in the impedance spectrum. [On the semilogarithmic plots of $Z(f)$ shown in Fig. 2 it falls halfway between the first maximum and minimum.] Z_0 was found to vary from approximately 0.2 to 1 MPa s m^{-3}. This corresponds to the expected values for cylindrical pipes with diameters in the approximate range 25–50 mm. Figure 3(a) shows there is a negative correlation between ranked quality and Z_0, higher quality instruments being associated with a lower Z_0. It could be argued that the effective value of Z_0 defined above is a slight overestimate because the magnitude of successive extrema in $Z(f)$ will decrease with increasing frequency. This can be a consequence of increased losses due to wall damping and/or radiation losses from the open end of the instrument. The decrease in extrema with increasing frequency could be described in the form

$$Z_i = Z_1 \exp[-\lambda(f_i - f_1)], \tag{1}$$

where λ was estimated by linear regression to the first 3–6 maxima of $\log Z(f)$. The regression was always significant to 5% or better. The effect of a decrease in extrema with increasing frequency upon Z_0 can then be allowed for using the relationship,

$$\log_{10}(Z_0) = [\log_{10}(Z_1) + \log_{10}(Z_{\min}) - \lambda f_{\text{diff}}/\log_e(10)]/2, \tag{2}$$

where f_{diff} is the separation in frequency between Z_1 and $Z_{\min}$. Correction of Z_0 for the decrease in extrema with increasing frequency was found to reduce Z_0 by an average value of 5% and to make no significant difference to any of the correlations studied.

2. Z_1: The maximum impedance of the first resonance

Z_1 was estimated from the largest magnitude measured around the first resonance peak and was found to range from approximately 5 to 40 MPa s m^{-3}. Figure 3(b) shows there is a negative correlation between ranked quality and Z_1, higher quality instruments being associated with a lower value of Z_1. The impedance spectrum was measured with a frequency resolution of 1.346 Hz and consequently a value of Z_1 estimated from the largest measured value can be an underestimate because the peak value of Z_1 is unlikely to be measured. Consequently Z_1 was also estimated by fitting a polynomial function to the $Z(f)$ data on each side of the measured maximum, and then finding their intersection. Z_1 estimated in this fashion was only an average of 12% greater than the maximum measured value. This made no significant difference to any of the correlations studied.

3. Z_1/Z_0: The ratio of the maximum value of the first resonance to the characteristic impedance

Both Z_0 and Z_1 are negatively correlated with ranked quality. The ratio Z_1/Z_0 was investigated as a measure of the "strength" of the resonance. It varied from approximately 20 to 70, and was positively correlated with ranked quality, but to a smaller degree than the negative correlations for Z_0 and Z_1 [see Fig. 3(c)]. The correction of Z_0 for a decrease in extrema with increasing frequency, and Z_1 for finite frequency resolution, only slightly reduced the correlation of the ratio Z_1/Z_0 with ranked quality.

4. Q_1: The quality factor of the first resonance

Q_1 was examined as a measure of the "sharpness" of the fundamental resonance. A precise estimation required applying a polynomial interpolating function to the $Z(f)$ data on each side of the first measured maximum, and then finding

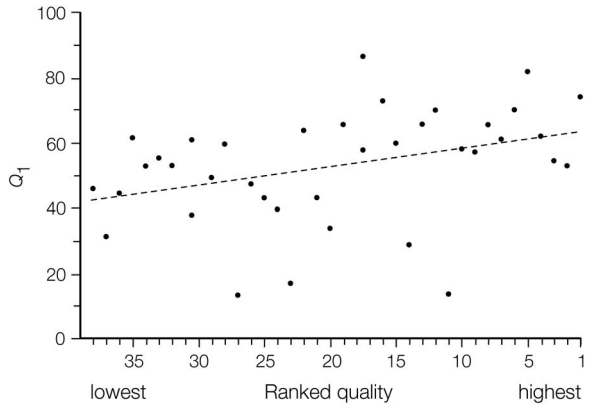


FIG. 4. The relationship between the Q_1, the measured Q factor of the first impedance maximum, and the ranked quality of the 38 traditional instruments. The dashed line indicates the correlation. The probability was $p=0.02$ that this correlation might occur by chance.

their intersection. Values ranged from 12 to 86. There was a modest positive correlation between Q_1 and ranked quality (see Fig. 4) with players preferring instruments with a sharper peak in $Z(f)$.

5. f_1: The frequency of the first resonance

f_1 ranged from approximately 50 to 105 Hz. No significant correlation was found between ranked quality and f_1 (data not shown). This is consistent with the results found with the 9 PVC pipes of different length, but the same diameter.

The average difference between the frequency of the nominal playing *pitch* and f_1 was 2%, with the *pitch* frequency usually slightly above f_1. In only one case was the *pitch* frequency significantly below f_1, this was instrument #5 which was noticeably anomalous in that it was the only instrument for which Z_1 was significantly lower than Z_2. In this case the *pitch* frequency (98 Hz) fell between f_1 (102 Hz) and $f_2/3$ (92 Hz).

The relationships between ranked quality and the characteristics of the first resonance show that players strongly prefer instruments with an overall low impedance and, to a lesser degree, a relatively narrow and strong fundamental resonance.

E. The relationship between quality and the low frequency impedance maxima

It has been suggested previously that the characteristic sound of a didjeridu is partly explained by the presence of a characteristic "hole" in the sound spectrum: in other words, the first few overtones above the fundamental are weak (Amir, 2003, 2004). Could this feature of the spectrum be related to the perceived quality? In most musical reed and lip-valve instruments, at least in the low region of their pitch range, several of the first few maxima in the impedance spectrum are in nearly harmonic ratios: either 1:2:3, etc. (oboe, saxophone etc.), 2:3:4 etc. (brass instruments) or 1:3:5 etc. (clarinet). This means that, at least for low notes, harmonics of the vibration of the reed or lips fall very close to the frequencies of impedance maxima. At these frequencies, the bore operates as an impedance transformer, efficiently matching the relatively high acoustic impedance at the lips or reed to the low impedance of the radiation field.

A simple cylindrical pipe has strong impedance maxima with frequencies that fall close to the ratio 1:3:5. Thus the third and fifth harmonics of a note played near the frequency of the first maximum fall near resonances of the bore and so are radiated efficiently. When such a pipe is used as didjeridu, the spectrum of the sound inside the player's mouth has many harmonics, in the ratio 1:2:3 etc., whose amplitudes fall almost monotonically as frequency increases. The spectrum of radiated sound, on the other hand, has relatively weak second, fourth, and sixth harmonics (Tarnopolsky *et al.*, 2006).

A simple cylindrical pipe is classed as a relatively poor didjeridu. Could the quality of a noncylindrical traditional didjeridu be related to how well the first several harmonics are efficiently radiated as sound? Efficient radiation of these low harmonics can be reduced in two distinct ways, or by a combination of them:

(i) a bore whose impedance maxima were harmonically related, but which decreased rapidly with increasing frequency, would be less efficient as an impedance transformer, or
(ii) in a bore whose impedance maxima were not close to harmonic ratios, the harmonics of a note played near the frequency of the lowest resonance would not fall at impedance maxima, and so there would be inefficient radiation, even if impedance maxima did not decrease significantly with increasing frequency.

We investigate the second effect first: examining the correlation between ranked quality and the resonance frequencies of the instrument bore.

1. f_2/f_1: The frequency ratio of the second to the first resonance

The ratio f_2/f_1 ranged from approximately 2.6 to 3.0. The third harmonic would be enhanced in the radiated sound of an instrument with a value close to 3 compared to an instrument with a value significantly less than 3. No significant correlation was found between ranked quality and f_2/f_1 [see Fig. 5(a)].

2. Harmonicity of impedance maxima

The degree of inharmonicity (Fletcher, 2002) was examined by comparing the didjeridu resonance frequencies to those expected for a cylinder. A parameter denoted as the harmonicity was estimated as the slope obtained by fitting a linear relationship between f_n and $(2n-1)f_1$. There was always a good fit with $p<10^{-5}$. Values for the harmonicity ranged from 0.72 to 1.0 (a value of 1 would be expected for a perfect cylinder without end effects). There was no significant correlation with ranked quality [see Fig. 5(b)]. This is interesting because it has been previously suggested that inharmonicity can play an important positive role in determining quality (Amir, 2004).

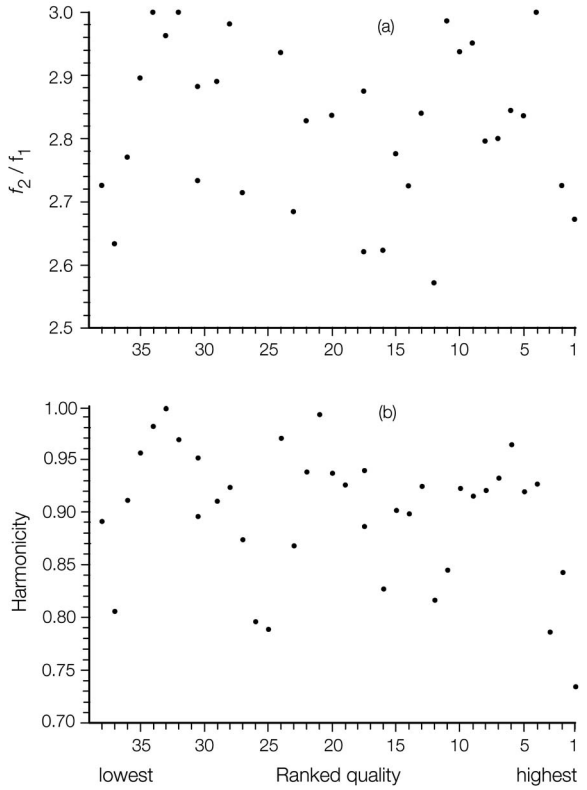


FIG. 5. The relationships between the frequencies of the maxima in $Z(f)$ and the ranked quality of the 38 traditional instruments. (a) and (b) indicate no significant correlation with f_2/f_1 (the frequency ratio of the second to the first resonance) or the harmonicity (see text), respectively.

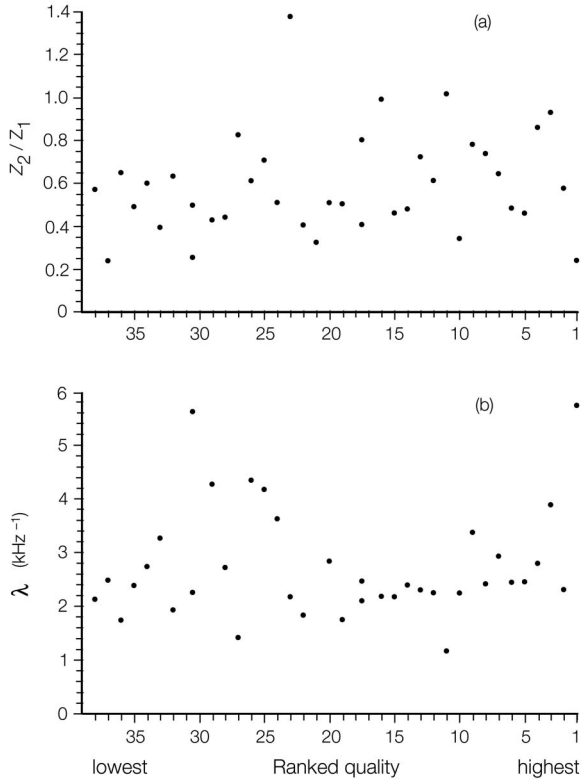


FIG. 6. The relationships between the magnitudes of the maxima in $Z(f)$ and the ranked quality of the 38 traditional instruments. No significant correlation was found for Z_2/Z_1 (the ratio of the magnitudes of the second to the first resonance) (a). No significant correlation was also found for λ (b). [λ describes how rapidly the magnitude of the low frequency extrema decrease with increasing frequency, see Eq. (1)]. The datum with an anomalously high value for Z_2/Z_1 was measured on didjeridu #5, the only didjeridu where Z_1 was significantly less than Z_2.

The ratio f_2/f_1 and the harmonicity can be significantly affected by the bore angle. An instrument with a larger bore angle (e.g., a truncated cone rather than a cylinder) might be expected to radiate more strongly, which might be a desirable characteristic. Although it is difficult to define a useful bore angle for such an irregular shape, f_2/f_1 is a measure of the acoustic property usually associated with bore angle.

Although the resonance frequencies appear to be uncorrelated with ranked quality, it is possible that the magnitudes of the maxima in $Z(f)$ might be important.

3. Z_2/Z_1: The ratio of the magnitudes of the second to the first resonances

The ratio Z_2/Z_1 ranged from approximately 0.2 to 1.4. No significant correlation was found between ranked quality and Z_2/Z_1 [see Fig. 6(a)].

4. The decrease in impedance extrema with increasing frequency

Could the ranked quality depend upon how rapidly the impedance extrema decrease with increasing frequency? Values for λ [see Eq. (1)] ranged from 1 to 6 kHz^{-1}. There was no significant correlation between λ and ranked quality [see Fig. 6(b)].

It could be argued that there is no need to look at the overall structure of the impedance spectrum. The shape of the first resonance peak (and perhaps one or two more) may influence the vibration regime but, although our measurements show the entire spectrum of $Z(f)$, the only frequencies present when a note is played at f_o are integral multiples of that frequency (nf_o).

5. Dependence on $Z(nf_o)$

The important parameters in determining the spectrum of the sound of the instrument playing at f_o will be the impedance at harmonics of the fundamental frequency, i.e., $Z(2f_o)$, $Z(3f_o)$,…, $Z(nf_o)$, etc. The value of f_o will probably vary from player to player, and may indeed be deliberately varied by a player to produce different interactions with $Z(f)$. To allow for this possibility the values of $Z(nf_o)$ quoted are the maximum values calculated when f_o was varied over a range of plus or minus one semitone in half cent steps from the measured value. There was no significant correlation between ranked quality and the individual magnitudes of either $Z(nf_o)$ or the normalized ratio $Z(nf_o)/Z(f_o)$ for $n=2-11$ (data not shown). However there was a good negative correlation with the sum $\Sigma Z(nf_o)$ for $n=2-11$; high quality instruments being associated with a low value for $\Sigma Z(nf_o)$ (see Fig. 7). The absolute value was important rather than the relative value: there was no correlation of ranked quality with the normalized sum $\Sigma Z(nf_o)/Z(f_o)$ (data not shown).

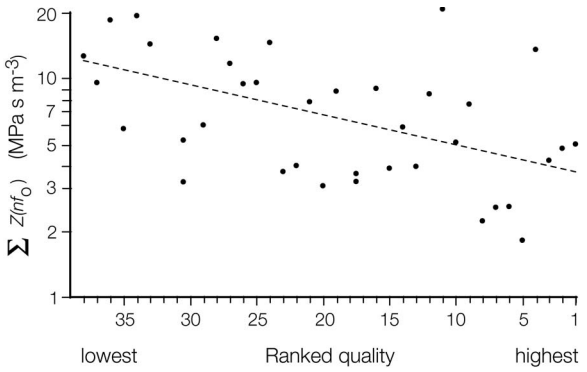


FIG. 7. A semilogarithmic plot showing the relationship between the sum $\Sigma Z(nf_o)$ for $n=2-11$ and the ranked quality. The dashed line indicates the correlation. There is a probability $p=0.001$ that the observed correlation might occur by chance.

F. Relationship between quality and mid range impedance

The vocal tract influences significantly the sound of the didjeridu when its impedance (measured just inside the player's lips, during performance) is comparable to or greater than that of the instrument (Tarnopolsky *et al.*, 2005, 2006; Fletcher *et al.*, 2006). The main influence of the tract on the sound appears to be the production of formants in the radiated sound in the frequency range from about 1 to 2 kHz. Sufficiently high peaks of acoustic impedance in the player's vocal tract inhibit acoustic airflow entering the instrument at those frequencies, by typically 20 dB (Tarnopolsky *et al.*, 2006, Figs. 4 and 6). These peaks in the tract impedance thus produce minima in the sound radiated from the instrument. The resultant formants in the instrument's sound (the frequency bands near tract impedance minima and not thus inhibited) are important in idiomatic performance. Consequently a high quality instrument might be expected to have a low acoustic input impedance in this frequency range. Comparison of the three examples shown in Fig. 2 suggests this might be the case. To quantify this hypothesis, Fig. 8 shows the relationship between the ranked quality and the maximum impedance in the range 1–2 kHz. A clear negative correlation between the ranked quality and the logarithm of maximum impedance is apparent; the highest ranking instruments have $Z \leqslant 1$ MPa s m^{-3}, whereas the instruments with lowest ranked quality have $Z \geqslant 3$ MPa s m^{-3}. This is consistent with our measurements that found the vocal tract impedance measured just inside the lips needed to be several MPa s m^{-3} before strong formants appeared in the output sound (Tarnopolsky *et al.*, 2005, 2006).

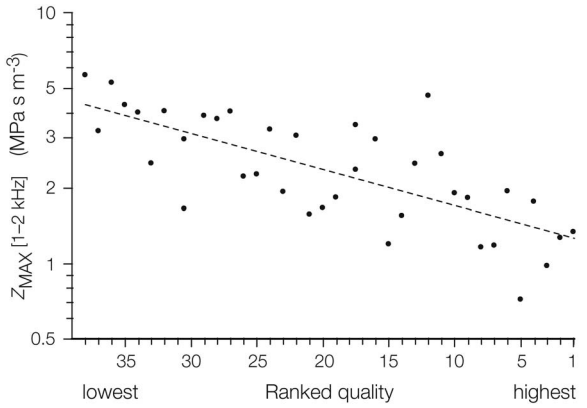


FIG. 8. A semilogarithmic plot showing the relationship between the maximum value of $Z(f)$ in the range 1–2 kHz and the ranked quality. The dashed line indicates the correlation. There is a probability $p=5\times10^{-7}$ that the observed correlation might occur by chance.

This observation suggests two more specific questions:

(i) Does a high quality instrument need to display a low value of $Z(f)$ in a particular narrow frequency range f_r?

(ii) Is it the maximum value Z_{MAX}, or the average value Z_{AVG}, or the rms variation Z_{RMS}, in $Z(f)$ over a particular frequency range that is important?

These were investigated by calculating the three above parameters for a number of narrower frequency ranges. A frequency range $f_r=200$ Hz was selected to ensure that at least one maximum in $Z(f)$ was always present. In each frequency range and for each instrument, the slope and intercept (both with errors) of the correlation of the logarithm of the calculated parameter with ranked quality were calculated for each different frequency range using a plot similar to Fig. 8. The slope and intercept were then used to calculate two values of Z predicted by this relationship and their associated errors; one that would be associated with the lowest ranked instrument and the other with the highest ranked instrument. These data are plotted in Fig. 9 on a logarithmic scale as a function of frequency. For each frequency range and parameter it is apparent that a decrease in Z_{MAX}, Z_{AVG}, or Z_{RMS} is always associated with an improvement in ranked quality. Figure 9(a) indicates that the improvement is most sensitive to fractional changes in Z_{max} around 1200 Hz. An improvement in ranked quality produced by a decrease in Z_{AVG} is significant over a wider frequency range (200–1600 Hz) [see Fig. 9(b)]. Figure 9(c) indicates that the ranked quality was most sensitive to fractional changes in Z_{RMS} around 1600 Hz.

The results shown in Fig. 9 indicate that, although a decrease in Z is always associated with an improvement in quality, low values in the frequency range 1–2 kHz are the best indicators of quality.

G. Effects of harmonic coincidence

We showed previously (Tarnopolsky *et al.*, 2006) that "harmonic coincidence" in the 1–2 kHz region can contribute in a modest way to the strength of formants in the output sound. A set of harmonically-related maxima in the impedance of the instrument in this frequency range might, in principle, be excited by the appropriate playing pitch, or "switched off" by a relatively small change in the playing pitch. In some of the sound samples studied, this effect contributed to the magnitude of formants, although the effect was rather smaller than that produced by vocal tract resonances. Because the formants in question typically involve the 20 to 30th harmonics, changes in the playing frequency of only a couple of percent (a fraction of a semitone) can produce or remove the harmonic coincidence effect.

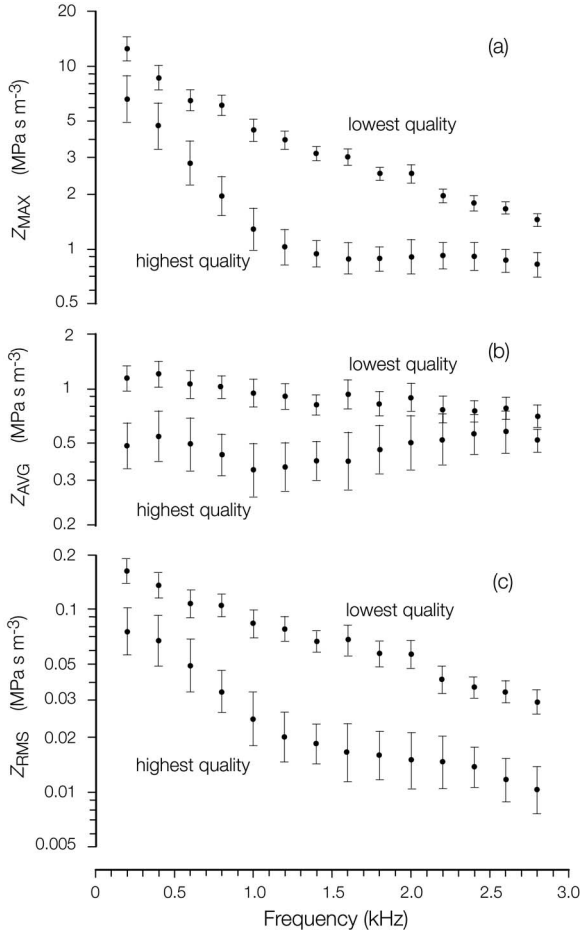


FIG. 9. Semilogarithmic plots showing the relationships between parameters associated with the impedance and the ranked quality as a function of frequency. Each figure shows the idealized behavior obtained from the correlations between that parameter and ranked quality over a frequency range of 200 Hz. Values towards the top of each figure indicate the high impedances associated with the lowest quality instruments. Values on the lower part of each figure indicate the low impedances associated with the highest quality instruments. Z_{MAX} indicates the maximum value, Z_{AVG} indicates the average value, and Z_{RMS} indicates the rms variation in $Z(f)$ over that frequency range. Error bars indicate the s.e.

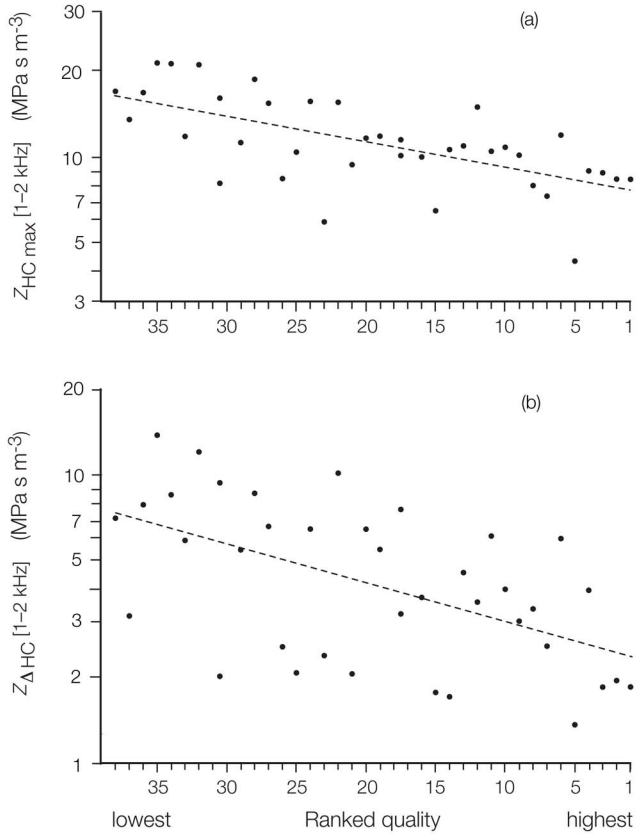


FIG. 10. The relationship between the magnitude of "harmonic coincidence" in the range 1–2 kHz and the ranked quality of the 38 traditional instruments. (a) shows Z_{HC}, the maximum value of the sum of $Z(nf_o)$ over the frequency range 1–2 kHz when f_o is varied over a range of plus or minus one semitone in half cent steps. (b) shows $Z_{\Delta HC}$, the difference between the maximum and minimum values of the sum of $Z(nf_o)$ under the same conditions. The dashed lines indicate the correlations. The probabilities are $p=2\times10^{-4}$ and $p=0.01$, respectively that the correlations might occur by chance.

Could the availability of resonances in the 1–2 kHz region contribute to perceived quality? This was tested by calculating Z_H defined as the sum of all the values of $Z(f)$ within a certain frequency range that corresponded to harmonics of the fundamental frequency f_0. [Interpolation between discrete values of $Z(f)$ was used where necessary]. Z_{HCmax} and Z_{HCmin} were defined as the maximum and minimum values of Z_H, respectively, when f_0 was varied over a frequency range of plus or minus a semitone in half cent steps. In all cases there was a clear negative correlation between Z_{HC} and ranked quality [see Fig. 10(a)]. Of course the difference in Z_H as f_o is varied might be the important parameter. Figure 10(b) shows $Z_{\Delta HC}$ defined as the difference $Z_{HCmax}-Z_{HCmin}$. Again there is a clear negative correlation between $Z_{\Delta HC}$ and ranked quality. It thus appears that any timbral variations produced by harmonic coincidence are more than countermanded by the necessary associated increase in $Z(f)$, which in turn reduces the effect of changes in the vocal tract impedance.

H. Relationships among geometry, quality, and acoustic impedance

The bore profiles of traditional, termite-eaten didjeridus are much more complicated than those of any other wind instrument. So it is difficult to estimate acoustic parameters precisely from geometry. However we can look at some factors.

A player might possibly be influenced by the size and heft of an instrument. However there was no statistically significant correlation between length and ranked quality.

Z_0 of a cylindrical pipe is inversely proportional to the cross-sectional area. The didjeridu bore-profile is much more complicated, but one might expect that Z_0 is proportional to d_{in}^{-2}. We found that Z_0 varies with d_{in}^{-2} with a slope 190 ± 30 MPa s m^{-5} ($p=5\times10^{-7}$). For a cylindrical tube the expected slope is around 500 MPa s m^{-5}.

Consequently a correlation between ranked quality and d_{in}, the bore diameter just inside the beeswax ring, might be expected. Figure 11(a) shows in this case, that higher quality instruments tend to have a larger diameter at the mouth end.

A correlation between ranked quality and d_{out} might also be expected. This is because waves in the bore are more readily transmitted to the radiation field if the instrument has

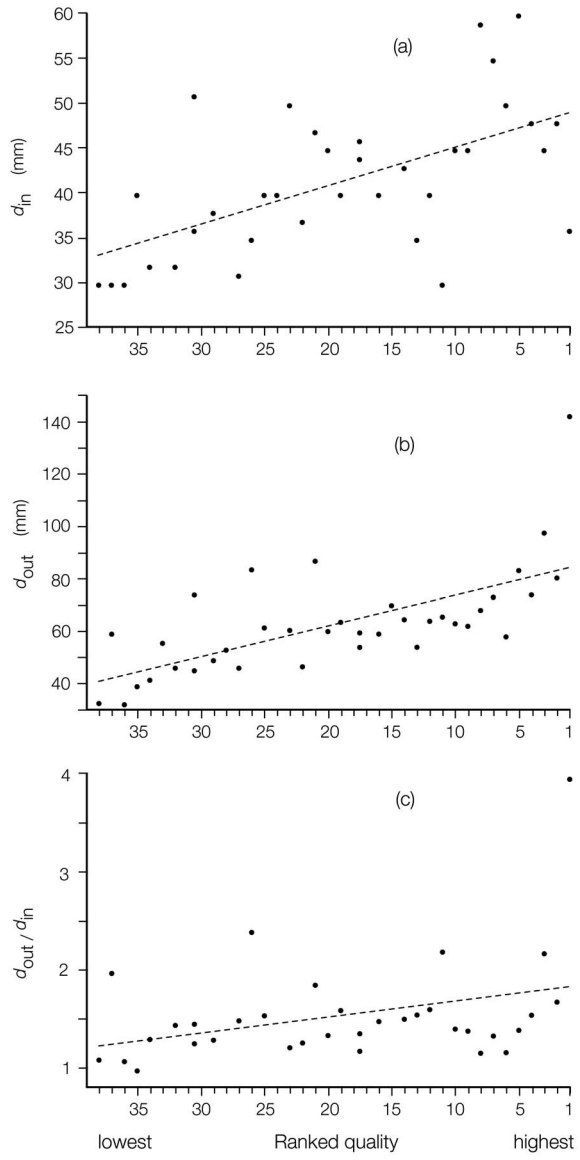


FIG. 11. The relationships between instrument geometry and ranked quality for the 38 traditional instruments. (a) shows d_{in}, the bore diameter just inside the beeswax ring. (b) shows the effective output diameter, d_{out}. (c) shows the overall flare ratio d_{out}/d_{in}. The dashed lines indicate the correlations. The probabilities were $p=2\times10^{-4}$, 7×10^{-6}, and 0.044, respectively, that the correlations might occur by chance. (Values of d_{in} were not available for 3 instruments, and d_{out} was not available for one instrument; hence the missing points.)

a large cross-sectional area at the output end. This has two effects: it reduces the magnitude of the standing wave and therefore the acoustic impedance associated with resonances. As explained above, this is a desirable consequence. Furthermore, the increased radiation increases the loudness, a property that is usually regarded by musicians as desirable in an instrument. Figure 11(b) shows that this is the case: higher quality instruments tend to have a larger diameter at the far end. There is also a positive correlation between the overall flare ratio d_{out}/d_{in} and ranked quality [see Fig. 11(c)].

IV. CONCLUSIONS

Figure 12 presents photographs of the seven didjeridus with the highest ranked quality of the 38 studied. The large

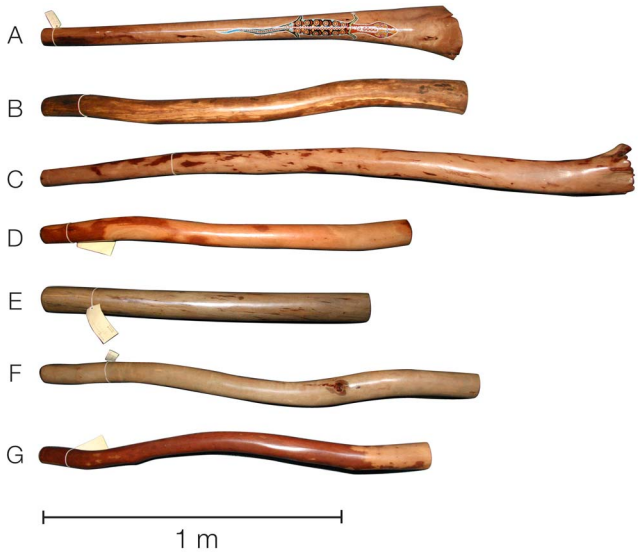


FIG. 12. (Color online) Photographs (to scale) illustrating the seven didjeridus with the highest ranked quality of the 38 studied. Instruments labelled A–G were ranked from 1 to 7, respectively. All instruments were classified by the Didjshop as being *high concert, medium-high concert, or medium concert*. This composite image was composed from images of the individual instruments that were adjusted for correct relative scale.

variations in external geometry illustrate just how difficult it is to judge a good didjeridu by inspection. The players' judgments of the overall quality of the instruments were significantly concordant and allow the following conclusions.

The dependence of ranked quality on a number of parameters is summarized in Table I. Most of the effects listed, however, are relatively small. Perhaps the most important conclusions are these:

(1) Ranked quality is not significantly related to pitch, nor to the frequencies of the resonances.
(2) Providing the characteristic input impedance at low frequencies is sufficiently low, players prefer a relatively narrow and strong (i.e., high Q) fundamental resonance.
(3) Quality is inversely related to the magnitude of the acoustic impedance in all frequency ranges. These results show that players strongly prefer instruments with an overall low impedance and the most important determinant of ranked quality is the impedance in the frequency range about 1–2 kHz, where strong resonances and high values of impedance are strongly negatively correlated with ranked quality.

This last point can be explained in terms of the techniques used for sound production and variation in the instrument (Tarnopolsky *et al.*, 2005, 2006; Fletcher *et al.*, 2006). At the frequencies around the maxima in the impedance spectra of the player's vocal tract, acoustic airflow into the instrument is inhibited and, consequently, output power from the instrument is weak in these frequency bands. The formants in the output sound are the remaining bands of frequencies—those not inhibited by high impedance in the tract. This inhibition effect requires that the maxima in the impedance of the vocal tract have magnitudes comparable with or greater than the input impedance of the didjeridu.

TABLE I. The correlations between the ranked quality and some acoustic and geometric parameters for the 38 traditional didjeridus (see text for definitions of symbols). A positive (+ve) or negative (−ve) sign indicates the sign of the correlation, whereas an asterisk indicates that there was no significant correlation at the 5% level. p indicates the probability that the observed correlation might arise by chance.

Parameter	Correlation	Probability p that correlation might arise by chance
pitch	*	
Z_0	−ve	5×10^{-6}
	Fundamental resonance	
f_1	*	
Z_1	−ve	1×10^{-4}
Z_1/Z_0	+ve	0.04
Q_1	+ve	0.02
	Harmonicity of first 6 harmonics	
f_2/f	*	
$f_n/(2n-1)f_1$	*	
	$Z(f)$ for harmonics 2 to 11	
Z_2/Z_1	*	
λ	*	
$Z(nf_o)$	*	
$Z(nf_o)/Z(f_o)$	*	
$\Sigma\ Z(nf_o)$	−ve	0.001
$\Sigma\ Z(nf_o)/Z(f_o)$	*	
	$Z(f)$ between 1 kHz and 2 kHz	
Z_{MAX} (1–2 kHz)	−ve	3×10^{-7}
$Z_{HC\ max}$	−ve	2×10^{-5}
$Z_{\Delta HC}$	−ve	4×10^{-4}
	Geometry	
L	*	
d_{in}	+ve	2×10^{-4}
d_{out}	+ve	7×10^{-6}
d_{out}/d_{in}	+ve	0.044

Thus an instrument with low impedance and relatively weak resonances in this frequency range would allow the player greater control of the formants in the output sound and thus lead to high perceived playing quality.

ACKNOWLEDGMENTS

We thank the Australian Research Council for support. Svargo Freitag and the staff of the Didjshop kindly made available the sets of instruments and information from their own records. We would also like to thank the players who volunteered for the evaluation panel. G. R. participated as an undergraduate research project.

Amir, N. (**2003**). "Harmonics: What do they do in the didjeridu?," in Proceedings of the Stockholm Music Acoustics Conference (SMAC 03), edited by R. Bresin, Stockholm, Sweden, pp. 189–192.

Amir, N. (**2004**). "Some insights into the acoustics of the didjeridu," Appl. Acoust. **65**, 1181–1196.

Amir, N. and Alon, Y. (**2001**). "A study of the didjeridu: Normal modes and playing frequencies," in Proceedings of the International Symposium on Musical Acoustics, Perugia, edited by D. Bonsi, D. Gonzalez, and D. Stanzial, pp. 95–98.

Bertsch, M. (**2003**). "Bridging instrument control aspects of brass instruments with physics-based parameters," in Proceedings of the Stockholm Music Acoustics Conference (SMAC 03), edited by R. Bresin, Stockholm, Sweden, pp. 193–196.

Caussé, R., Goepp, B., and Sluchin, B. (**2004**). "An investigation on 'tonal' and 'playability' qualities of eight didgeridoos, perceived by players," in Proceedings of the International Symposium on Musical Acoustics (ISMA2004), Nara, Japan, pp. 261–264.

Epps, J., Smith, J. R., and Wolfe, J. (**1997**). "A novel instrument to measure acoustic resonances of the vocal tract during speech," Meas. Sci. Technol. **8**, 1112–1121.

Fletcher, N. H. (**1983**). "Acoustics of the Australian didjeridu," Australian Aboriginal Studies **1**, 28–37.

Fletcher, N. H. (**1996**). "The didjeridu (didgeridoo)," Acoust. Aust. **24**, 11–15.

Fletcher, N. H., (**2002**). "Harmonic? Anharmonic? Inharmonic?," Am. J. Phys. **70**, 1205–1207; (**2003**) **71**, 492(E).

Fletcher, N. H., Hollenberg, L. C. L., Smith, J., Tarnopolsky, A. Z., and Wolfe, J. (**2006**). "Vocal tract resonances and the sound of the Australian didjeridu (yidaki): II. Theory," J. Acoust. Soc. Am. **119**, 1205–1213.

Gibiat, V. and Laloe, F. (**1990**). "Acoustical impedance measurements by the two-microphone-three-calibration (tmtc) method," J. Acoust. Soc. Am. **88**, 2533–2544.

Hollenberg, L. (**2000**). "The didjeridu: Lip motion and low frequency harmonic generation," Aust. J. Phys. **53**, 835–850.

Inta, R., Smith., J., and Wolfe, J. (**2005**). "Measurement of the effect on violins of ageing and playing," Acoust. Aust. **33**, 25–29.

Jang, S.-H. and Ih, J.-G. (**1998**). "On the multiple microphone method for measuring in-duct acoustic properties in the presence of mean flow," J. Acoust. Soc. Am. **103**, 1520–1526.

Kendall, M. G. and Babington Smith, B. (**1939**). "The problem of m rankings," Ann. Math. Stat. **10**, 275–287.

Knopoff, S. (**1997**). "Accompanying the dreaming: Determinants of didjeridu style in traditional and popular Yolngu song," in *The Didjeridu: From Arnhem Land to Internet*, edited by K. Neuenfeld (John Libbey, Sydney), pp. 39–67.

Plitnik, G. R. and Lawson, B. A. (**1999**). "An investigation of correlations between geometry, acoustic variables, and psychoacoustic parameters for French Horn mouthpieces," J. Acoust. Soc. Am. **106**, 1111–1125.

Pratt, R. L. and Bowsher, J. M. (**1978**). "The subjective assessment of trombone quality," J. Sound Vib. **57**, 425–435.

Pratt, R. L. and Bowsher, J. M. (**1979**). "The objective assessment of trombone quality," J. Sound Vib. **65**, 521–547.

Smith, J. R., Henrich, N., and Wolfe, J. (**1997**). "The acoustic impedance of the Bœhm flute: Standard and some nonstandard fingerings," Proc. Inst. Acoustics **19**(5), 315–320.

Tarnopolsky, A. Z., Fletcher, N. H., Hollenberg, L. C. L., Lange, B. D., Smith, J., and Wolfe, J. (**2005**). "The vocal tract and the sound of the didgeridoo," Nature (London) **436**, 39.

Tarnopolsky, A. Z., Fletcher, N. H., Hollenberg, L. C. L., Lange, B. D., Smith, J., and Wolfe, J. (**2006**). "Vocal tract resonances and the sound of the Australian didjeridu (yidaki): I. Experiment," J. Acoust. Soc. Am. **119**, 1194–1204.

Wiggins, G. C. (**1988**). "The physics of the didgeridoo," Phys. Bull. **39**, 266–267.

Widholm, G., Linortner, R., Kausel, W., and Bertsch, M. (**2001**). "Silver, gold, platinum—and the sound of the flute," in Proceedings of the International Symposium on Musical Acoustics, Perugia, edited by D. Bonsi, D. Gonzalez, and D. Stanzial, pp. 277–280.

Wolfe, J., Smith, J., Tann, J., and Fletcher, N. H. (**2001**). "Acoustic impedance of classical and modern flutes," J. Sound Vib. **243**, 127–144.

Experimental study of A0 and T1 modes of the concert harp

J-L. Le Carrou[a] and F. Gautier
Laboratoire d'Acoustique de l'Université du Maine, UMR-CNRS 6613, Université du Maine, Avenue Olivier Messiaen, 72085 Le Mans Cedex 09, France

E. Foltête
Institut FEMTO ST, Université de Franche-Comté, Laboratoire de Mécanique Appliquée, 24 rue de l'Epitaphe, 25000 Besançon, France

(Received 20 January 2006; revised 5 October 2006; accepted 5 October 2006)

String instruments are usually composed of a set of strings, a soundboard, and a soundbox with sound holes, which is generally designed to increase the sound level by using the acoustic resonances of the cavity. In the case of the harp, the soundbox and especially the sound holes are primarily designed to allow access to the strings for their mounting. An experimental modal analysis, associated to measurements of the acoustic velocity in the holes, shows the importance of two particular modes labeled A0 and T1 as it was done for the guitar and the violin. Their mode shapes involve coupled motions of the soundboard's bending and of the oscillations of the air pistons located in the sound holes. The A0 mode is found above the frequency of the lowest acoustically significant structural mode T1. Thus, the instrument does not really take advantage of the soundbox resonance to increase its radiated sound in low frequencies. However, contribution of mode A0 is clearly visible in the response of the instrument, confirming the importance of the coupling between the soundboard and the cavity. © *2007 Acoustical Society of America.*
[DOI: 10.1121/1.2384842]

PACS number(s): 43.75.Gh [NHF] Pages: 559–567

I. INTRODUCTION

The harp is one of the oldest string instruments. Its evolution from the prehistoric instrument to the modern concert harp led to the elaboration of constitutive elements, which are designed to efficiently radiate the sound. The modern concert harp is composed of a set of strings directly connected to a long thin flat soundboard attached to a fairly solid soundbox with several sound holes. These three elements are coupled in a complex manner and are attached to a base, a pillar, and an arm as shown in Fig. 1.

In a string instrument, the mechanical characteristics of the strings define the note to be played and the soundboard is designed as a sound radiator. Unfortunately, this sound radiator is not efficient in the low-frequency range when the acoustic wave length is greater than the size of the soundboard. An acoustical resonator, called the soundbox, is generally added in order to increase the sound level. The first acoustic resonance of the cavity can be used to reinforce the sound radiation of the instrument. This effect is used in the design of bass-reflex enclosures. The acoustical resonator has been the subject of many studies[1] on the guitar and on the violin: the acoustic motion inside the cavity interacts with the motion of the soundboard to produce two coupled modes. The first mode is called the plate mode and is associated with a strong bending motion of the soundboard. In the case of the guitar[2] and of the violin,[3] this mode is commonly labeled T1. The second mode is called the Helmholtz mode or A0 air mode and corresponds to a strong motion of an air piston located in the hole. The A0 air mode contributes to a significant increase of the sound radiation in the low frequency range. In order to well understand this low-frequency behavior of the guitar or of the violin, simple discrete models[4–6] have been carried out. The parameters of these models can be obtained from transfer functions measurements on these instruments.

In the case of the harp, the cavity and holes' sizes and shapes are not particularly designed to amplify the sound in the low-frequency range. One of the main reasons for the choice of sizes and locations of the holes is the facility for string mounting. The acoustic role of the holes is not well understood because the harp, and especially the soundbox, has not been the subject of many vibroacoustic studies.

The first study[7] was carried out on the small harp of Scotland. Modal analysis has been performed on the soundboard at different steps of its manufacturing. Air resonances were also investigated in the soundbox alone by burying it into sand in order to damp wall vibrations. No evidence of the presence of a Helmholtz resonance was found. Moreover, in playing configuration, the relationship between vibration modes of the instrument and radiated sound was not investigated. This study was later carried out on a Spanish harp of the baroque period,[8] close in size to the current concert harp. In this study, it was found that vibroacoustic interactions between soundboard vibrations and the acoustic motions of the air cavity lead to two coupled modes (112 and 146 Hz) having similar shapes and corresponding to A0 and T1 modes, respectively. This kind of result was also found on an unstrung Salvi Orchestra Concert Harp[9] by using holographic interferometric analysis of the soundbox. The author of this last study identifies A0 and T1 modes by measuring

[a]Electronic mail: jean-loic.le_carrou@univ-lemans.fr

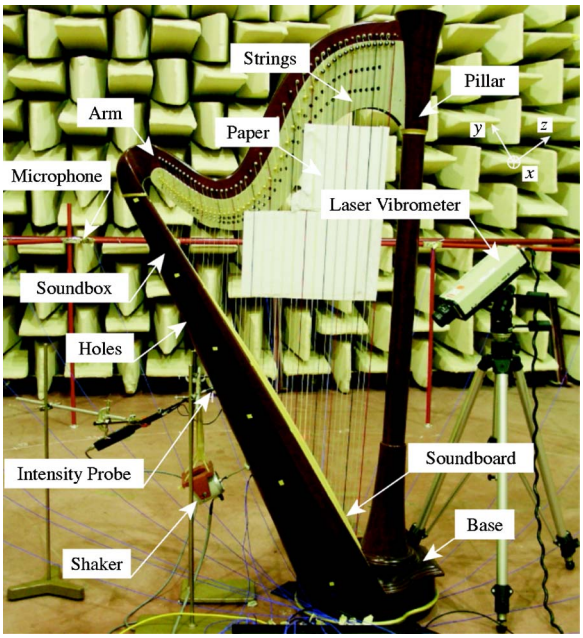


FIG. 1. (Color online) Experimental setup.

the changes of the structural response when sound holes are closed. For the Salvi Orchestra's soundbox, another study[10] has confirmed that if wall vibrations are damped by sand, the Helmholtz mode is clearly present in the acoustic response. A semi-empirical formula was proposed to predict its eigenfrequency. Another conclusion by Bell[9] is that one of the two coupled modes' (A0 and T1) presence in the response weakens when the soundboard is stressed by the strings. Thus, for a strung harp, the A0 mode is particularly difficult to identify. The reason for that is not clear. This difficulty was also pointed out on a Celtic harp.[11]

The aim of this paper is to identify the A0 air mode for the concert harp and to investigate the importance of its contribution to the instrument's response. For this purpose, the paper is divided into two parts. A study of the response functions of the instrument is first performed through the experimental modal analysis of the instrument's body and through an investigation of the acoustic field in the cavity. Then, the identification of the A0 and T1 modes is achieved by studying a modified instrument.

TABLE I. Dimensions of the five elliptical sound holes. The two dimensions correspond to the major axis and minor axis of each ellipse.

No.	Major axis (cm)	Minor axis (cm)
1	16.6	4.8
2	17.2	5.6
3	17.7	7
4	18.1	8
5	18.5	9.3

II. EXPERIMENTAL STUDY OF THE CONCERT HARP

A. Experimental procedure

The vibroacoustic behavior of a concert harp is experimentally investigated. All measurements are performed on an *Atlantide Prestige* concert harp lent by a French harp maker, Camac Harps. A schematic diagram is proposed in Fig. 2 with the principal dimensions of the instrument. The soundbox of the studied concert harp consists of a 6-mm-thick semi-conical shell with a total volume of the enclosed air of 0.029 m^3. On the back of the soundbox, there are five elliptical sound holes whose dimensions are shown in Table I. The concert harp is studied in playing configuration: all strings are mounted and tuned. For these measurements, strings are damped with paper to prevent their vibration while keeping the static deformation and load imposed by them on the soundboard. So, the string modes, including

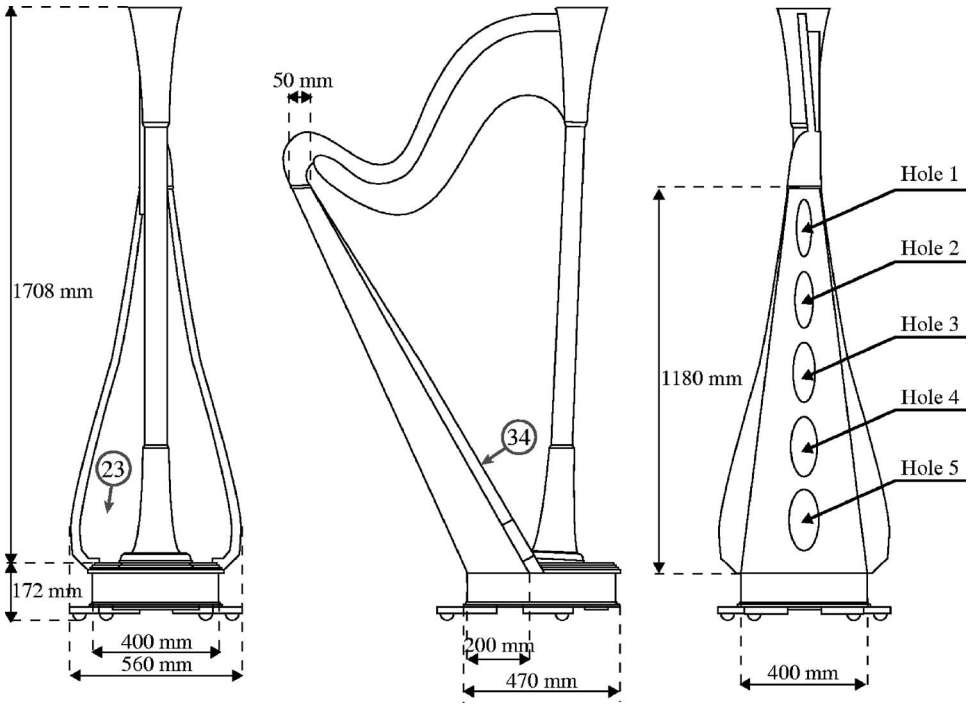


FIG. 2. Schematic diagram with dimensions of the *Atlantide Prestige* concert harp. The locations of two characteristic points 34 and 23 and hole number are also shown.

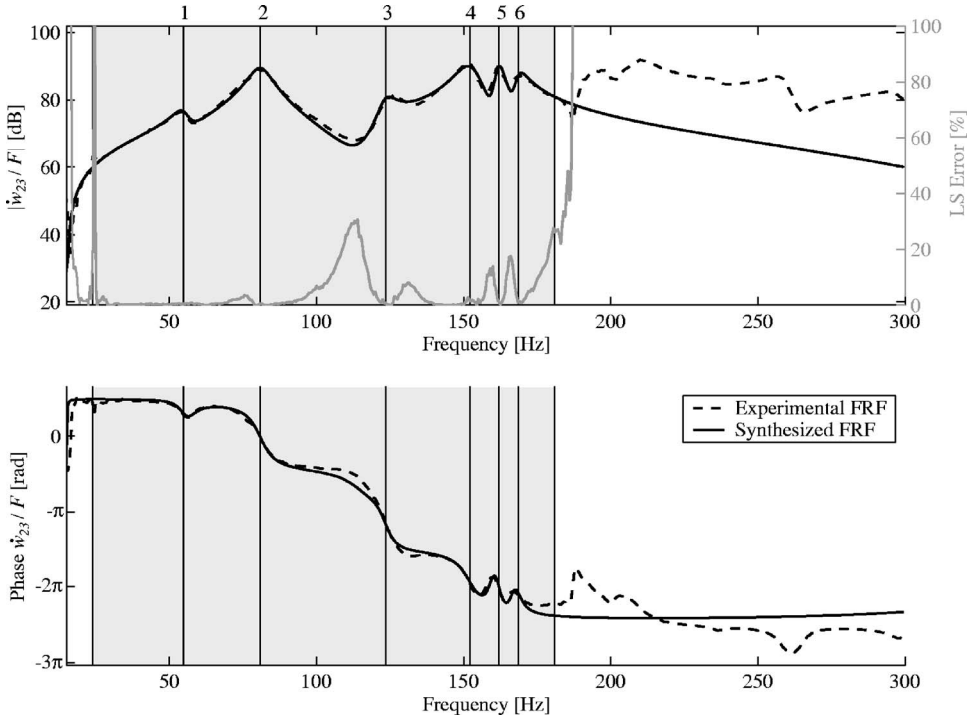


FIG. 3. Measured FRF $\dot{w}_{23}/F$, synthesized FRF $\dot{w}_{23}/F$, and least square error ϵ are shown versus frequency (Ref. 1 dB: 5×10^{-8} m s^{-1} N^{-1}). The grayed area corresponds to the frequency range in which modes have been identified. Numbers associated to vertical lines indicate the modal frequencies given in Fig. 4.

sympathetic modes,[12] are highly damped and are not evident in the instrument's response.

The experimental setup is shown in Fig. 1: the instrument is excited by a shaker driven by a white noise connected via a rod, through sound hole 4, to the back of the soundboard. The excitation force F is measured with an appropriate force sensor directly glued to the back of soundboard. The excitation point is labeled 34, as shown in Fig. 2, and is located between the Ab and the Bb string (respective fundamental frequencies at 103.8 and 116.5 Hz) attachment points. The vibratory velocity $\dot{w}$ is measured with a laser vibrometer. The acoustic velocity V in the middle of the sound holes is measured with an intensity probe. The far field acoustic pressure P is measured with microphones placed around the concert harp. Frequency response functions (FRFs) $H=\dot{w}/F$, $H_V=V/F$, and $H_P=P/F$ are then computed by a standard analyzer.

B. Experimental modal analysis of the instrument's body

The identification of structural modes of the soundbox in the low-frequency range is carried out by modal testing: eigenfrequencies, mode shapes, and damping parameters can be extracted from response functions measured at different points of the structure. The experimental mesh is composed of 60 points on the soundboard and of 18 points on the curved surface at the back of the harp, as shown on each modal shape in Fig. 4. The laser vibrometer is adjusted to measure the normal velocity on the soundboard. For each point on the curved surface, both the velocity along the z axis and along the x axis, defined in Fig. 1, are measured. Measurements are performed at each mesh point in the frequency range 0–300 Hz. A typical example of the measured frequency response functions is shown in Fig. 3.

The modal identification is carried out using the least square complex exponential method[13] implemented in the LMS software. Only six consecutive modes in the frequency range 24–181 Hz are identified because of the high modal density above 181 Hz as shown by the typical measurement at point 23 ($H_{23}=\dot{w}_{23}/F$) in Fig. 3. In this figure the synthesized response function ($\hat{H}_{23}$) and the least square error ϵ, defined by

$$\epsilon=\frac{|\hat{H}_{23}-H_{23}|^2}{|\hat{H}_{23}|^2}, \tag{1}$$

are plotted in order to validate the modal identification. According to this indicator ϵ, a good agreement between the measurement and the model can be found. Parameters obtained from this modal analysis are shown in Fig. 4. The following conclusions can be drawn for each identified mode.[14]

(i) Mode 1 has no nodes on its mode shape: the modal displacement is close to a global motion of the body depending on its connections to the arm and to the bottom of the pillar.

(ii) Modes 2 and 3 have common characteristics: The axial profiles of soundbox's displacements are similar to the first two mode shapes of a simply supported free beam. Note that as for mode 1, the shapes of modes 2 and 3 do not induce a change in the volume of the cavity: a weak coupling of these modes with the fluid inside the cavity can be expected.

(iii) Modes 4 and 6 have very similar mode shapes. The soundboard's displacement field corresponds to the first bending mode of a quasi-clamped plate. A slight

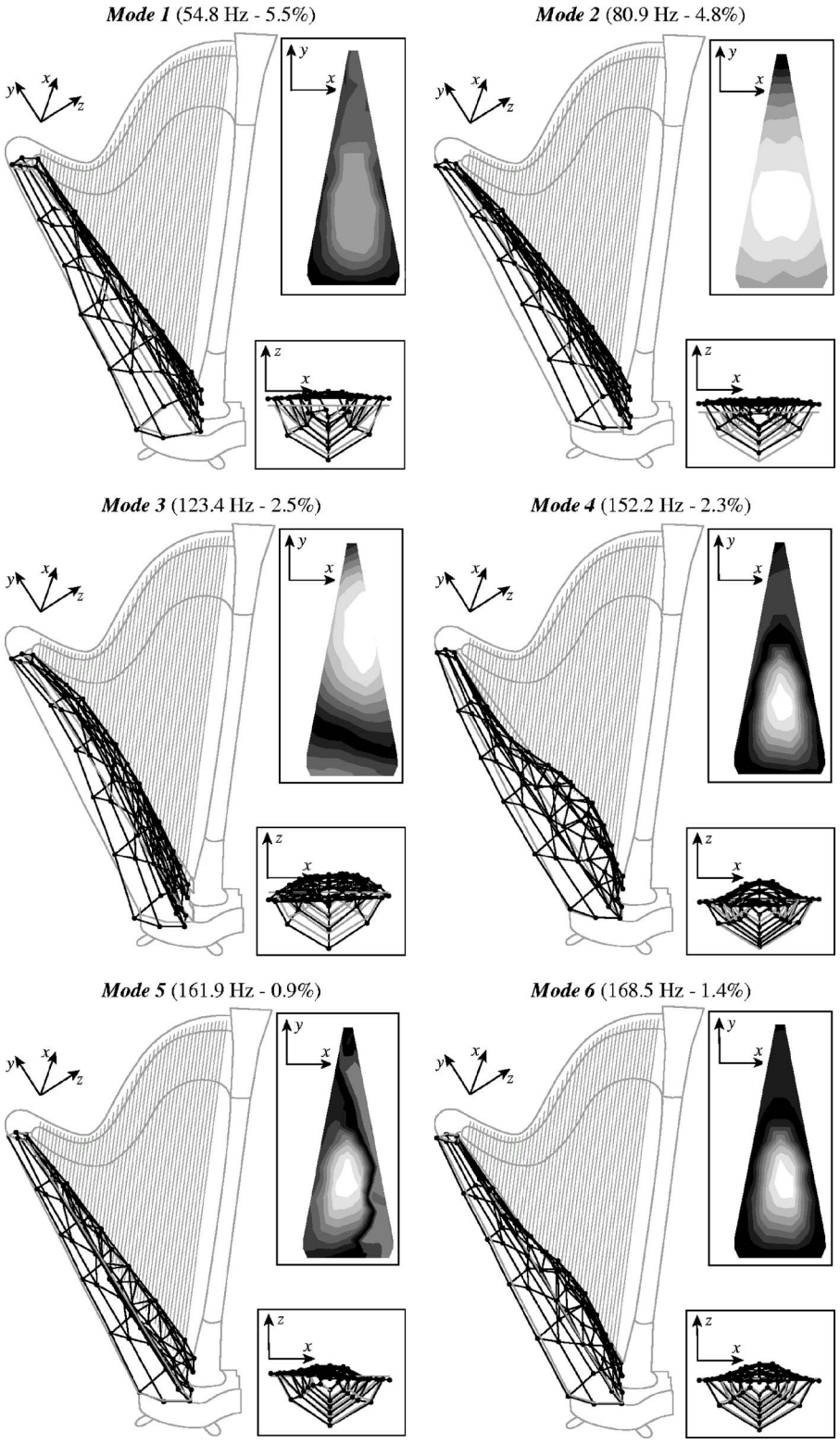


FIG. 4. Eigenfrequencies, damping coefficients and mode shapes of identified modes.

breathing motion of the soundbox is also observed. Shapes of modes 4 and 6 lead to an important change in the volume of the cavity.

(iv) Mode 5 is a pitch mode. In the measured response functions, this mode is not clearly present. It is actually not well excited since the shaker is connected close to the central line of the soundboard, which exactly corresponds to its nodal line. Since the strings are also attached on this nodal line, the role of this mode is not important when the instrument is played. For this reason, it will not be considered afterwards.

The two modes 4 and 6, which have similar shapes, have also been found on an unstrung concert harp[9] and on a strung Spanish harp.[8] However, when the harp is strung it seems difficult[9] to extract these two similar mode shapes. Nevertheless, in our study, these two modes were found in the playing configuration. Moreover, it should be noticed that the dis-

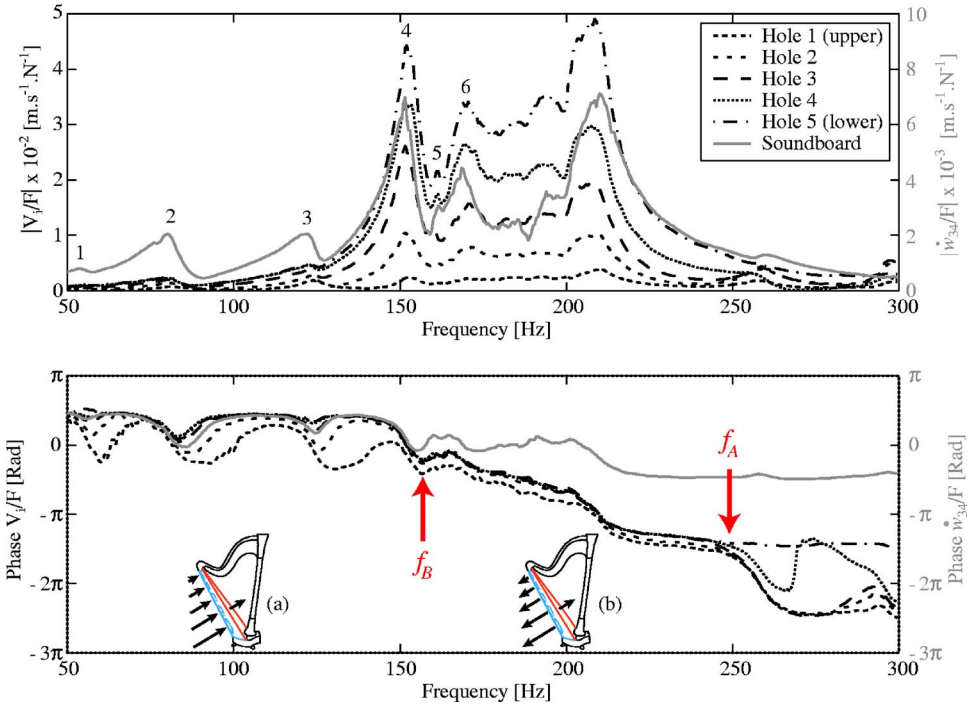


FIG. 5. (Color online) Magnitude and phase of FRF V_i/F (V_i: acoustic velocity in hole i) and of $\dot{w}_{34}/F$ ($\dot{w}_{34}$: velocity at point 34 on the soundboard). Note that magnitude scales for these two kinds of FRF are different. Numbers indicate the modal frequencies given in Fig. 4.

placements of the soundbox have the same order of magnitude as those of the soundboard; this is unexpected because the cavity seems to be much more rigid than the soundboard. Such a result was already mentioned for a Celtic harp.[15]

C. Analysis of the acoustic response functions of the instrument

In order to characterize the acoustic field inside the soundbox, the acoustic velocity in each hole has been measured in the low-frequency range 50–300 Hz, as shown in Fig. 5. Measurements are performed using the two microphones of an acoustic intensity probe. After an accurate calibration of the microphones, the acoustic velocity can be computed from the pressure measured at two points close to each other. Each sound hole can be described as an air piston of which the velocity is measured. It is found that the lower the hole is, the higher the magnitude of the velocity of the piston will be. Since the level of acoustic velocity for the upper hole 1 is far smaller than that of the four others, it cannot be considered as significant and will be ignored afterwards. So, in the studied frequency range the four other air pistons are found to be in phase below f_A (=250 Hz) and are no longer above. Thus, these measurements show that the acoustic field inside the cavity is mostly governed by the first acoustical mode below f_A=250 Hz. Above this particular frequency, other acoustical modes like longitudinal or pipe modes are present.

The mobility at excitation point 34 of the soundboard $\dot{w}_{34}/F$ is also plotted in Fig. 5. Its phase can be compared with the V_i/F phase, V_i being the acoustic velocity measured in hole i. It is found that below a second characteristic frequency, f_B (=160 Hz), the soundboard and all air pistons are in phase. Above this frequency f_B, but below f_A, the phase difference between FRF $\dot{w}_{34}/F$ and FRF V_i/F increases from 0° to 180°. This shows that in the frequency range f_A–f_B, the soundboard and the air pistons are out-of-phase. These particular phase relationships are schematically represented in Fig. 5 by arrows in harp drawings (a) and (b). The direction and length of the arrows that are plotted in these diagrams represent the phase and the magnitude of the velocity of the soundboard and of the air pistons below and above f_B.This result had already been found on another Camac concert harp in a previous paper[16] where the characteristic frequency f_B was found to equal 175 Hz.

In order to find out the implication of the acoustic field inside the soundbox on the acoustic far field of the instrument, we investigate the acoustic pressure around the concert harp. The pressure is measured in an anechoic room by 32 microphones regularly placed around the harp on a 2.35 m radius circle at 1.2 m in height. In the frequency range 50–220 Hz, the directivity patterns are found to be nondirectional, as shown in Fig. 6 for two selected frequencies corresponding to the eigenfrequencies of modes 4 and 6. The acoustic transfer function P_C/F measured in front of the harp (at the point labeled C defined in Fig. 6) is also shown in Fig. 7. As for afterwards measurements, the shaker used for the excitation is connected exactly on the central line of the soundboard and the acoustic effect of the pitch mode is then canceled out. In Fig. 7, we note that for a same force applied by the shaker, the acoustic pressure is much more important in the range 140–230 Hz than in the rest of the studied frequency range. Therefore, in a playing configuration, the harp seems to radiate the sound more efficiently in the range 140–230 Hz. The first two important peaks of acoustic pressure correspond to the eigenfrequencies of modes 4 and 6. Moreover, modes whose eigenfrequencies are above 200 Hz cannot be individually distinguished and their contribution to the response below 200 Hz is probably not negligible.

To conclude, six structural modes have been identified in the low-frequency range. Among these six modes, two play

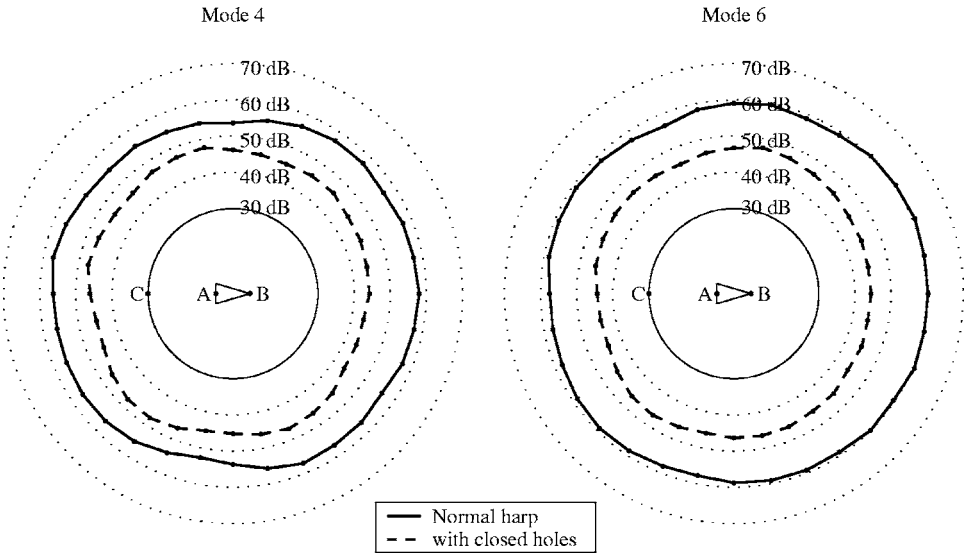


FIG. 6. Directivity patterns for modes 4 and 6 in two configurations: normal and with all holes closed by stoppers. Points A and B correspond to the bottom of the pillar and to the top of the soundboard respectively. Point C is the measurement point in front of the harp.

an important part in the sound radiated by the concert harp. They lead to a strong acoustic radiation, associated to a nondirectional directivity pattern.

III. IDENTIFICATION OF A0 AND T1 MODES OF THE CONCERT HARP

A. Frequency response functions of a modified instrument

Considering only frequency response measurements on the instrument's body, modes 4 and 6 have similar mode shapes (see Fig. 4). However, the air piston motions are different for these two modes. To identify the nature (A0 or T1) of modes 4 and 6, the study of frequency response functions of a slightly modified instrument is performed. Mobilities measured at point 34 on the soundboard for three different configurations are compared with the normal configuration. The amplitude and frequency shifts of peaks are shown in Fig. 8 and in Table II.

The first modification consists of closing the sound holes of the concert harp as shown in Fig. 9 and labeled (1). Those are closed by using stoppers made with small tar plates. This configuration prevents all fluid motions inside the sound holes. This modification has heavily affected the instrument. Eigenfrequencies of modes 1, 2, and 3 undergo a shift of approximately −2 Hz due to the additional mass loading induced by the stoppers. Two additional peaks are seen below 200 Hz and in the rest of the frequency range the level is lower than in the normal configuration. The peak for mode 6

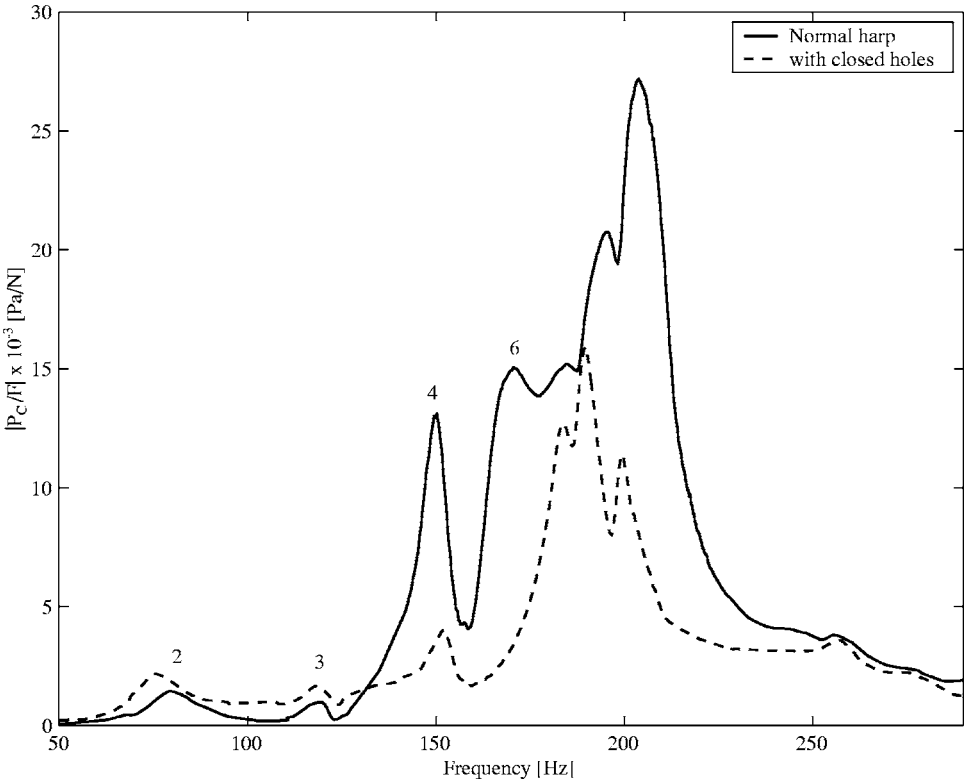


FIG. 7. Frequency response function at point C defined in Fig. 6. Numbers indicate the modal frequencies given in Fig. 4.

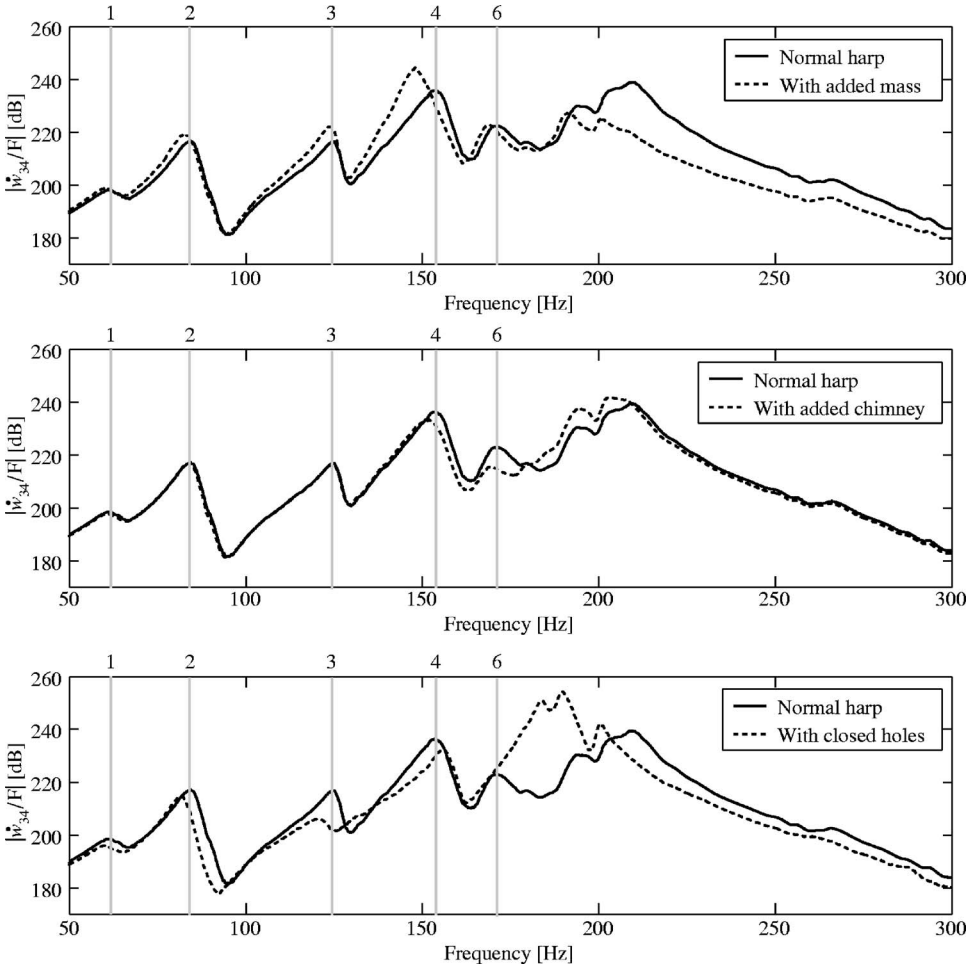


FIG. 8. Magnitude of the FRF $\dot{w}_{34}/F$ on the soundboard depending on four configurations: normal, mass added on the soundboard, chimney inserted in the lower hole, and with all holes closed. Numbers associated to vertical lines indicate the modal frequencies given in Fig. 4.

is no longer distinct. This result is also confirmed by the acoustic pressure measurement in front of the harp as shown in Fig. 7 when the sound holes are closed. This observation can be interpreted by the fact that the resonance of the open cavity does not exist anymore (see Sec. II B).

The second modification (2) consists of inserting a 2-cm-high chimney in the lower hole of the harp as shown in Fig. 9. This change induces an increase of the mass of the first air piston. Only two structural modes are affected: modes 4 and 6. This proves that these two modes are coupled to the fluid inside the cavity. The other modes are weakly coupled to the air cavity and do not participate in the acoustic response function as shown in Fig. 7.

The third modification (3) consists of adding a mass (m=200 g) on both sides of the central line of the soundboard as shown in Fig. 9. All eigenfrequencies of structural modes are lowered but modes 4 and 6 more than modes 1, 2, and 3. This is probably due to the fact that the mass is located on the maximum displacement area of these two modes.

TABLE II. Resonance frequencies for the first six modes according to four configurations of the instrument: normal (f), mass loaded on the soundboard (f_m), chimney inserted in the lower hole (f_c), and holes closed (f_{cl}).

	Resonance frequencies (Hz)				Deviations (Hz)		
Modes	f	f_m	f_c	f_{cl}	$f-f_m$	$f-f_c$	$f-f_{cl}$
1	61.5	60	61.5	60	1.5	0	1.5
2	84.5	82	84.5	81.5	2.5	0	3
3	124.5	123.5	124.5	120.5	1	0	4
4	153.5	148.5	152	156	5	1.5	2.5
6	172	168.5	169	…	3.5	3	…

B. Discussion

The most important effects of the modifications (1), (2), and (3) on modes 4 and 6 can be summarized as follows: on one hand, when sound holes are closed, mode 6 disappears. On the other hand, when the mass of the soundboard is increased, the eigenfrequency of mode 4 is lowered whereas the eigenfrequency of mode 6 is nearly stable. When the mass of the air pistons is increased, the eigenfrequency of mode 6 is lowered whereas the eigenfrequency of mode 4 undergoes smaller modifications. By considering these experimental results, it can be concluded that modes 4 and 6 involve a coupling between the bending motion of the soundboard mode and the oscillation of the air piston. These two modes can respectively be labeled, with the common notation, T1 and A0.

The fact that the A0 mode is present in the instrument's response clearly depends on the modal density and on the damping coefficients of the acoustical and structural modes. For some configurations, these parameters are such that the contribution of the A0 mode can be a minor one.[17] In our configuration, although sound holes are designed to ease the

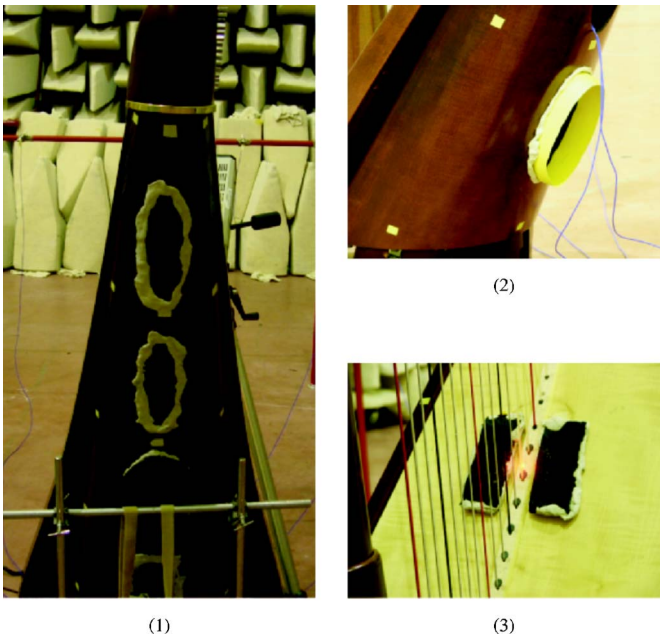


FIG. 9. (Color online) Different configurations of the modified instrument: (1) with all holes closed, (2) with chimney inserted in the lower hole, and (3) with mass added on the soundboard.

string mountings, they are found to have a significant influence on the vibroacoustic response of the concert harp. This is confirmed by the measurements of the far field acoustic pressure performed with opened and closed sound holes, as shown in directivity patterns (Fig. 6) and in the frequency response functions (Fig. 7).

Contrary to the violin and to the guitar, the A0 mode is above rather than below the frequency of the lowest acoustically significant structural mode T1. The concert harp does not take advantage of the soundbox resonance to increase its sound radiated in low frequencies, below the T1 mode.

Since these two modes, T1 and A0, are dominant in the low-frequency range, the response of the instrument can be approximated by using a two degrees of freedom oscillator model in which one degree is due to the soundboard and the other one to the fluid inside the cavity as it was done for the guitar.[4]

IV. CONCLUSION

This paper deals with the vibroacoustic behavior of a concert harp in the low-frequency range. The nature of the modes of the soundbox coupled to the internal fluid is investigated.

A classic experimental modal analysis has permitted the identification of six modes in the frequency range 24–181 Hz. Since the modal density increases with the frequency, mode identification at higher frequencies was not possible. Among the six identified modes, four correspond to global motions of the soundbox, which do not induce a change in the volume of the cavity and are thus weakly coupled to the internal acoustic field. These modes, which mostly depend on the characteristics of the connection of the soundbox to the arm and to the bottom of the pilar, lead to the weakening of the acoustic radiation. The two remaining modes, called T1 and A0, play an important acoustic role and have the following characteristics. (1) They are associated to coupled motions of the bending vibration of the soundboard and to the oscillations of the air pistons located in the sound holes. They correspond to the first two modes of a Helmholtz resonator with yielding walls. The labels T1 and A0 were used for the guitar and the violin for which this Helmholtz effect is known. (2) Modes T1 and A0 lead to important acoustic radiation: the acoustic pressure radiated by the harp takes high values in the range 140–230 Hz and the first two peaks of the pressure amplitude correspond to the resonance frequencies of T1 and A0. (3) The mode shapes of T1 and A0 are such that the displacement of the air pistons located in the five holes are all in phase. For T1, the displacements of these pistons are approximately in phase with the bending displacement of the soundboard. For A0, these motions are approximately out of phase. (4) Contrary to the violin and to the guitar, the A0 mode is above rather than below the frequency of the lowest acoustically significant structural mode T1. Thus, the concert harp does not take advantage of the soundbox resonance to increase its sound radiated in low frequencies. However, the study reveals the importance of the contribution of mode A0 in the response of the instrument, confirming the importance of the coupling between the soundboard and the cavity. This result is valid for the studied harp: *Atlantide Prestige* concert harp. Future works may concern others harps with different characteristics on which the eigenfrequencies of modes A0 and T1 depend: cavity volume, sound holes sizes, and soundboard material.

ACKNOWLEDGMENTS

The authors acknowledge financial support from the CNRS and the Région des Pays de la Loire for Jean-Loïc Le Carrou's PhD scholarship. They also acknowledge the instrument maker CAMAC Harps for the lending of the concert harp and the CTTM (Le Mans Centre for Technology Transfer) where measurements were performed.

[1]N. H. Fletcher and T. D. Rossing, *The Physics of Musical Instruments*, 2nd ed. (Springer, New York, 1998).
[2]I. M. Firth, "Physics of the guitar at the Helmholtz and first top plate resonances," J. Acoust. Soc. Am. **61**, 588–593 (1977).
[3]J. A. Moral and E. V. Jansson, "Eigenmodes, Input Admittance, and the Function of the Violin," Acustica **50**, 329–337 (1982).
[4]O. Christensen and B. B. Vistisen, "Simple model for low-frequency guitar function," J. Acoust. Soc. Am. **68**, 758–766 (1980).
[5]O. Christensen, "Quantitative models for low frequency guitar function," J. Guitar Acoust. **6**, 10–25 (1982).
[6]L. Cremer, *The Physics of the Violin* (MIT, Cambridge, MA, 1984).
[7]I. M. Firth, "On the acoustics of the harp," Acustica **37**, 148–154 (1977).
[8]I. M. Firth, "Harps of the baroque period," J. Catgut Acoust. Soc. **1**(3), 52–61 (1989).
[9]A. J. Bell, "An acoustical investigation of the Concert Harp," Ph.D. dissertation, University of St Andrews, UK, 1987.
[10]A. J. Bell, "The Helmholtz resonance and higher air modes of the harp soundbox," J. Catgut Acoust. Soc. **3**(3), 2–8 (1997).
[11]A. Le Pichon, "Méthode de prediction du rayonnement acoustique de structures volumiques composées d'une ou de plusieurs faces planes vibrantes, application aux instruments de musique à cordes (Prediction method of acoustical radiation of volumic structures composed of one or many vibrating faces, application to string instruments)," Ph.D. disserta-

tion, University of Paris XI, Paris, F, 1998 (in French).
[12]J-L. Le Carrou, F. Gautier, N. Dauchez, and J. Gilbert, "Modelling of sympathetic string vibrations," Acta. Acust. Acust. **91**, 277–288 (2005).
[13]D. J. Ewins, *Modal Testing: Theory and Practice* (Wiley, Somerset, England, 1994).
[14]J-L. Le Carrou, F. Gautier, and N. Dauchez, "Acoustic radiation of the concert harp in the low frequency range," in Proc. ICSV12 2005, Lisbon, Portugal (2005).
[15]G. Kergoulay and E. Balmès, "Dynamic behaviour of a harp soundboard and soundbox," in Proc. SCI 2001, Orlando, FL (2001).
[16]F. Gautier and N. Dauchez, "Acoustic intensity measurement of the sound field radiated by a concert harp," Appl. Acoust. **65**, 1221–1231 (2004).
[17]B. E. Richardson, "Stringed instruments: plucked," Encyclopedia Acoust., 1627–1634 (1997).

Investigation of an anisotropic tortuosity in a Biot model of ultrasonic propagation in cancellous bone

Elinor R. Hughes, Timothy G. Leighton,[a] and Paul R. White
Institute of Sound and Vibration Research, University of Southampton, Southampton, SO17 1BJ, United Kingdom

Graham W. Petley
Department of Medical Physics and Bioengineering, Southampton University Hospitals NHS Trust, Tremona Road, Southampton, SO16 6YD, United Kingdom

(Received 10 June 2006; revised 12 October 2006; accepted 12 October 2006)

The modeling of ultrasonic propagation in cancellous bone is relevant to the study of clinical bone assessment. Historical experiments revealed the importance of both the viscous effects of bone marrow and the anisotropy of the porous microstructure. Of those propagation models previously applied to cancellous bone, Biot's theory incorporates viscosity, but has only been applied in isotropic form, while Schoenberg's anisotropic model does not include viscosity. In this paper we present an approach that incorporates the merits of both models, by utilizing the tortuosity, a key parameter describing pore architecture. An angle-dependent tortuosity for a layered structure is used in Biot's theory to generate the "Stratified Biot Model" for cancellous bone, which is compared with published bone data. While the Stratified Biot model was inferior to Schoenberg's model for slow wave velocity prediction, the proposed model improved agreement fast wave velocity at high propagation angles, particularly when sorted for porosity. An attempt was made to improve the fast wave agreement at low angles by introducing an angle-dependent Young's Modulus, which, while improving the agreement of predicted fast wave velocity at low angles, degraded agreement at high angles. In this paper the utility of the tortuosity in characterizing the architecture of cancellous bone is highlighted. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2387132]

PACS number(s): 43.80.Cs, 43.80.Qf, 43.20.Jr [CCC] Pages: 568–574

I. INTRODUCTION

The skeleton contains two types of calcified tissue: cortical and cancellous. Cortical bone is dense and compact, while cancellous bone is a porous network of calcified "trabeculae," filled with fatty bone marrow (Fig. 1). There is much variety in the trabecular arrangement throughout the skeleton, and hence in the mechanical properties of cancellous bone, which may be isotropic at some sites (e.g., femoral head), or highly anisotropic at others (vertebrae, femur).[1]

Osteoporosis is a skeletal disease that reduces bone density and erodes the trabecular microstructure, leading to increased bone fragility. While a reduction in density can be assessed by Dual X-Ray Absorptiometry,[2] microstructure may potentially be examined using ultrasonic techniques. Quantitative Ultrasound[2] (QUS) analyses information from ultrasonic signals traveling through a skeletal site, often a site containing cancellous bone, such as the calcaneus. However, QUS parameters are not firmly linked to physical parameters, such as bone strength or porosity, other than through statistical means. Current research aims to establish such relations through a validated predictive propagation model. Biot's theory, and Schoenberg's theory for propagation in porous media, have both been applied to the problem.

Biot's theory[3] of wave propagation in a porous solid saturated with fluid, predicts two compressional waves (fast and slow waves) and a shear wave. The theory was first applied to ultrasonic propagation in cancellous bone in 1991 by McKelvie and Palmer,[4] and has since been used by several authors, with varying degrees of success.[5–12] Two compressional waves in cancellous bone at ultrasonic frequencies were first observed by Hosokawa and Otani,[8] and later confirmed by the present authors.[9] Further workers have applied modified Biot models to the problem.[10,11] One disadvantage of Biot's theory is that it requires knowledge of over a dozen input parameters, many of which are not easily be evaluated for biological tissue.

The present authors used Schoenberg's theory[9] to predict the well-documented anisotropic behavior of ultrasonic properties in cancellous bone.[13–15] By imagining the microstructure of highly oriented cancellous bone as an array of bone-marrow layers, the use of Schoenberg's theory for a stratified composite medium provided the first anisotropic model, giving qualitative agreement with measured fast and slow wave phase velocities. One disadvantage of the Schoenberg model lies in its simplicity; for example, it takes no account of fluid viscosity. Replacing interstitial marrow by another fluid, such as water, is known to affect wave properties in cancellous samples,[16] owing to a difference in viscosity ($\eta_{\text{marrow}} = 0.15$ Pa s; $\eta_{\text{water}} = 0.0001$ Pa s, at 20 °C). Therefore any propagation model for bone should also account for this factor.

In this paper we present a simple approach that incorporates the merits of both Biot's theory and Schoenberg's

[a]Electronic mail: T.G.Leighton@soton.ac.uk

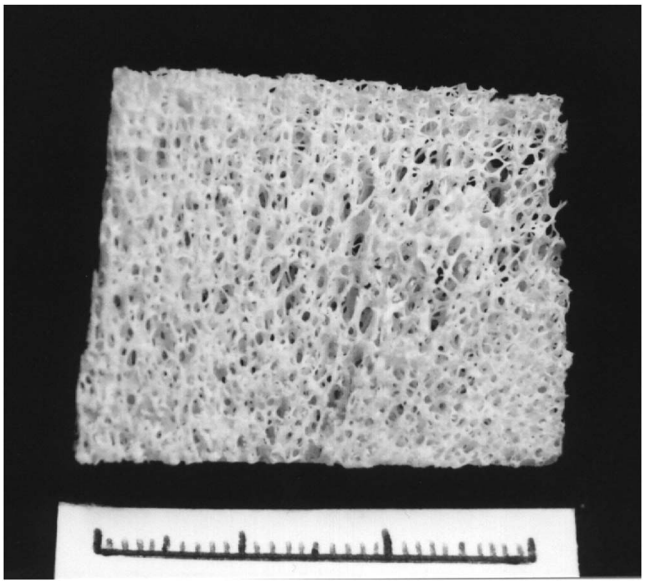


FIG. 1. The porous microstructure of cancellous bone from a bovine femur, showing a centimeter scale.

theory. This is done by introducing anisotropy into Biot's theory through an angle-dependent tortuosity for layers, to create a "stratified Biot" model. The performance of the models is compared with each other and with experimental data. Analysis is limited to the consideration of phase velocity, owing to complex problems in the comparison of predicted absorption with measured signal loss.[9]

II. BIOT'S THEORY

Biot's theory[3] considers acoustic wave propagation in an isotropic porous elastic solid (of density, ρ_s; moduli, E_s, K_s, and N; porosity, β), saturated with a viscous fluid (of density, ρ_f; moduli, E_f and K_f; and viscosity η). The theory is valid for frequencies where the wavelength is large relative to the size of the discontinuities. For a harmonic pressure wave, phase velocity and attenuation are found from

$$V_{\text{fast,slow}} = \sqrt{\frac{\Delta \pm [\Delta^2 - 4(PR - Q^2)(\rho_{11}\rho_{22} - \rho_{12}^2)]^{1/2}}{2(\rho_{11}\rho_{22} - \rho_{12}^2)}}, \tag{1}$$

$$V_{\text{shear}} = \sqrt{N/[(1-\beta)\rho_s + (1 - 1/\alpha)\rho_f]}, \tag{2}$$

where $\Delta = P\rho_{22} + R\rho_{11} - 2Q\rho_{12}$. The terms A, Q, P, and R are generalized elastic coefficients that are defined in terms of the bulk moduli of the solid frame, K_b, the solid material, K_s, and the fluid, K_f, which can be found from standard expressions for an isotropic medium. The mass density terms, ρ_{ij}, are defined in terms of porosity and density, as given in the literature.[3] Equation (1) governs the propagation of two compressional waves, known as fast and slow waves, and Eq. (2) describes shear wave propagation. The fast wave is a bulk wave where the fluid and solid are locked together through viscous or inertial coupling, where as the slow wave corresponds to there being a relative motion between the fluid and the solid.

Below a specific "critical frequency" viscous coupling dominates, allowing fast wave propagation, but impeding the slow wave, which becomes diffusive. Above the critical frequency, viscous coupling weakens and inertial coupling dominates. In cancellous bone *in vitro*, the critical frequency is typically below[17] 1 kHz, hence at ultrasonic frequencies of interest, inertial coupling dominates.

The extent of inertial coupling, ρ_{12}, depends on the geometry of the porous solid, as described by

$$\rho_{12} = -(\alpha - 1)\beta\rho_f, \tag{3}$$

where α is the tortuosity, a crucial parameter in Biot's theory. It represents the squared ratio of the mean pathlength through the porous frame to that of the direct path. In a nonviscous fluid, the tortuosity is real-valued, being between 1 (for $\beta = 1$), and tending to infinity, as $\beta \to 0$. It may be defined for a specific porous geometry, noted here as α_∞. For example, for a geometry of fused spheres,[18] the tortuosity, α_∞, is

$$\alpha_\infty = 1 - s\left(1 - \frac{1}{\beta}\right), \tag{4}$$

where the structure factor, $s = 1/2$. Also, for a medium of cylindrical tubes,[19] at a propagation angle θ to the pores, the tortuosity is

$$\alpha_{\text{tubes}} = 1/\cos^2\theta, \tag{5}$$

which equals unity for propagation along the pores.

The importance of α_∞ is apparent when considering propagation in porous media having a rigid frame ($K_b \gg K_f$, $K_s > N$), where the velocity of the slow wave is[20] $V_{\text{slow}} = V_{\text{fluid}}/\sqrt{\alpha_\infty}$.

Equation (3) may be written to account for frequency-dependent effects, as

$$\rho_{12}(\omega) = -[\alpha(\omega) - 1]\beta\rho_f, \tag{6}$$

where $\alpha(\omega)$ is the dynamic tortuosity. Johnson *et al.*[20] defined $\alpha(\omega)$ in terms of measurable parameters as

$$\alpha(\omega) = \alpha_\infty + j\eta\beta\left(1 - \frac{4j\alpha_\infty^2 k_0^2 \rho_f \omega}{\beta^2 a^2}\right) \Big/ \omega\rho_f k_0, \tag{7}$$

where a is a pore size parameter (often the pore radius), and k_o is the permeability for static conditions.

The geometric tortuosity α_∞ in Eq. (7) can be evaluated by different methods. First, the "electrical conductivity" method involves measuring the ratio of the conductivity of saline alone to a sample saturated with saline, F, and using the relationship $\alpha_\infty = F\beta$. Tortuosity may also be measured using the nonviscous superfluid ^{4}He in the pores[20] or by a slow wave method, involving relation $V_{\text{slow}} = V_{\text{fluid}}/\sqrt{\alpha_\infty}$ for rigid frame media. The reflection and transmission of pulses of ultrasonic[21] and audio[22] frequencies have also been used to deduce the tortuosity and porosity of air-filled rigid frame media.

III. TORTUOSITY IN CANCELLOUS BONE

The tortuosity of cancellous bone has been defined and evaluated in different ways. Table I summarizes some results

TABLE I. Input parameters for the stratified Biot and Schoenberg models.

Authors	Bone type	Exp. Tortuosity	Method
Williams *et al.*, 1996	Human calcaneus	1.6 for a=0.5 mm	Electrical conductivity
Lauriks *et al.*, 1994	Bovine	1.26 to 2.64 (various porosities)	Electrical conductivity
Strelitzki *et al.*, 1999	Human calcaneus	1.04	Transmitted slow wave, at 400 kHz
Fellah *et al.*, 2004	Human femur	1.02 to 1.05	Reflected slow wave method at 2.25 MHz
Attenborough *et al.*, 2005	Phantoms of femoral head	1.3 to 1.8	Reflected slow wave method at 1 kHz
	calcaneus	1.1 to 1.7	

of tortuosity in cancellous bone, with values varying between 1.01 and 2.6. Initial consideration of Table I suggests that 'slow wave' methods at ultrasonic frequencies yield lower results when compared with data taken at audio frequencies or by the electrical conductivity method. This may be because ultrasonic tests are more likely to be affected by scattering. Although the data in Table I does not allow skeletal sites to be classified by tortuosity, Attenborough *et al.*[22] recently demonstrated that average tortuosity differs between sites, with highly oriented microstructures (lumbar spine) yielding tortuosities closer to unity than those less oriented sites (e.g., femoral head). More significantly, tortuosity was orthogonally anisotropic.

In propagation models, some workers[5,8,12] used the purely geometric definition of Eq. (4), while others applied the dynamic tortuosity of Eq. (7).[6,7,9,11] In Biot simulations, Fellah *et al.*[11] demonstrated that a 20% increase in tortuosity produced a 10% decrease in wave velocities, a 300% increase in the fast wave amplitude, and a 60% decrease in the slow wave amplitude, highlighting the sensitivity of wave properties to this factor.

Using the tortuosity to characterize the cancellous structure has significant potential, since it may be able to describe architecture with a greater value and significance than parameters such as porosity or pore size. Recognizing the angular dependence of the tortuosity, with respect to trabecular orientation, is the subject of the remainder of this paper, after a discussion of Schoenberg's plate model.[23]

IV. SCHOENBERG'S THEORY FOR STRATIFIED MEDIA

In Schoenberg's theory,[24] for propagation in periodically alternating fluid-solid layers, strata are parallel to x and y directions, with spatial period, H, in the z direction (Fig. 2) and porosity, β. The solid is isotropic and elastic, with density ρ_s, compressional speed V_s, and shear speed, V_{sh}, while the fluid is ideal, nonviscous, with sound speed V_f.

Wave propagation is expressed in terms of the slowness vector, $\mathbf{s}=(s_x,s_y,s_z)$ and the components parallel to the layers, s_x, and normal to the layers, s_z, (x and y are interchangeable) are related by

$$(s_z^2/\langle\rho\rangle)-\left(\frac{\beta(V_f^{-2}-S_x^2)}{\rho_f}+\frac{(1-\beta)(V_s^{-2}-s_x^2)}{\rho_s(1-V_{\mathrm{pl}}^2 s_x^2)}\right)=0, \qquad (8)$$

where $\langle\rho\rangle=\beta\rho_f+(1-\beta)\rho_s$, and V_{pl} represents the plate speed in the solid

$$V_{\mathrm{pl}}=2\left(1-\frac{V_{\mathrm{sh}}^2}{V_s^2}\right)^{1/2}V_{\mathrm{sh}}. \qquad (9)$$

Phase velocity is the inverse of the magnitude of the slowness vector, $|\mathbf{s}|^{-1}$, at an angle $\theta=\tan^{-1}(s_z/s_x)$, where 0° is normal to the layers, and 90° is parallel to the layers. Schoenberg's theory predicts two compressional waves equivalent to the fast and slow waves of Biot's theory. Their phase velocities vary with propagation angle, owing to an anisotropic inertial coupling, which tends to infinity at 0° (normal to the layers where $\alpha_\infty\to\infty$), varying to zero at 90° (parallel to the layers where $\alpha_\infty=1$).

In *in vitro* studies, Schoenberg's theory gave qualitative agreement with measured data in bovine cancellous bone[9] and further success was reported by other authors[25–27] for various bone types.

While Schoenberg offers insight into the role of inertial coupling in layers, the theory clearly oversimplifies the cancellous structure. In addition, the omission of fluid viscosity in Schoenberg's model prevents it from accounting for viscous absorption and restricts its application to Biot's "high" frequency region. Biot's theory, while accounting for viscous effects, has only been applied to cancellous bone in isotropic form. In the following section we develop a model that incorporates viscous effects and anisotropic effects into a Biot model in a simple and straightforward manner.

V. AN ANISOTROPIC BIOT MODEL FOR CANCELLOUS BONE

The angular variation in wave properties with respect to trabecular direction originates from two characteristics of

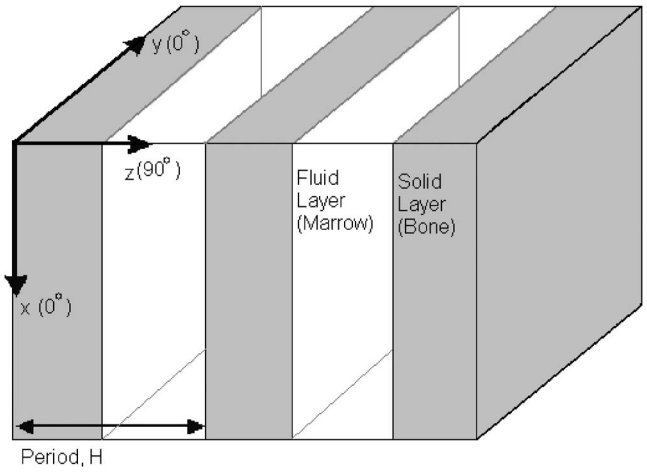


FIG. 2. Idealization of the structure of cancellous bone as an array of parallel bone-marrow layers.

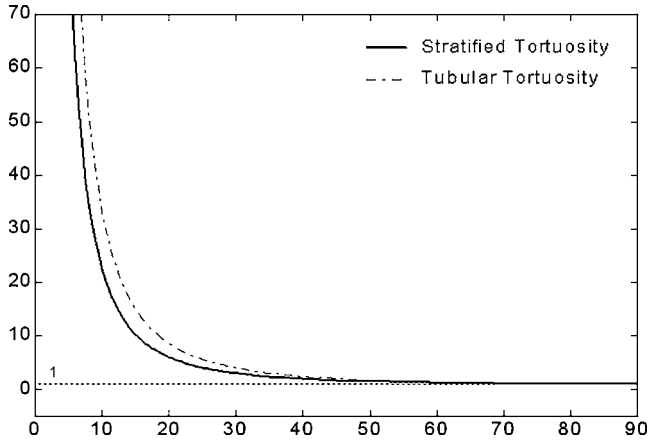


FIG. 3. Predicted stratified tortuosity of Eq. (20) (solid line), and tubular tortuosity, $\alpha_{\text{tubes}}=1/\cos^2\theta$ (dashed line), versus the angle of propagation. The angle 0° is normal to the stratification.

cancellous bone: the mechanical properties of the bone matrix and the motion of the fluid within irregularly shaped pores. Previous models of propagation in anisotropic porous media have either been extremely generalized[28] and therefore complex, or specific to particular conditions of structure, frequency or constituent media.[29] The current approach introduces a pragmatic way of modeling an anisotropic stratified structure within the confines of a Biot model. This is done by using an angle-dependent tortuosity defined for layers of periodic bone-marrow plates, and inserting this back into Biot's theory. This approach was named the "Stratified Biot model," following the work of Schoenberg and Sen.[30]

The dependence of velocity on the direction in a stratified model reflects the variation in inertial coupling. The degree of inertial coupling within Schoenberg's layers may be assumed to be equivalent to that occurring in an arbitrary anisotropic porous medium as described by Biot's theory. Therefore, describing a layered structure in Biot's theory requires finding an expression for the tortuosity for a simple stratified structure as a function of the angle of propagation. Such an expression may be obtained by equating the compressional phase velocity in terms of α from Biot's theory [Eqs. (1)], with that from Schoenberg's model in terms of propagation angle, θ [Eqs. (8) and (9)]. In the first instance, this calculation may be substantially simplified if shear in the solid is neglected ($K_b=N=0$ and $PR-Q^2=0$). While effectively yielding a fluid-fluid system, this is a fruitful starting point in developing a tortuosity for a stratified medium. The tortuosity as a function of propagation angle, $\alpha_{\text{layers}}(\theta)$, can then be found as

$$\alpha(\theta) = 1 + [(1-\beta)\rho_s/\langle\rho\rangle]\cot^2\theta. \tag{10}$$

Equation (10) may then be inserted back into Biot's theory, substituting for α_∞, to predict wave properties that vary with orientation. This approach will be known for the remainder of this paper as the "Stratified Biot" model.

Figure 3 shows the stratified tortuosity of Eq. (10) plotted as a function of angle, using values from Table II. The angle $\theta=0°$ is normal to the layers and $\theta=90°$ is parallel to the layers. For comparison, the tortuosity for an array of cylinders from Eq. (5) is plotted, with angle convention dictating correspondence between the dominant axes of both oriented systems (i.e., $\theta=0°$ being normal to layers and tubes, Fig. 3). Figure 3 shows that both stratified and tubular tortuosities tend to infinity as $\theta\to0°$, and to unity as $\theta\to90°$. The two curves show a marked similarity in shape, however, the stratified tortuosity is a function of porosity, which thus provides added power to model changes in structure over Eq. (5).

TABLE II. Input parameters for the stratified Biot and Schoenberg models.

Parameter	Value	Source
Density of Solid Bone, ρ_s	1960 kg/m^3	Williams, 1992
Density of Marrow, ρ_f	990 kg/m^3	Duck, 1990
Young's Modulus of Bone, E_s	20 GPa	Williams, 1992
Bulk Modulus of Marrow, K_f	2.2 GPa	Hosokawa and Otani, 1997
Poisson's ratio of solid, ν_s	0.32	Williams, 1992; Duck 1990
Poisson's ratio of frame, ν_b	0.32	Williams, 1992
Porosity, β	0.65–0.82	By experiment
Power index, n	1.23	Williams, 1992
Viscosity of marrow, η	0.04 Pa s	Bryant *et al.*, 1989
Pore radius, a	5×10^{-3} m	By experiment
Permeability, k_0	5×10^{-9} m^3	McKelvie and Palmer, 1991
Frequency, f	1 MHz	By experiment
Solid compressional speed, V_s	3200 m/s	Williams, 1992
Fluid compressional speed, V_f	1500 m/s	Estimate
Shear speed, V_{sh}	1800 m/s	Wu and Cubberly, 1997 (Ref. 32)

VI. RESULTS FOR CANCELLOUS BONE

Figure 4 shows the predictions of fast and slow wave velocities versus angle with respect to the structural orientation for the Stratified Biot model and Schoenberg's theory. The input parameters are listed in Table II. Stratified Biot fast and slow wave phase velocities were found from Eq. (1), using Eq. (10). Schoenberg's velocities were predicted as a function of the propagation angle, using Eqs. (8) and (9). The angle 0° is normal to the layers and 90° is parallel to the layers.

Each model predicts a pair of velocity contours that vary with angle. The fast wave corresponds to the upper curve of

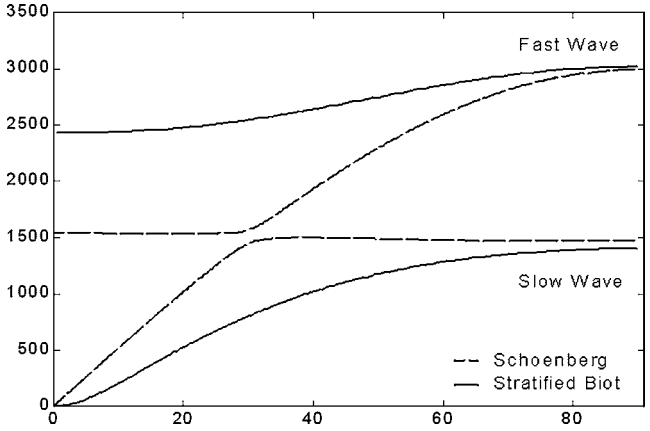


FIG. 4. Predicted fast and slow wave phase velocities for the stratified Biot model (solid line) and the Schoenberg model (dashed line). Angles 0° and 90° are normal and parallel to the stratification, respectively.

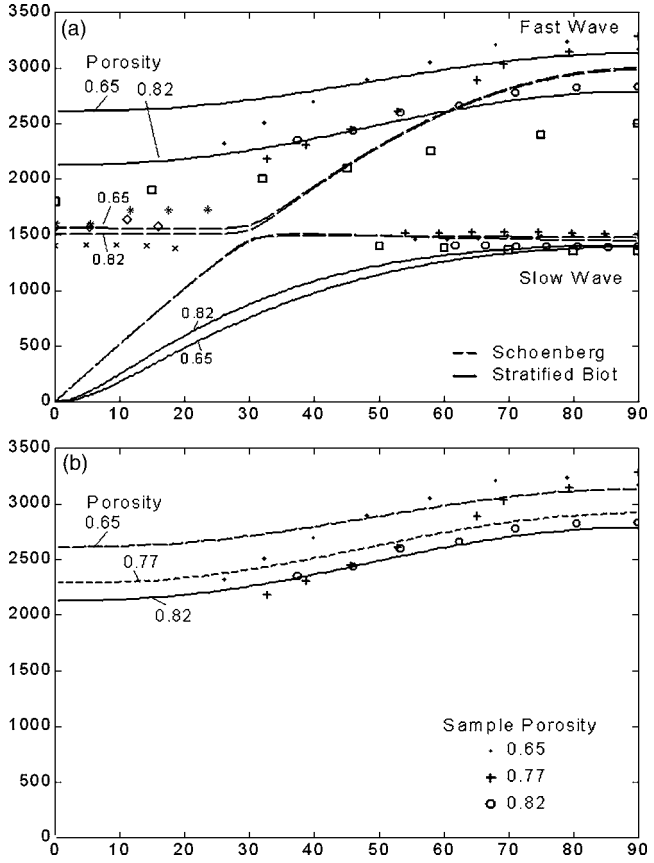


FIG. 5. (a) Predicted fast and slow wave velocities versus the angle of propagation for the stratified Biot model (solid lines) and the Schoenberg model (dashed lines). The angle 0° is normal to layering. Models are evaluated at porosities of 0.82 and 0.65, corresponding to the range within samples. Data from Hughes *et al.* (Ref. 6) (·, o, +, ∗, x, ◇), and Hosokawa and Otani (Ref. 16) (□). (b) Predicted fast wave velocities versus angle of propagation for the Stratified Biot model for input porosities of 0.65, 0.77, and 0.82 [data from Hughes *et al.* (Ref. 6)].

the pair, while the slow wave corresponds to the lower of the pair. Predictions from Schoenberg's theory (dashed lines) show a characteristic piecewise linear shape over most of the angular range, with the fast wave curve becoming nonlinear above 60°. By contrast, predictions from the stratified Biot model (solid lines) vary less sharply, converging with Schoenberg velocities at 0° and 90° for the slow wave, and 90° for the fast wave. A significant discrepancy between the models is seen in the fast wave, disagreeing by almost 50% at 0°. It should be noted that a degree of disagreement between the two approaches is to be expected, because the inclusion of anisotropy and viscosity in the stratified Biot model, for interpenetrating fluid and solid phases, will not make it identical to the Schoenberg model, for infinitely extending uninterrupted plates. Perhaps a better test of performance is to compare it with experiment.

In Figs. 5(a) and 5(b), previously published data from ultrasonic measurements of bovine femur[9,15] are compared with the predictions of Schoenberg (dashed lines) and the Stratified Biot model (solid lines). Full details of the experimental methods may be found in work by Hughes *et al.*[9] and Hosokawa and Otani.[15] Data from Hughes *et al.* was taken from six marrow-filled bovine femur samples of porosity between $\beta=0.65$ and 0.82, and subject to errors in the measurement of thickness, transducer separation, and insonation angle. The intersample precision was 6.0%. The porosity of the bovine femur sample tested by Hosokawa and Otani was $\beta=0.82$.

Figures 5(a) and 5(b) demonstrate the reproducible result that measured fast and slow wave phase velocities depend on the angle of propagation within cancellous samples. The fast wave velocity increases with angle, from around 2000–2200 m/s at 30° to between 2800–3200 m/s at 90° in Hughes' data, and from 1800 m/s at 0° to 2500 m/s at 90° in the data of Hosokawa and Otani data. By contrast, the slow wave velocity, for those angles where it was observed, remains around 1500 m/s in both datasets. Discrepancies between the two sources of data for similar bone type and porosities may arise from the fact the Hosokawa samples had been saturated with water prior to testing, which is likely to affect measured velocities,[16] as well as differences in experimental method.

Figure 5(a) shows the predictions of the Stratified Biot and Schoenberg models calculated for the minimum and maximum porosity values of the samples tested, these being $\beta=0.65$ and 0.82. This yielded a pair of curves for each wave predicted, such that a tested sample with a porosity within this range would yield a velocity within these bounds. This approach is preferable to comparing all data to one prediction with a single porosity value. Clearly, the effect of varying input porosity between these bounds (roughly a 20% change) has a greater effect on the Stratified Biot model, which displays a change in fast wave velocity of roughly 500 m/s (also 20%), than on the Schoenberg model, which, at its greatest, displays a change of less than 100 m/s (6%).

The performance of the Stratified Biot model is varied. First, the Stratified Biot model gives poorer agreement with slow wave data than the Schoenberg model, both quantitatively and in the curvature of its angular variation. Some discrepancy may arise because the omission of shear in the development of the stratified tortuosity reduces the slow wave velocity in Eq. (1) to zero, and thus only the expression for the Biot fast wave is equated with Eq. (8). This may reduce the influence of the slow wave in this approach. Nevertheless, at 0°, where the slow wave velocity is reduced to 0 m/s, the models are in agreement.

In the case of the fast wave, two differing outcomes are seen. First, between 0° and 20°, Schoenberg's theory continues to provide better agreement with the Hughes *et al.* data, although not with that of Hosokawa and Otani. However, the Stratified Biot model improves agreement with data over a larger angular range from 30° to 90°. This effect is illustrated more clearly in Fig. 5(b). Figure 5(b) shows fast wave data selected between 30° and 90° only, where the Stratified Biot model is evaluated for the porosity of each sample tested. Agreement is considered to have been improved in two areas. First, the Stratified Biot model enhances quantitative agreement to data with porosity, an effect not predicted by Schoenberg's theory to any similar degree. Second, the shallower curvature of the Stratified Biot model improves agreement over this wide angular range. This is particularly seen

in the case of the data from the sample with a porosity of 0.82, where the prediction is, to within 4%, inside the experimental error.

Hence, it may be concluded that the introduction of a stratified tortuosity within Biot's theory enhances the correlation with the current fast wave data at high propagation angles, although Schoenberg's model remains superior for low angles and the slow wave.

As an aside, these results may be further judged in the context of clinical utility. Many QUS systems test the calcaneus in the mediolateral axis, where the dominant trabecular microstructure may be assumed to be perpendicular to the direction of insonation, i.e., close to 0°. It is at low angles, however, where the Stratified Biot model does not perform as well as Schoenberg's theory. Thus, a limited attempt was made to improve agreement to fast wave data at low angles.

VII. INTRODUCTION OF DIRECTIONAL INPUT PARAMETERS

While the Stratified Biot model introduces anisotropy through its description of the porous geometry, anisotropy also arises through mechanical properties of the cancellous matrix. Parameters such as the Young's modulus and Poisson's ratio of bone are dependent on direction;[31] indeed, highly anisotropic cancellous bone can display a ten fold difference in moduli depending on the loading direction with respect to trabecular orientation.[1] Much bone is also transversely isotropic, where properties are similar in a plane perpendicular to a longitudinal axis, and are significantly different from properties in the longitudinal direction.

Mechanical anisotropy was introduced into the Stratified Biot model in a simple empirical way. Gibson and Ashby[1] reported the Young's modulus of human cortical bone along the longitudinal axes as 18.1–22.6 GPa, and perpendicular to this, as 10.2–11.5 GPa. Taking rounded averages for each axes, with 90° representing the trabecular direction in testing, an angle-dependent Young's modulus for an anisotropic medium may be expressed as

$$E_s(\theta) = E_s(0^\circ)\sin^2(\theta) + E_s(90^\circ)\cos^2(\theta), \tag{11}$$

where $E_s(0^\circ)=10$ GPa and $E_s(90^\circ)=20$ GPa.

The results of introducing $E_s(\theta)$ into the Stratified Biot model can be seen in Fig. 6. Figure 6 shows predictions at the maximum and minimum porosities from both models, and data is as in previous figures.

The inclusion of $E_s(\theta)$ provides no improvement in agreement with slow wave data, possibly because solid properties such as moduli are more associated with fast wave propagation. One significant shortcoming with the modification is that quantitative agreement at higher angles is somewhat degraded for some data. However, the fast wave predictions of the stratified Biot model and Schoenberg's theory become closer in shape, with the former being reduced by roughly 500 m/s at 0° and slightly flattening over 10°–30°. Furthermore, it should be noted that for individual samples, the consistent error in the Stratified Biot $E_s(\theta)$ model of roughly 9% is comparable with the worst error in Schoenberg's model (of ~10% against porosity-matched curves).

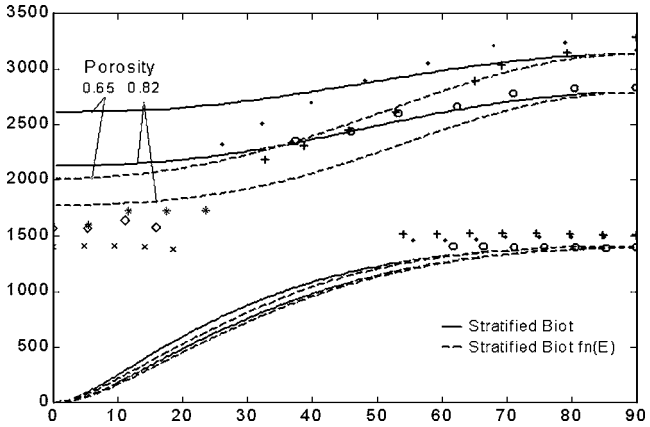


FIG. 6. Predicted and measured fast wave phase velocities versus propagation angle for the Stratified Biot model (solid line) and a stratified Biot model including directional Young's modulus (dashed line), evaluated at the upper and lower porosity limits. The angle 0° is normal to the structure [data from Hughes *et al.* (Ref. 6)].

Hence, although this approach would benefit from further study, the inclusion of angle-dependent mechanical properties such as $E_s(\theta)$, as well as structural ones such as $\alpha(\theta)$, appears to have some value in anisotropic propagation models.

VIII. CONCLUSION

In this paper we have shown that the introduction of an angle-dependent tortuosity for a layered medium in a Biot model leads to the prediction of direction-dependent wave properties. Thus, the anisotropic response of ultrasound in cancellous bone may be partly attributed to the angular variation in inertial coupling.

While work on an anisotropic propagation model for cancellous bone requires further development, it has nonetheless been a key aim of this paper to highlight the fundamental importance of the tortuosity. Expressions for tortuosity may be easily manipulated and adapted to describe the varying architecture of cancellous bone. Where the use of tortuosity may be most valuable is in describing the deterioration of the cancellous structure, owing to the onset of osteoporosis. Changes such as the erosion of calcified plates and rods, and loss of anisotropy, cannot be completed described by the porosity or factors such as pore size, plate separation, or trabecular thickness. What remains to be investigated is whether an analysis of an adaptable tortuosity may give valuable information in the assessment of bone, whether *in vitro* or with clinical utility.

ACKNOWLEDGMENTS

The authors would like to acknowledge the contribution of Professor Robert C Chivers, Visiting Professor at ISVR (1998–2001), who died in November 2004. Bob was a key advisor during the research, providing invaluable guidance drawn from his vast knowledge of physical acoustics. His groundbreaking research paved the way for our current understanding of ultrasonic propagation in biological tissue.

The authors also appreciate the collaboration of Dr. Kang II Lee, of ISVR and the Department of Physics and Institute of Basic Science, Sungkyunkwan University, Republic of Korea.

[1]L. J. Gibson and M. Ashby, "Cancellous bone," in *Cellular Solids—Structure and Properties* (Pergamon, Oxford, 1998), Chap. 11, pp. 316–331.
[2]C. F. Njeh, D. Hans, T. Fuerst, C.-C. Glüer, and H. K. Genant, "Basic sciences," in *Quantitative Ultrasound: Assessment of Osteoporosis and Bone Status* (Martin Dunitz Ltd, London, 1999), Chap. 1, pp. 1–101.
[3]M. A. Biot, "Theory of propagation of elastic waves in a fluid saturated porous solid, II. High frequency range," J. Acoust. Soc. Am. **28**, 179–191 (1956).
[4]M. L. McKelvie and S. B. Palmer, "The interaction of ultrasound with cancellous bone," Phys. Med. Biol. **36**, 1331–40 (1991).
[5]J. L. Williams, "Ultrasonic wave propagation in cancellous and cortical bone: predictions of some experimental results by Biot's Theory," J. Acoust. Soc. Am. **92**, 1106–1112 (1992).
[6]W. Lauriks, J. Thoen, I. Van Asbroek, G. Lowet, and G. Van der Perre, "Propagation of ultrasonic pulses through trabecular bone," J. Phys. IV **4**, 1255–1258 (1994).
[7]J. L. Williams, M. J. Grimm, F. W. Wehrli, K. R. Foster, and H.-W. Chung, "Prediction of frequency and pore-size dependent attenuation of ultrasound in trabecular bone using Biot's Theory," in *Mechanics of Poroelastic Media* (Kluwer Academic, Netherlands, 1996), pp. 263–71.
[8]A. Hosokawa and T. Otani, "Ultrasonic wave propagation in bovine cancellous bone," J. Acoust. Soc. Am. **101**, 558–562 (1997).
[9]E. R. Hughes, T. G. Leighton, G. W. Petley, and P. R. White, "Ultrasonic propagation in cancellous bone: A new stratified model," Ultrasound Med. Biol. **25**, 811–821 (1999).
[10]K. I. Lee, H.-S. Roh, and S. W. Yoon, "Acoustic wave propagation in bovine cancellous bone: Application of the Modified Biot-Attenborough model," J. Acoust. Soc. Am. **114**, 2284–2293 (2003).
[11]Z. E. A. Fellah, J. Y. Chapelon, S. Berger, W. Lauriks, and C. Depollier, "Ultrasonic wave propagation in human cancellous bone: Application of Biot theory," J. Acoust. Soc. Am. **116**, 61–73 (2004).
[12]K. A. Wear, A. Laib, A. P. Stuber, and J. C. Reynolds, "Comparison of measurements of phase velocity in human calcaneus to Biot theory," J. Acoust. Soc. Am. **117**, 3319–3324 (2005).
[13]C.-C. Gluer, C. Y. Wu, and H. K. Genant, "Broadband ultrasound attenuation signals depend on trabecular orientation," Osteoporosis Int. **3**, 185–191 (1993).
[14]P. H. F. Nicholson, M. J. Haddaway, and M. W. J. Davie, "The dependence of ultrasonic properties on orientation in human vertebrae," Phys. Med. Biol. **39**, 1013–24 (1994).
[15]A. Hosokawa and T. Otani, "Acoustic anisotropy in bovine cancellous bone," J. Acoust. Soc. Am. **103**, 2718–2722 (1998).
[16]P. H. F. Nicholson and M. L. Bouxsein, "Bone marrow influences quantitative ultrasound measurements in human cancellous bone," Ultrasound Med. Biol. **28**, 369–375 (2002).
[17]E. R. Hughes, T. G. Leighton G. W. Petley, P. R. White, and R. C. Chivers, "Estimation of critical and viscous frequencies for Biot theory in cancellous bone," Ultrasonics **41**, 365–368 (2003).
[18]J. G. Berryman, "Confirmation of Biot's theory," Appl. Phys. Lett. **37**, 382–384 (1980).
[19]C. Zwikker and C. W. Kosten, *Sound Absorbing Materials* (Elsevier, New York 1949), Chap. 8, p. 21.
[20]D. L. Johnson, J. Koplik, and R. Dashen, "Theory of dynamic permeability and tortuosity in fluid-saturated porous media," Fluid Dyn. **176**, 379–402 (1987).
[21]Z. E. A. Fellah, S. Berger, W. Lauriks, C. Depollier, C. Aristegui, and J.-Y. Chapelon, "Measuring the porosity and the tortuosity of porous materials via reflected waves at oblique incidence," J. Acoust. Soc. Am. **113**, 2424–2433 (2003).
[22]K. Attenborough, H.-C. Shin, Q. Qin, M. Fagan, and C. M. Langton, "Measurement of tortuosity in stereolithographical bone replicas using audiofrequency pulses," J. Acoust. Soc. Am. **118**, 2779–2782 (2005).
[23]R. Strelitzki, V. Paech, and P. H. F. Nicholson, "Measurement of airborne ultrasonic slow waves in calcaneal cancellous bone," Med. Eng. Phys. **21**, 215–233 (1999).
[24]M. Schoenberg, "Wave propagation in alternating solid and fluid layers," Wave Motion **6**, 303–320 (1984).
[25]F. Padilla and P. Laugier, "Phase and group velocities of fast and slow compressional waves in trabecular bone," J. Acoust. Soc. Am. **108**, 1949–1952 (2000).
[26]K. A. Wear, "A stratified model to predict dispersion in trabecular bone," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **48**, 1079–1083 (2001).
[27]M. Kaczmarek, J. Kubik, and M. Pakula, "Short ultrasonic waves in cancellous bone," Ultrasonics **40**, 95–100 (2002).
[28]M. A. Biot, "Generalized theory of acoustic propagation in porous dissipative media," J. Acoust. Soc. Am. **34**, 1254–1264 (1962).
[29]M. D. Sharma and M. L. Gogna, "Wave propagation in anisotropic liquid-saturated porous solids," J. Acoust. Soc. Am. **90**, 1068–1073 (1991).
[30]M. Schoenberg and P. N. Sen, "Properties of a periodically stratified acoustic half-space and its relation to a Biot fluid," J. Acoust. Soc. Am. **73**, 61–67 (1983).
[31]F. A. Duck, "Elastic moduli of bone and teeth," in *Physical Properties of Tissue: A Comprehensive Reference Book* (Cambridge, University Press, Cambridge, 1990), Chap. 5, pp. 139—146.
[32]J. R. Wu and F. Cubberly, "Measurement of velocity and attenuation of shear waves in bovine compact bone using ultrasonic spectroscopy," Ultrasound Med. Biol. **23**, 129–34 (1997).

Acoustics and behavioral contexts of "gecker" vocalizations in young rhesus macaques (*Macaca mulatta*)

Erik R. Patel[a] and Michael J. Owren[b]
Department of Psychology, 211 Uris Hall, Cornell University, Ithaca, New York, 14853

(Received 21 February 2006; revised 28 August 2006; accepted 13 October 2006)

Loud, pulsed "gecker" vocalizations are commonly produced by young rhesus macaques in distressful circumstances. The acoustics, usage, and responses associated with these calls were examined using audio recordings and observational data from captive, socially living rhesus up to 24 months old. One-hundred-eleven gecker bouts were recorded from ten individuals (six males, four females), with most geckers produced during the first 6 months of age. A gecker call consisted of a bout of up to 28 pulses of spectrally structured noise with a single prominent frequency peak. Nine contexts of calling were identified, but little evidence of context-specific acoustic variation was found. While geckering often triggered responses by the vocalizer's mother, the most common outcome was the absence of any reaction. Females geckered longer and at higher rates than did males, while also showing acoustic evidence of greater vocal effort. Mothers nonetheless responded more often and more positively to males. Overall, results show that gecker acoustics vary somewhat with vocalizer sex, age, and likely arousal level, but do not reflect detailed aspects of behavioral context. Circumstances of production suggest that geckers function primarily to draw the attention of mothers, who in turn are selective in responding.
[DOI: 10.1121/1.2390662]

PACS number(s): 43.80.Ka [JAS] Pages: 575–585

I. INTRODUCTION

Young nonhuman primates produce a variety of vocalizations when in distress, which in rhesus macaques (*Macaca mulatta*) include "coos," "squeaks," "screams," "pant threats," "girneys," and "geckers" (Maestripieri and Call, 1996; Maestripieri *et al.*, 2000; Tomaszycki *et al.*, 2001). Geckers in particular have been described as a distress-related vocalization in rhesus monkeys, occurring in large numbers when infants are ignored, rejected, or otherwise separated from their mothers. Often accompanied by dramatic, convulsive bodily jerking, these loud, explosive, and distinctive-sounding vocalizations have received some attention and mention from a variety of researchers, but predominantly through qualitative rather than quantitative description. Thus, while geckers have also been documented in a number of other macaque species, details of their form and function remain unclear (Newman, 1995; Green, 1981).

Altmann (1962) first named geckers onomatopoetically as "ik, ik, ik." He noted that the call is accompanied by spasmodic jerking and occurs during weaning and other circumstances of evident infant distress. The sounds were subsequently referred to as "geckers" and "geckering screeches" by Rowell (1962; Rowell and Hinde, 1962), who described them as a series of 0.1 s squeaks separated by 0.5–1.0 s intervals (illustrated in Fig. 1). Several subsequent studies have shown geckers to be among the most common vocalizations produced by young rhesus. For example, Owren *et al.* (1993) reported that geckers comprised approximately 38% of the total number of vocalizations produced by rhesus monkeys during the first year of life, and 12% in the second year. Figures from Maestripieri *et al.* (2000) were comparable, with geckers making up 43.3% of all vocalizations in the first three months of life. Tomaszycki *et al.* (2001) similarly noted that geckers were the most common vocalization over the first 8 months, accounting for 34% of vocal output.

In spite of the prevalence of these sounds, little quantitative information is available concerning their acoustics or function. On the one hand, Newman (1995) notes that geckers can routinely occur without apparent cause, and suggests that they "may largely be a reflection of the immature state of the central nervous systems underlying vocal expression" (p. 79). On the other hand, many researchers are likely to agree with Maestripieri and Call's (1996) characterization that geckers, like most primate infant distress calls, occur "when the infant is not in contact with its mother" and "signal the infant's need for nursing, transport, or protection" (Maestripieri and Call, 1996, p. 620). Critical issues thus include whether rhesus geckers are in fact occurring primarily or exclusively in instances of distress, whether geckers have specific, context-dependent signaling value, and how other animals respond to these sounds.

Relatively few studies of primate infant vocalization have examined the co-occurrence of call acoustics and context (e.g., Kalin *et al.*, 1992; Jovanovic and Gouzoules, 2001), or found evidence of context-specific vocal subtypes (e.g., Bayart *et al.*, 1990). Although also limited, some information is available concerning responses to geckers and other distress-related calls. Across primate species, for instance, it has generally been found that mothers are most responsive to young infants that vocalize at relatively low

[a]Author to whom correspondence should be addressed. Electronic mail: erp8@cornell.edu
[b]Michael J. Owren is now at the Department of Psychology, Georgia State University, PO Box 5010, Atlanta, Georgia 30302-5010.

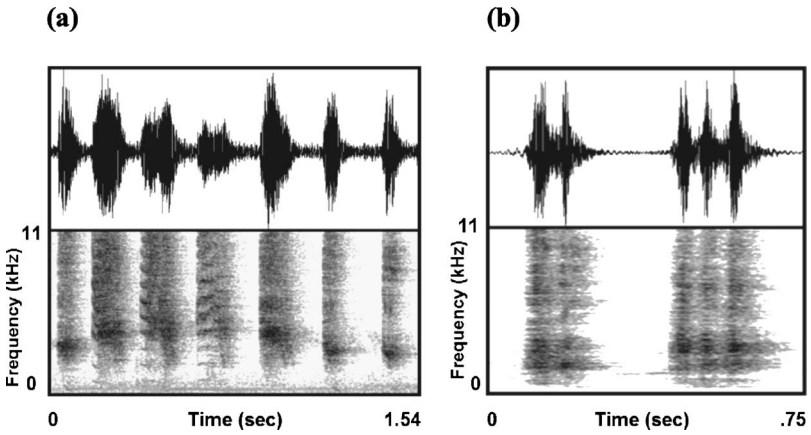


FIG. 1. (a) A representative gecker bout consisting of seven pulses, shown as (top) a wave form and (bottom) a narrowband FFT spectrogram (20 ms Hanning analysis window, 22.05 kHz sampling rate). The third and fourth pulse shown produced higher HNR values than elsewhere, but are nonetheless still likely to reflect deterministic chaos in underlying vocal-fold vibration. (b) A gecker bout illustrating pulse reduplication in the wave form and a narrowband FFT spectrogram. Pulses are first doubled and then tripled.

rates (Maestripieri, 1995; Hauser, 1993). Furthermore, females in several species have been reported to produce distress-related calls comprised of more call types, emitted at higher rates, and for longer durations than males (Locke and Hauser, 1999; Wallen, 2005; Green, 1981). Overall, rhesus females do show greater volubility than males, not only in infancy (Newman *et al.*, 1990; Tomaszycki *et al.*, 2001), but also in adolescence and adulthood (Erwin, 1975). Thus, while sex differences in maternal treatment of primate infants tends to be rare (Wallen, 2005), Tomaszycki *et al.* (2001) have nonetheless reported that rhesus mothers are more likely to respond to vocalizations by male than by female infants. Despite this evidence, rather little is as yet known about the ontogeny or extent of sex differences in vocal behavior (Gautier and Gautier 1977; Green, 1981). Furthermore, the few studies that are available have typically examined time periods of less than a year (Hammerschmidt *et al.*, 2000; Hammerschmidt *et al.*, 2002; Seyfarth and Cheney, 1986; Newman, 1995; Snowdon *et al.*, 1997; Owren *et al.*, 1993).

The current study sought to extend previous work on rhesus geckers by examining the acoustics and contexts of these vocalizations as systematically as possible in animals up to 24 months of age. The work began with a detailed characterization of gecker acoustics and the contexts in which infants produce these calls. The data were then used to test for possible specificity in acoustics and contexts, age and sex-related differences, and factors that might affect the likelihood of responding by mothers or other animals. These questions were examined using a large database of calls recorded during a long-term study of two groups of socially housed rhesus monkeys living in outdoor cages at a national primate center.

II. METHODS

A. Study site and subjects

Data were collected during the course of a long-term project on vocal development (Owren *et al.* 1992a, 1992b; Owren *et al.*, 1993). The subset analyzed here drew on audio recordings and behavioral data collected between 1986 and 1989 from ten rhesus macaques (six males and four females) ranging in age from newborn to 24 months of age (see Table I). The overall project involved cross fostering a small number of neonatal infants between different-species foster mothers (Owren and Dieter, 1989), but no such offspring are represented in the data analyzed here. The ten subjects included here represented all the rhesus offspring born into the study groups during the course of the project and from whom there were any usable gecker recordings. The subjects were offspring of a total of seven different mothers, all of whom were multiparous. Five mothers were represented by a single study subject, one by two study subjects, and one by three study subjects. All animals were located at the California National Primate Research Center (CNPRC; University of California, Davis), as part of two socially living groups. Group composition varied somewhat over time, with each of the two cages including 1–2 adult males, one or no unrelated juvenile males, 4–6 adult females, and 6–8 offspring of those females.

Each group was housed outdoors in a separate cage constructed from two modified commercial corn cribs connected by a rectangular inter-cage unit (Hoffman and Stowell, 1973). The cages were approximately $13.6 \times 4.3 \times 3.0$ m (with additional space created by conical roofs), and contained capture chutes, perches, and pea gravel used as replaceable ground cover. During the winter months, screens were installed as wind shields and infrared lamps provided heat. The monkeys had continuous access to fresh water and received Purina monkey chow twice daily in quantities that ensured a plentiful supply for all. Detailed physical examinations of the animals were conducted three times per year by the CNPRC veterinary staff and general health was checked daily.

TABLE I. Subject demographics and representation in the gecker sample.

Subject	Sex	Age (mo)	No. of Bouts	No. of Pulses
JN	M	1–5	3	9
JS	M	2–5	16	76
LN	F	4–14	27	281
ML	F	2–4	7	47
MS	M	1–16	28	98
MY	F	3–17	6	55
NK	M	4–7	9	33
RK	M	3–17	4	15
SM	M	5–14	8	18
SV	F	5	3	19

TABLE II. The 24 variables used in acoustic analysis.

Acoustic measure	Definition
Bout-Dur	Duration of a gecker bout (s)
Pulse-Dur	Duration of a single gecker pulse within a bout (s)
Pulses-Bout	Number of gecker pulses within a bout
Pulse-Redup	Number of re-duplicated pulses within a bout
SNR (signal-to-noise ratio)	Difference between rms amplitude of a pulse and adjacent background noise (dB)
Peak1	Frequency of the first LPC peak (Hz)
Spectral-Tilt	Overall slope of the LPC function
LPC1 to LPC12	Values of the 12 coefficients of the LPC function
Spectral-Mean	Mean of the normalized frequency spectrum (Hz)
Spectral-StDev	Standard deviation of the normalized spectrum
Spectral-Skew	Relative symmetry of the normalized spectrum
Spectral-Kurtosis	Relative peakedness of the normalized spectrum
HNR (harmonic-to-noise ratio)	Ratio of wave form periodicity to noisiness (dB)

B. Apparatus and procedure

1. Observation and recording

Behavioral observations were conducted in 10 min continuous focal sessions (Altmann, 1974) occurring in the morning (7:30 AM to noon) in the spring, summer, and fall, and in the afternoon (noon to 5 PM) in winter, up to six days per week. The daily sampling order was determined quasi-randomly. During sampling, behaviors exhibited by the focal animal and others it interacted with were entered as codes on a handheld computer, with each entry noting animal identities, any one or more of 63 concomitant behaviors, as well as the onset time of occurrence. This approach allowed documentation of all vocalizations produced by a focal animal during the session, as well as the behavior associated with these calls.

Audio recording was conducted with a four-channel Tascam 234 cassette deck (Teac Corporation, Tokyo), and two Sennheiser ME88 microphones (Sennheiser Electronic, Old Lyme, CT) placed on stands in fixed positions outside the cage. Calls from all animals were routinely recorded, including ad libitum samples that were described through supplementary commentary using a lapel microphone connected to a separate channel of the cassette deck. Vocalizations were later transferred to $\frac{1}{4}$ in. audio tape using either a Uher 4200 Report Monitor (Uher Werke Munchen, Munich) or a Fostex A-2 reel-to-reel deck (Fostex America, Norwalk, CA). Calls were separated, numbered, and annotated during this process using the coded behaviors and any additional, audio-recorded comments.

2. Call selection and analysis

The vocalization sample was identified by examining the entire database of more than 10,000 separately cataloged calling episodes available from the project. Gecker vocalizations were identified based on previous annotations in the database (Owren *et al.*, 1992a), by matching acoustic features to published descriptions of gecker calls (e.g., Rowell and Hinde 1962; Jovanovic and Gouzoules 2001), and by ear. A total of 111 analyzable gecker bouts comprising 651 individual gecker pulses from the ten vocalizers were identified in this manner (see Table I). A bout was defined an as uninterrupted emission of pulsed vocalization by a single animal with less than a 5 s pause between pulses.

Vocalizations were digitized and archived using a sampling rate of 44.1 kHz with 16 bit accuracy. Calls were subsequently downsampled to 22.05 kHz for analysis, which was conducted using the ESPS/waves+ 5.3 "xwaves" package (Entropic Research, Washington, DC), running on an SGI O2 workstation (Silicon Graphics Incorporated, Mountain View, CA). The Praat speech analysis package was also used in some analyses (Boersma, 2001; available at www.praat.org). Statistical analyses were conducted using NCSS 2004 (Jerry Hintze, Kaysville, UT) and SPSS 13.0 (SPSS Inc., Chicago, IL).

3. Acoustic variables

Twenty-four acoustic measures were extracted (definitions are provided in Table II), spanning temporal, amplitude, and frequency-spectrum aspects of the calls. Temporal measures of geckering at bout and pulse levels included absolute durations (Bout-Dur, Pulse-Dur), and number of pulses per bout (Pulses-Bout). As illustrated in Fig. 1(b), gecker pulses sometimes occur in doubled, tripled, quadrupled, and even quintupled form without intervening silent intervals. This phenomenon was referred to as "twinning" by Maestripieri *et al.* (2000), and here is considered as a single phenomenon labeled "reduplication" (Pulse-Redup). Call amplitude was characterized based on calculating a signal-to-noise ratio (SNR) for individual bouts, defined as the root-mean-square (rms) amplitude of the pulse minus the rms amplitude of a representative, adjacent segment of background noise. No attempt was made to measure the absolute amplitude of calls. This indirect approach was used as an attempt to obtain relative amplitude information for comparisons such as possible age-, sex-, and context-related differences. It was arguably justified by the fact that microphones were placed in the same locations outside the cages for every session, and based on the assumption that general background noise levels remained roughly constant over time. Sources of error for SNR

TABLE III. Codes used to characterize behavior occurring before/during and after a geckering event.

Code	Type	Definition
Affil	Before/during	Infant engages in "friendly" prosocial behavior with mother or others, such as touching, grooming, and play.
Aggr	Before/during	Infant engages in or receives aggressive behavior with mother or others, such as threatening, pushing, hitting, biting, and nipple rejections.
Inf-Follow	Before/during	Infant follows its locomoting mother.
Mth-Leave	Before/during	Mother moves out of proximity to her infant.
Inf-Leave	Before/during	Infant moves out of proximity to her mother.
Spon	Before/during	Absence of any evident antecedent or co-occurrence.
Pos-Outcome	After	Infant experiences Affil and/or attention from mother or others.
Neg-Outcome	After	Infant experiences Aggr from its mother or others.
No-Outcome	After	Infant experiences neither Aggr nor Affil from mother or others.
Rsp	Response	Any immediate discernible maternal action occurring in response to her infant's gecker (includes looking, touching, shifting position, threatening, pushing, hitting, or biting an infant that is in contact, and vocalizing, looking, approaching, reaching, touching, threatening, pushing, hitting, biting, as well as Affil or Aggr toward an infant that is not in contact).
No-Rsp	Response	Absence of discernible maternal behavior in response to her infant's gecker.

measurements would then be any moment-to-moment variation in background noise, and the vocalizer's distance from and relative orientation to the microphone.

Frequency-spectrum (spectral) measurements were made from all 651 total pulses in the sample, based on a spectral slice located midway through each pulse. This slice was characterized by overlaying a ten-coefficient, autocorrelation-based linear predictive coding (LPC) function (0.023 s Hanning window) on a 512-point fast Fourier transform (FFT) spectrum (Hanning window) of the same segment (further described by Owren and Bernacki, 1998), with frequency and amplitude values recovered from the LPC envelope through cursor-based measurement. Variables used were the first LPC peak (Peak1), the overall slope of the LPC spectrum (Spectral-Tilt), values of each of the 12 coefficients used in the LPC function (LPC1 to LPC12), and the first four spectral moments of the Fourier spectrum (Spectral-Mean, Spectral-StDev, Spectral-Skew, Spectral-Kurtosis). Spectral-moment measures are particularly well suited to characterizing global features of noisy, broadband frequency spectra (Forrest *et al.*, 1988), and are computed by normalizing the Fourier power spectrum and treating it as a statistical distribution. The first four moments of this distribution are its mean, standard deviation, skewness, and kurtosis, in the normal sense of these terms. The last measure was the harmonic-to-noise ratio (HNR) of each pulse, extracted using Praat's pitch-based algorithm (Boersma, 1993).

4. Behavioral variables

The behavioral contexts of each of the 111 gecker bouts were coded in binary fashion according to whether the offspring's mother exhibited any discernible response to the vocalizations (Rsp/No-Rsp), as well as being scored as to which of six possible behaviors were occurring just before or during the bout (Before/During) and whether vocalizing was followed by aggression, affiliation, or neither (After). Behavioral codes are defined in Table III. Coding for Before/During and After variables was based on behavioral codes recorded just before, during, and after each gecker bout. Behaviors included as the context of each vocalization event were determined by the observer, who also transcribed the relevant codes for each episode on a daily basis during the course of the project. Contexts were thus not inherently limited in time either before or after the vocalization, but rarely exceeded a few seconds on either side. In four cases, data were insufficient to determine the Before/During context and those gecker bouts were excluded from analyses involving this variable. In 34 cases, mothers were out of view, and Rsp/No-Rsp could not be coded.

III. RESULTS

A. Gecker acoustics

1. General features

Descriptive statistics for the acoustic features of gecker vocalizations are shown in Table IV. On average, geckers were emitted in 1 s bouts of four pulses, though both bout duration (Bout-Dur) and number of pulses per bout (Pulses-Bout) were highly variable. The longest bout observed was nearly 9 s in duration, and as many as 28 pulses were observed in a single bout. Reduplication (Pulse-Redup) occurred in 34% of all bouts, could include as many as five pulses, but usually consisted of just two (72% of total reduplication events). Pearson's correlation coefficients revealed that Pulse-Redup was strongly correlated with Bout-Dur ($r=0.83$, $p<0.001$) and Pulses-Bout ($r=0.78$, $p<0.001$).

Although geckers are likely among the loudest vocalizations emitted by young rhesus macaques, mean SNR values were relatively modest (just over 6 dB), while showing high overall variability with the loudest geckers at 19.1 dB. The broadband noisiness of the sounds was reflected in generally low HNR values and spectral slopes that were virtually flat.

TABLE IV. Descriptive features of gecker vocalizations, based on 111 total bouts and 651 total pulses.

Variable	M	SE	Min	Max	Range	COV_{mean}
Age (mo)	5.32	0.41	1	17	16.0	0.80
Bout-Dur (s)	0.98	0.12	0.03	8.96	8.9	1.26
Pulse-Dur (s)	0.18	0.01	0.03	0.71	0.68	0.52
Pulses-Bout	4.29	0.37	1.00	28	27.0	0.92
Pulse-Redup	1.06	0.28	0.00	26	26.0	2.79
SNR (dB)	6.40	0.44	−1.35	19.1	20.5	0.72
Peak 1 (Hz)	2923	74.9	1208	4488	3280	0.27
Spectral-Tilt	−0.01	0.01	−0.22	0.33	0.55	-12.6
Spectral-Mean (Hz)	2087	127.2	114.2	5932	5818	0.64
Spectral-StDev	1480	66.1	136.5	3661	3524	0.47
Spectral-Skew	2.70	0.40	−0.92	22.7	23.6	1.55
Spectral-Kurtosis	41.5	14.5	−1.53	1066	1068	3.67
HNR (dB)	4.13	0.24	−0.67	10.4	11.0	0.60
HNR-SD	3.75	0.11	1.25	7.24	5.99	0.31
LPC1	0.63	0.07	0.03	6.1	6.10	1.18
LPC2	0.66	0.07	0.05	6.1	6.10	1.14
LPC3	0.30	0.03	−0.64	0.92	1.60	1.09
LPC4	−0.35	0.03	−0.84	0.43	1.27	−0.86
LPC5	−0.17	0.03	−0.66	0.54	1.20	−1.69
LPC6	−0.38	0.02	−0.75	0.33	1.08	−0.61
LPC7	−0.10	0.02	−0.57	0.53	1.10	−2.24
LPC8	−0.20	0.02	−0.71	0.31	1.01	−1.02
LPC9	−0.01	0.02	−0.47	0.59	1.06	−15.47
LPC10	−0.09	0.02	−0.60	0.49	1.09	−1.93
LPC11	0.02	0.01	−0.42	0.42	0.84	6.16
LPC12	−0.05	0.01	−0.40	0.27	0.67	−2.57

HNR values did become significantly higher in a few cases (see, for example, the third and fourth pulses shown in Fig. 1(a)), but were never clearly or purely harmonically structured. LPC analysis revealed a one consistent frequency peak just below 3000 Hz (Peak1), with the spectral-moment mean falling just above 2000 Hz. LPC analysis revealed a second peak occurring above 7000 Hz in some cases, but associated frequency values were highly variable. This peak was not discernible in the majority of bouts, and therefore was also not included as a reliable component of gecker acoustic structure. When it did occur, the peak was not harmonically related to Peak1.

2. Age and sex

The effects of age and sex on geckering are shown in Fig. 2, based on 2-mo age blocks during which gecker bouts were available from at least three different individuals (i.e., months 1–2, 3–4, 5–6, 7–8, and 13–14, respectively). Repeated-measures General Linear Model analysis of variance (ANOVAs) were used to test for possible age effects on acoustics, but independent of individual variation. Results revealed that younger infants emitted the fewest pulses per bout, $F(4,96)=6.75$, $p<0.05$, the shortest pulses, $F(4,96)=8.39$, $p<0.05$, and the shortest bouts, $F(4,96)=7.55$, $p<0.01$. Results from associated Tukey-Kramer multiple-comparison posthoc tests are displayed in Fig. 2.

A number of sex differences were also found, even when controlling as well as possible both for body-size differences between males and females, and for differences among individuals. As could be expected, analysis of covariance (ANCOVA) with vocalizer age entered as a covariate showed males ($M=1.62$ kg) to be significantly heavier than females ($M=1.41$ kg), $F(1,110)=14.2$, $p<0.001$. Infant body weight was therefore entered as a covariate in further repeated-measures ANCOVAs. With vocalizer sex as the independent variable, these tests revealed sex differences in four temporal, amplitude, and frequency-spectrum aspects of the calls (see Table V). The strongest effect sizes were found for Peak1, with frequency peaks in female being almost 700 Hz higher than in male calls, and for Bout-Dur, where female bouts were more than twice as long as male bouts. Other outcomes showed male geckers to be higher in amplitude (SNR), and also less noisy (HNR).

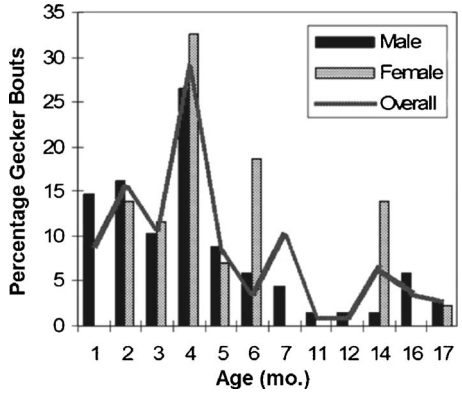


FIG. 2. Mean percentage of gecker bouts occurring by sex and age, based on 2 month age blocks, tallied separately for each individual caller.

TABLE V. Statistically significant ANCOVA tests for sex differences in gecker acoustics with body weight entered as a covariate.

Acoustic measure	Males (M/SE)	Females (M/SE)	Direction of difference	F(1,5)	Effect Size (d)	p value
Bout-Dur (s)	0.57/0.07	1.60/0.25	F>M	19.3	0.86	.007
SNR (dB)	7.53/0.56	4.59/0.63	M>F	8.74	0.67	.032
HNR (dB)	4.58/0.29	3.40/0.38	M>F	8.73	0.48	.032
Peak1 (Hz)	2654/91.3	3349/99.3	F>M	17.7	0.99	.008

B. Gecker usage

1. Age and sex

During the first two years of life, gecker bouts were emitted by infants from 1 to 17 months of age. Over this period, 75.7% of all gecker bouts occurred during the first 6 months of life (see Fig. 3). Gecker usage peaked at 4 months in both males, 26.5% ($X^2_{(6)}=51.7$, $p<0.001$), and females, 32.6% ($X^2_{(6)}=16.7$, $p<0.05$). However, the percentage of gecker bouts used in each month differed between the sexes ($X^2_{(11)}=25.1$, $p<0.01$), with male geckers tending to occur at younger ages. The greatest proportions of male gecker bouts occurred during months four (26.5%), two (16.2%), and one (14.7%), respectively. Females produced the most gecker bouts during months four (32.6%), two (14.0%), and 14 (14.0%), respectively.

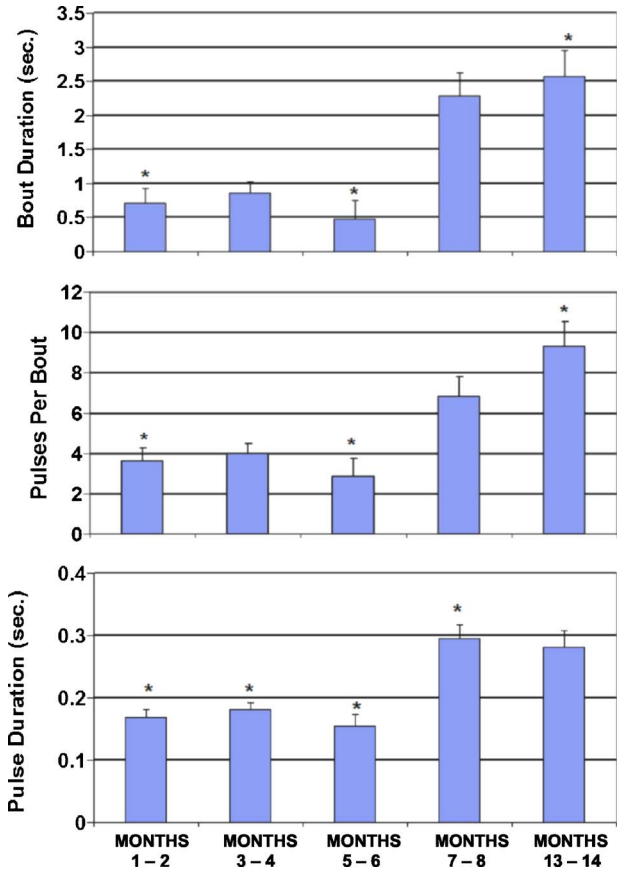


FIG. 3. (Color online) Bout duration (Bout-Dur), the number of pulses per bout (Pulse-Bout), and pulse duration (Pulse-Dur) are shown as a function of offspring age in 2 month blocks. Asterisks signify Tukey-Kramer posthoc tests in which at least one comparison to other outcomes shown in the panel was statistically significant.

2. Contexts

The behavior occurring immediately before or during a gecker bout was clear in all but four of the 111 total cases. Although none of the six categories of Before/During behavior could be considered predominant, relative proportions did show statistically significant heterogeneity ($X^2_{(5)}=26.7$, $p<0.001$). Inf-Follow (25.2%) was the most prevalent, followed by Aggr (22.3%) and Spon (22.3%), Mth-Leave (18.7%), and Affil (9.4%). Inf-Leave (1.9%) could also occur before or during a gecker bout, but was rare. Males and females were found to differ in the proportion of geckers emitted across these categories ($X^2_{(5)}=23.9$, $p<0.001$). Male geckers were most common in Affil (male: 15.4%, female: 0%) and Aggr (male: 27.7%, female: 14.3%) contexts, while female geckers were most common in the Inf-Follow context (male: 10.8%, female: 47.6%). Differences in the Inf-Leave, Mth-Leave, and Spon categories were all less than five percentage points. The influence of infant age on sex differences in Before/During gecker usage was examined in infants up to 6 months old, and results paralleled the overall pattern of sex differences ($X^2_{(5)}=22.0$, $p<0.001$). Males geckered most in the Affil (male: 17.0%, female: 0%) and Aggr (male: 28.3%, female: 14.3%) contexts, while females geckered most in the Inf-Follow (male: 5.44%, female: 10.3%) context. However, the male-female difference in the Inf-Follow context was less pronounced in these younger infants.

C. Acoustics and contexts

A cross-validated, multinomial discriminant-function analysis was conducted with Before/During as an independent variable to determine whether the acoustic measures considered collectively could be used to discriminate among the six contexts associated with geckering. Only 34.8% of the 107 available cases were successfully classified, which was nonetheless statistically significant, $F(52,343)=2.10$, $p<0.001$. Canonical discriminant analysis (Tabachnik and Fidell, 2001) revealed that the first canonical correlation ($r_c=0.59$) was significant, $F(52,343)=2.06$, $p<0.0001$, with Pulse-Redup (1.02), SNR (0.91), and Bout-Dur (−0.67) being the most influential in discriminating among the contexts.

A cross-validated, multinomial discriminant-function analysis was conducted with After as an independent variable to determine whether the acoustic measures considered collectively could be used to discriminate among the three outcome categories associated with geckers. Only 26.6% of the 111 cases were correctly classified, which was not a statistically significant proportion, $F(26,184)=0.94$, $p>0.55$.

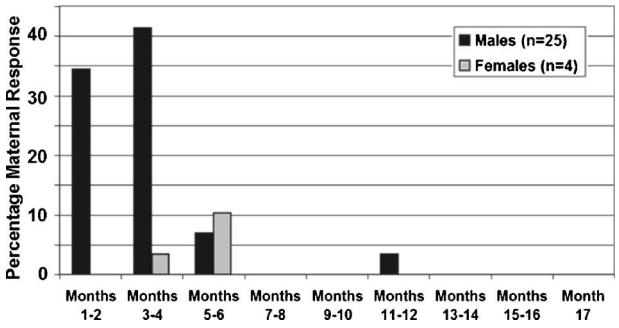


FIG. 4. Percentage occurrence of immediate maternal response by infant age block (mo.) to each gecker bout that received an immediate maternal response. Percentages shown sum to 100%, representing all instances that the mother showed an immediate response to her offspring's geckering (scored as Rsp, as described in TABLE III).

One-way ANOVA did reveal that the means of some acoustic measures differed as a function of the After context, including HNR, $F(2,104)=3.95$, $p<0.05$, Bout-Dur, $F(2,104)=3.91$, $p<0.05$, and Pulse-Dur, $F(2,104)=4.0$, $p<0.05$. Tukey posthoc tests demonstrated that the geckers with the longest Bout-Dur and Pulse-Dur values were associated with negative outcomes significantly more often than either positive outcomes or no outcomes. Geckers with higher HNR values were more often associated with positive outcomes than no outcome.

D. Responses to geckers

Immediate maternal response coding (Rsp/No-Rsp) was available for 76 geckering episodes, with mothers found to respond in only 38.2% of these cases, and doing so predominantly when the vocalizers were relatively young ($M=3.52$ mo). Although offspring as old as 17 months were still producing geckers, no responses were noted to individuals older than 12 months (see Fig. 4). A clear sex difference also emerged, with 47.2% of the gecker bouts emitted by males eliciting an immediate maternal response (25/53), compared to only 17.4% from females (4/23), $X^2_{(1)}=6.03$, $p<0.014$. Male geckers were also responded to ($M=3.32$ mo) earlier in life than females (M=4.75 mo), Mann-Whitney U, $z=2.44$, $p<.02$.

Gecker usage did not deviate significantly from expected proportions across the three outcome categories scored for geckers ($X^2_{(2)}=2.65$, $p>0.25$). Furthermore, no sex difference was found in the outcomes experienced when examining offspring across the entire 24 months, $X^2_{(2)}=0.72$, $p>0.68$. However, significant differences were apparent for infants that were 7 months of age or older. Here, 50% of male bouts were ultimately associated with positive outcomes, but only 26.7% of female bouts. Conversely, 66.7% of female bouts but only 8.3% of male bouts ultimately resulted in negative outcomes, $X^2_{(2)}=10.2$, $p<0.01$.

Multivariate logistic regression analysis was used to identify independent acoustic predictors of maternal response as coded through Rsp/No-Rsp. Two subjects, female infant LN and male infant MS, contributed a disproportionately high number of bouts to this sample, 17 and 23 bouts, respectively. Using a random number generator, 15 bouts from each were selected for analysis, which reduced the sample to 66 bouts from nine individuals (six males, three females), with equivalent numbers of male ($M=7.7$, SD=6.68, range 1–15) and female bouts ($M=7.0$, SD=6.93, range 3–15). All variables that achieved univariate statistical significance at the 0.2 level were entered into a forward stepwise model selection procedure for the multivariate logistic regression analysis. Two variables, HNR and Peak1, were determined to be statistically significant predictors of maternal response in a multivariate model. Like HNR, Peak1 can be interpreted as a measure of relative noisiness, as this LPC coefficient tends to co-vary with the overall slope of the LPC function. However, the measures were nonetheless largely independent. After adjusting for Peak1, the odds of a maternal response were 1.42 times greater for each unit increase in HNR (95% confidence interval 1.2–2.1). After adjusting for HNR, the odds of a maternal response were 3.3 times greater for each unit increase in Peak1 (95% confidence interval 1.2–9.1). Overall, the likelihood of a maternal response significantly increased for vocalizers with higher HNR and Peak1 values. The multivariate logistic regression model correctly predicted the presence or absence of maternal response in 80.3% of cases, although prediction accuracy was substantially higher for instances of no response (92.9%) compared to when a response did occur (58.3%).

IV. DISCUSSION

Acoustic analyses of rhesus monkey gecker vocalizations during the first 24 months of life revealed age, sex, and maternal response differences across several temporal, spectral, and amplitude measures, but little context-specific acoustic differentiation. Geckers of the youngest infants had the shortest bout durations, pulse durations, and fewest pulses per bout. Female geckers showed higher spectral peaks and bout durations, while male geckers were higher in amplitude and less noisy. Developmentally, gecker usage peaked at four months of age for both sexes, with male geckers nonetheless tending to occur at younger ages than those of females. More than 75% of gecker bouts were produced by infants 6 months of age or less, but offspring as old as 17 months could also gecker. Over 20% of gecker bouts appeared to be spontaneous, while the majority of these calls were emitted when the offspring were following their mothers, receiving aggression from their mothers, or had been left behind. Acoustic measures showed some statistical power in discriminant-function classification of calls according to six different behavioral contexts associated with geckering, but modestly so. The measures discriminating most among these contexts were bout duration, pulse reduplication, and signal-to-noise ratio. Mothers responded most, and most positively to the geckers of young infants, while also clearly favoring males. Maternal response was also more likely when geckers showed a pronounced spectral peak and less noisiness.

A. Acoustics

1. Structure and function

Acoustic analysis confirmed that geckers are composed solely of multiple short pulses. With just one reliable fre-

quency peak (at just below 3000 Hz), geckers can be considered "spectrally structured noise" (Beeman, 1998). The noisiness of gecker pulses was reflected in a virtually flat spectral slope (Spectral-Tilt), low tonality (HNR), and high spectral standard deviation (Spectral-StDev) relative to spectral mean (Spectral-Mean). While the absolute amplitude of geckers was not measured, they are likely among the loudest vocalizations produced by young rhesus. Their noisiness therefore almost certainly reflects chaotic vocal-fold vibration (e.g., Tokuda *et al.*, 2002) rather than simple airflow turbulence. The occasional occurrence of periodic components within this noise-based spectral structure is also consistent with interpreting the sounds as example of deterministic chaos (Wilden *et al.*, 1998). This kind of chaos is often a hallmark of elevated vocal effort, for example, occurring as subglottal air pressures and vocal-fold tensions increase during vocal production (Wilden *et al.*, 1998; Fitch *et al.*, 2002; Brown *et al.*, 2003; Riede *et al.*, 2004). The occurrence of pulse reduplications in 34% of all gecker bouts can thus be taken as an indicator of additional vocal-fold instability due to vocal effort, with these events being particularly prevalent in longer geckers with a large number of pulses.

This acoustic structure suggests that geckers are both salient and localizable to listeners. Several features, including their abrupt, high-amplitude pulses and broadband atonal spectra, likely make geckers particularly difficult to ignore as an auditory event (see Owren and Rendall, 1997, 2001). These same features also suggest that geckers should be easy to localize in both vertical and horizontal dimensions (Brown, 1982; Heffner, 2004; Recanzone and Beckermann, 2004). Moreover, auditory localization is facilitated when sounds are produced in conjunction with salient visual events (e.g., Heffner, 2004), such as the dramatic, spasmodic whole-body jerking that can accompany geckering. Geckers are thus well designed to serve as signals of distress, and communicative significance likely adds additional salience for species-specific listeners. However, these same perceptual features can contribute to these sounds becoming annoying as well, particularly with prolonged use (e.g., Todt, 1988). In humans, for instance, there is ample evidence both that infant distress vocalizations are aversive to caregivers, and that the sounds become significantly more noxious when produced in long bouts (Frodi and Senchak, 1990; Levitzky and Cooper, 2000; Soltis, 2004). Noisy, so-called "dysphonia" is also prominent among the features found to have the greatest negative impact in human infant cries (Wood and Gustafson, 2001; Gustafson and Green, 1989; Dessureau *et al.*, 1998), a phenomenon now shown to be chaotic vocal-fold vibration (Herzel, 1993).

2. Age and sex

Ontogenetic trends mainly involved producing longer pulses, longer bouts, and more pulses per bout. Consistent with previous work on primate vocal production, these changes likely reflect maturational and growth processes rather than vocal learning per se (Hammerschmidt *et al.*, 2000; Hammerschmidt *et al.*, 2002). Sex differences, such as female gecker bouts lasting more than twice as long as those of males, were similar to outcomes reported for other distress calls in rhesus (Tomaszycki *et al.*, 2001; Erwin, 1975) and Japanese macaques (Green, 1981). The single characteristic spectral peak in geckers was also nearly 700 Hz higher in females than in males. Male geckers were somewhat less noisy (HNR), as well as being substantially higher in amplitude (SNR). These differences were apparent even after statistically controlling for body weight, suggesting that the critical factors may involve neuroanatomy, hormones, or vocal-fold size and shape, rather than body size and associated differences in vocal-tract length (Fitch, 1997).

Gonadal sex steroids are particularly likely to play a critical role, with Tomaszycki *et al.* (2001) finding that female rhesus up to 8 months emitted longer call bouts and used a greater variety of call types than did males of comparable age. However, when female fetuses were treated with androgen during the second trimester of gestation, the sex difference in later calling behavior was eliminated. The hormone treatment was likely to be operating via effects both on inferior temporal cortex (Newman and Bachevalier, 1997; Newman *et al.*, 1990) and on the vocal folds themselves (Aufdemorte *et al.*, 1983; Saez and Martin, 1976; Hollien *et al.*, 1994). Vocal-fold dimensions are largely unrelated to overall body size (Fitch, 1997; Rendall *et al.*, 2005), consistent with the current finding that sex differences in gecker acoustics persisted after statistically controlling for body size.

B. Usage

1. Age and sex

Although rhesus from 1 to 17 months of age emitted geckers, 75.7% of these calls occurred during the first 6 months of life. Geckering peaked at 4 months of age in both males and females, similar to results from Berman *et al.*'s (1994) more general study of rhesus distress calling. Ontogenetic peaks have also been found in other primate infant distress calls: stumptail macaque "trilled whistles" at 8 weeks (Maestripieri *et al.*, 1995), vervet monkey "care-elicitation" calls at 8–10 weeks (Hauser, 1993), chimpanzee "crying" at 6–8 weeks (Bard, 2000), and human infant crying at 6 weeks (Barr, 1990).

The time course involved may reflect the changing quality of the mother-infant relationship, with both free-ranging and captive rhesus infants beginning to spend time away from their mothers at about 4 months of age (Berman, 1980). A mother's first postpartum estrus also occurs around the same point, potentially causing significant increases in separation, distress calling and infant tantrums (Berman *et al.*, 1994). Males both began and stopped geckering earlier than females (also see Green, 1981), consistent with Newman *et al.*'s (1990) finding that female rhesus vocalize more than males during social separation in the second half of their first year. Erwin (1975) has argued that this sex difference in rhesus vocalization rates is very general, extending to "every age other than the period when the females have reached puberty and the males have not" (p. 376).

2. Accompanying context

Gecker production was not associated with any single behavioral context, with most geckers occurring when an offspring received maternal aggression (22%), when there was a proximity change such as mother moving away (19%), or when the vocalizer was following its mother (25%). Spontaneous geckers (22%) were also common. This finding is compatible with Newman's (1995) argument that at least some geckers are artifacts of basic nervous-system development rather than having signaling function per se. In a similar vein, Blass (1994) argues that human infant crying may function to maintain ideal levels of brain activation during early development. Nonetheless, spontaneous geckers may also reflect distress with purely internal causes, or could be related to external circumstances that the observer cannot readily see. At present, this issue must be considered unresolved.

Male geckering was found to be significantly more likely in the context of experiencing either aggression or affiliative behavior than was female geckering. However, the most striking sex difference appeared for offspring following their mothers, a context that accounted for almost half of all gecker bouts in females, but only 11% in males. One interpretation of this discrepancy is that females are less independent of their mothers than are males, particularly in early infancy (Lindell *et al.*, 2003; Simpson *et al.*, 1986; although see Brown and Dixon, 2000). If so, females might also become more distressed and vocal when following an inattentive mother (Lovejoy and Wallen, 1988).

C. Acoustics and context

The contexts found to be associated with geckering are consistent with interpreting these sounds as "separation-rejection" vocalizations, but the lack of context-specificity suggests they are not differentiated signals of specific need. While gecker acoustics supported statistically significant discriminant-function classification according to preceding or co-occurring context, categorization success was only 35% correct overall. The upshot is that the degree of acoustic differentiation is unlikely to attain the "just noticeable difference" (sensu Nelson and Marler, 1990) necessary for conspecifics to reliably infer the context of calling from acoustics alone. Gecker variation appears more likely to reflect differences in vocalizer arousal, particularly as pulse reduplication, signal-to-noise ratio, and bout duration were the primary variables contributing to successful statistical classification. Each of these measures probably reflects overall vocal effort, with increasing arousal plausibly associated with higher subglottal air pressures, greater vocal-fold tension, and more prolonged calling. This interpretation is consistent with Bayart *et al.*'s (1990) compelling results with coo calls produced by rhesus infants being separated from their mothers. Both infant behavior and associated coo acoustics were differentially affected by the degree of isolation, with concomitant effects found on hormone levels, monoamine neurotransmitter measures, and behavioral arousal. Total isolation produced increases in each of these domains, and triggered coos that were longer, harmonically richer, and more frequency modulated than those produced when mothers were visible to the infants in an adjacent cage.

D. Maternal response

Mothers showed an immediate response to only 38% of gecker bouts, and these reactions depended on a combination of gecker acoustics, offspring age, and whether the vocalizers were male or female. Mothers were most responsive to infants aged 4 months or less, and mothers were never observed to react to offspring older than 12 months. Maestripieri (1995) and Hauser (1993) have reported similar outcomes for stump-tailed macaques (*Macaca arctoides*) and vervet monkeys (*Cercopithecus aethiops*), respectively. Overall maternal responsiveness here was lower than in those studies, but probably because the current subjects were observed over a wider age range. Mothers were also more likely to react to male than to female geckers, and began to show responses earlier in the infant males' lives. Tomaszycki *et al.* (2001) similarly found that rhesus mothers responded more to males than to females across a variety of distress vocalizations, consistent with a general pattern of rhesus females showing somewhat greater parental investment toward sons than daughters (Bercovitch *et al.*, 2000).

Overall, the outcomes experienced by geckering infants were about equally likely to be positive or negative, and based again on discriminant-function analysis, gecker acoustics could not be used to predict the nature of the outcome, or whether there would be any discernible outcome. However, univariate analyses did show that geckers with longer bout and pulse durations tended to be associated with more negative outcomes such as aggression, while less noisy geckers were associated with more positive outcomes. An additional, striking difference was that for offspring 7 months or older, geckering was followed by aggression from mothers or other group members more than 50% of the time for females, but less than 10% of the time for males. Similarly, about half the gecker bouts from older male infants were associated with positive outcomes such as affiliation and attention, while the comparable figure for females was only about one-quarter.

These differences probably reflect a combination of factors, including the vocalizer's age and sex, as well as the potential aversiveness of geckers as auditory events. On the one hand, rhesus mothers could be less responsive to, but more negative toward females due to their higher calling rates, noisier geckers, and older ages when calling. Reinforcement learning has often been proposed as the mechanism by which rat pups (D'amato *et al.*, 2005), bird nestlings (Lotem, 1998; Stamps *et al.*, 1989), or rhesus monkey infants (Berman *et al.*, 1994) adjust their vocalizing towards optimal rates of effectiveness. Female rhesus infants may vocalize more often and intensively than males because mothers are rarely responding. Mothers may respond less to females because females are geckering excessively beyond the ages at which maternal response to geckers is crucial to offspring well being.

On the other hand, females may be more likely to exhibit these calling characteristics precisely because their mothers are less responsive and behave less positively to-

ward them. If the latter, mothers are behaving differently toward males and females for other reasons, and can in fact likely distinguish their respective geckers due to differences such as spectral peak frequencies. In addition, mothers were simply less likely to respond to calls from older individuals, which in most instances of which were females. Overall, then, while sex differences in both geckering and maternal responsiveness are apparent in these animals, the direction of causality involved between remains unclear.

V. CONCLUSIONS

Taken together, gecker vocalizations are likely to be highly salient and localizable to receivers. Gecker acoustics appear to be well designed to draw the attention of mothers and other listeners, while also potentially becoming aversive. These sounds become even more salient by virtue of regularly being accompanied by spasmodic jerking, occurring in lengthy bouts, and being associated with situations of evident vocalizer distress. While thus potentially being an effective distress cue, geckers are not highly specific to particular circumstances, and only infrequently elicit maternal response. The likelihood of maternal response was greatest when geckers exhibited pronounced spectral peaks and less noisiness (HNR).

Most gecker bouts were emitted within the first 6 months of life, with peak occurrence when infants were 4 months of age and mothers were experiencing their first postpartum estrus. After this age, maternal responsiveness diminished markedly. Male infants both began and stopped geckering earlier in life than females, and mothers responded more to their calls. Sex differences were also found in gecker acoustics, albeit based on an imperfectly balanced sample. Female vocalizations were characterized by longer durations and higher-frequency spectral peaks, while male geckers were generally less noisy (HNR) and higher in amplitude. Differences in neuroanatomy, hormones, and vocal-fold size and shape are all likely to contribute to these acoustic differences. Sex-biased maternal responsiveness is likely attributable to sex differences in gecker acoustics as well as the fact that females are emitting geckers at older ages than males.

ACKNOWLEDGMENTS

This research was supported in part by NIH Grant No. RR00160 to the California National Primate Research Center, by NICHHD Grant No. NS 19826 to Robert M. Seyfarth, Dorothy L. Cheney, and Michael J. Owren, by postdoctoral fellowship 5 F32 HD06878-03 to Michael J. Owren, and by NIMH Grant No. R01MH65317 to Robert D. Hienz, Michael J. Owren, and Drew Rendall. We thank C. Cornelius, R. Henrickson, A. Hendrickx, and W. A. Mason for facilities and support provided throughout the cross-fostered macaque vocal development project, and to J. Adams, A. Cabrera, P. Telfer and other members of the technical staff for expert assistance. Thanks to Rachel Pulverman for digitizing sound files, Steve Creech for lending his statistical expertise, Nick Lerche for finding crucial information about the rhesus mothers involved in the project, and two anonymous reviewers for providing thoughtful and constructive comments on the initial manuscript.

Altmann, J. (**1974**). "Observational study of behavior: Sampling methods," Behaviour **49**, 227–265.

Altmann, S. (**1962**). "A field study of the sociobiology of rhesus monkeys (*Macaca mulatta*)," Ann. N.Y. Acad. Sci. **102**, 338–435.

Aufdemorte, T. B., Sheridan, P. J., and Holt, G. R. (**1983**). "Autoradiographic evidence of sex steroid receptors in laryngeal tissues of the baboon," Laryngoscope **93**, 1607–1611.

Bard, K. A. (**2000**). "Crying in infant primates: Insights into the development of crying in chimpanzees," *In Crying As a Sign, a Symptom, & a Signal: Clinical Emotional and Developmental Aspects of Infant and Toddler Crying*, edited by R. G. Barr, B. Hopkins, and J. A. Green (Cambridge University Press, New York), pp. 157–175.

Barr, R. G. (**1990**). "The normal crying curve: What do we really know?," Dev. Med. Child Neurol. **32**, 356–362.

Bayart, F., Hayashi, K. T., Faull, K. F., Barchas, J. D., and Levine, S. (**1990**). "Influence of maternal proximity on behavioral and physiological responses to separation in infant rhesus monkeys (*Macaca mulatta*)," Behav. Neurosci. **104**, 98–107.

Beeman, K. (**1998**). "Digital signal analysis, editing and synthesis," in *Animal Acoustic Communication: Sound Analysis and Research*, edited by S. L. Hopp, M. J. Owren, and C. S. Evans (Springer-Verlag, Berlin), pp. 59–104.

Bercovitch, F. B., Widdig, A., and Nuernberg, P. (**2000**). "Maternal investment in rhesus macaques (*Macaca mullata*): Reproductive costs and consequences of raising sons," Behav. Ecol. Sociobiol. **48**, 1–11.

Berman, C. M. (**1980**). "Mother infant interaction among free ranging rhesus monkeys: A comparison with captive pairs," Anim. Behav. **28**, 860–873.

Berman, C. M., Rasmussen, K. R., and Suomi, S. J. (**1994**). "Responses of free ranging rhesus monkey infants to a natural form of social separation: Parallels with mother infant separation in captivity," Child Dev. **65**, 1028–1041.

Blass, E. M. (**2004**). "Changing brain activation needs determine early crying: A hypothesis," Behav. Brain Sci. **27**, 460–461.

Boersma, P. (**1993**). "Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound," Proc. Institut. Phonet. Sci. Amsterdam **17**, 97–110.

Boersma, P. (**2001**). "PRAAT, a system for doing phonetics by computer," Glot Internat. **5**, 341–345.

Brown, C. H. (**1982**). "Auditory localization and primate vocal behavior," in *Primate Communication*, edited by C. T. Snowdon, C. H. Brown, and M. R. Peterson (Cambridge University Press, New York), pp. 144–164.

Brown, C. H., Alipour, F., Berry, D. A., and Montequin, D. (**2003**). "Laryngeal biomechanics and vocal communication in the squirrel monkey (*Saimiri boliviensis*)," J. Acoust. Soc. Am. **113**, 2114–2126.

Brown, G. R., and Dixon, A. F. (**2000**). "The development of behavioural sex differences in infant rhesus macaques (*Macaca mullata*)," Primates **41**, 63–77.

D'amato, F. R., Scalera, E., Sarli, C., and Moles, A. (**2005**). "Pups call, mothers rush: Does maternal responsiveness affect the amount of ultrasonic vocalizations in mouse pups?," Adv. Food Res. **35**, 103–112.

Dessureau, B. K., Kurowski, C. O., and Thompson, N. S. (**1998**). "A reassessment of the role of pitch and duration in adults' responses to infant crying," Infant Behav. Dev. **21**, 367–371.

Erwin, J. (**1975**). "Rhesus monkey vocal sounds," in *The Rhesus Monkey, Vol.1: Anatomy and Physiology*, edited by G. H. Bourne (Academic, New York), pp. 365–380.

Fitch, W. T. (**1997**). "Vocal tract length and formant frequency dispersion correlate with body size in rhesus macaques," J. Acoust. Soc. Am. **102**, 1213–1222.

Fitch, W. T., Neubauer, J., and Herzel, H. (**2002**). "Calls out of the chaos: The adaptive significance of nonlinear phenomena in mammalian vocal production," Anim. Behav. **63**, 407–418.

Forrest, K., Weismer, G., Milenkovic, P., and Dougall, R. N. (**1988**). "Statistical analysis of word-initial obstruents: Preliminary data," J. Acoust. Soc. Am. **84**, 115–123.

Frodi, A., and Senchak, M. (**1990**). "Verbal and behavioral responsiveness to the cries of atypical infants," Child Dev. **61**, 76–84.

Gautier, J. P., and Gautier, A. (**1977**). "Communications in old world monkeys," in *How Animals Communicate*, edited by T. A. Sebeok (Indiana University Press, Bloomington), pp. 890–964.

Green, S. M. (**1981**). "Sex differences and age gradations in vocalizations of Japanese and lion-tailed monkeys (*Macaca fuscata* and *Macaca silenus*)," Am. Zool. **21**, 165–183.
Gustafson, G. E., and Green, J. A. (**1989**). "On the importance of fundamental-frequency and other acoustic features in cry perception and infant development," Child Dev. **60**, 772–780.
Hammerschmidt, K., Newman, J. D., Champoux, M., and Suomi, S. J. (**2000**). "Changes in rhesus macaque 'coo' vocalizations during early development," Ethology **106**, 873–886.
Hammerschmidt, K., Freudenstein, T., and Juergens, U. (**2002**). "Vocal development in squirrel monkeys," Behaviour **138**, 1179–1204.
Hauser, M. D. (**1993**). "Do vervet monkey infants cry wolf?," Anim. Behav. **45**, 1242–1244.
Heffner, R. S. (**2004**). "Primate hearing from a mammalian perspective," Anat. Rec. **218A**, 1111–1122.
Herzel, H. (**1993**). "Bifurcation and chaos in voice signals," Appl. Mech. Rev. **46**, 399–413.
Hoffman, R. A., and Stowell, R. E. (**1973**). "Outdoor housing of nonhuman primates," Lab. Anim. Sci. **23**, 74–83.
Hollien, H., Green, R., and Massey, K. (**1994**). "Longitudinal research on adolescent voice change in males," J. Acoust. Soc. Am. **96**, 2646–2654.
Jovanovic, T., and Gouzoules, H. (**2001**). "Effects of maternal restraint on the vocalizations of infant rhesus monkeys (*Macaca mulatta*)," Am. J. Primatol **53**, 33–45.
Kalin, N. H., Shelton, S. E., and Snowdon, C. T. (**1992**). "Affiliative vocalizations in infant rhesus macaques (*Macaca mulatta*)," J. Comp. Psychol. **106**, 254–261.
Levitzky, S., and Cooper, R. (**2000**). "Infant colic syndrome—maternal fantasies of aggression and infanticide," Clin. Pediatr. (Phila) **39**, 395–400.
Lindell, S. G., Shannon, C., Suomi, S. J., and Higley, J. D. (**2003**). "Sex differences in rhesus macaque mother-infant behavior," Am. J. Primatol **60**, 130–131.
Locke, J. L., and Hauser, M. D. (**1999**). "Sex and status effects on primate volubility: Clues to the origin of vocal languages?," Evol. Hum. Behav. **20**, 151–158.
Lotem, A. (**1998**). "Differences in begging behaviour between barn swallow, *Hirundo rustica*, nestlings," Anim. Behav. **55**, 809–818.
Lovejoy, J., and Wallen, K. (**1988**). "Sexually dimorphic behavior in group-housed rhesus monkeys (*Macaca mulatta*) at 1 year of age," Psychobiol. **16**(4), 348–356.
Maestripieri, D. (**1995**). "Maternal responsiveness to infant distress calls in the stumptail macaque," Folia Primatol. **64**, 201–206.
Maestripieri, D., and Call, J. (**1996**). "Mother-infant communication in primates," Adv. Stud. Behav. **25**, 613–642.
Maestripieri, D., Jovanovic, T., and Gouzoules, H. (**2000**). "Crying and infant abuse in rhesus monkeys," Child Dev. **71**, 301–309.
Nelson, D. A., and Marler, P. (**1990**). "The perception of birdsong and an ecological concept of signal space," in *Comparative Perception, Vol. 2, Complex Signals*, edited by W. D. Stebbins and M. A. Berkley (Wiley, New York), pp. 443–478.
Newman, J. D. (**1995**). "Vocal ontogeny in macaques and marmosets: Convergent and divergent lines of development," in *Current Topics in Primate Vocal Communication*, edited by E. Zimmermann, U. Jürgens, and D. Symmes (Plenum, New York), pp. 73–97.
Newman, J. D., and Bachevalier, J. (**1997**). "Neonatal ablations of the amygdala and inferior temporal cortex alter the vocal response to social separation in rhesus macaques," Brain Res. **758**, 180–186.
Newman, J. D., Bachevalier, J., Michjeda, M., and Suomi, S. J. (**1990**). "A possible neural substrate for gender differences in vocal behavior by rhesus macaque infants during brief periods of social separation," Soc. Neuro. Abstr. **16**, 599.
Owren, M. J., and Bernacki, R. H. (**1998**). "Applying linear predictive coding (LPC) to frequency-spectrum analysis of animal acoustic signals," in *Animal Acoustic Communication: Sound Analysis and Research*, edited by S. L. Hopp, M. J. Owren, and C. S. Evans (Springer-Verlag, Berlin), pp. 129–161.
Owren, M. J., and Dieter, J. A. (**1989**). "Infant cross-fostering between Japanese (*Macaca fuscata*) and rhesus macaques (*M. mulatta*)," Am. J. Primatol, **18**, 245–250.
Owren, M. J., Dieter, J. A., Seyfarth, R. M., and Cheney, D. L. (**1992a**). "'Food' calls produced by adult female rhesus (*Macaca mulatta*) and Japanese (*M. fuscata*) macaques, their normally-raised offspring, and offspring cross-fostered between species," Behaviour **120**, 218–231.
Owren, M. J., Dieter, J. A., Seyfarth, R. M., and Cheney, D. L. (**1992b**). "Evidence of limited modification in the vocalizations of cross-fostered rhesus (*Macaca mulatta*) and Japanese (*M. fuscata*) macaques," in *Topics in Primatology, Vol. 1: Human Origins*, edited by T. Nishida, W. C. McGrew, P. Marler, M. Pickford, and F. de Waal (University of Tokyo Press, Tokyo), pp. 257–270.
Owren, M. J., Dieter, J. A., Seyfarth, R. M., and Cheney, D. L. (**1993**). "Vocalizations of rhesus (*Macaca mulatta*) and Japanese *(M. fuscata*) macaques cross-fostered between species show evidence of only limited modification," Dev. Psychobiol. **26**, 389–406.
Owren, M. J., and Rendall, D. (**1997**). "An affect-conditioning model of nonhuman primate vocal signaling," in *Perspectives in Ethology: Vol. 12. Communication*, edited by D. H. Owings, M. D. Beecher, and N. S. Thompson (Plenum, New York), pp. 299–346.
Owren, M. J., and Rendall, D. (**2001**). "Sound on the rebound: Bringing form and function back to the forefront in understanding nonhuman primate vocal signaling," Evol. Anthro. **10**, 58–71.
Recanzone, G. H., and Beckermann, N. S. (**2004**). "Effects of intensity and location on sound location discrimination in macaque monkeys," Hear. Res. **198**, 116–124.
Rendall, D., Kollias, S., Ney, C., and Lloyd, P. (**2005**). "Pitch (F_0) and formant profiles of human vowels and vowel-like baboon grunts: The role of vocalizer body size and voice-acoustic allometry," J. Acoust. Soc. Am. **117**, 944–955.
Riede, T. R., Owren, M. J., and Clark Arcadi, A. (**2004**). "Nonlinear acoustics in the pant-hoot vocalizations of common chimpanzees (*Pan troglodytes*): Frequency jumps, subharmonics, biphonation, and deterministic chaos," Am. J. Primatol **64**, 277–291.
Rowell, T. E. (**1962**). "Agonistic noises of the rhesus monkey (*Macaca mulatta*)," Sym. Zool. Soc. London **8**, 91–96.
Rowell, T. E., and Hinde, R. A. (**1962**). "Vocal communication by the rhesus monkey (*Macaca mulatta*)," Proc. R. Soc. London **138**, 279–294.
Saez, S. J., and Martin, P. M. (**1976**). "Androgen receptors in human pharyngo-laryngeal mucosa and pharyngo-laryngeal epithelium," J. Steroid Biochem. **7**, 919–921.
Seyfarth, R. M., and Cheney, D. L. (**1986**). "Vocal development in vervet monkeys," Anim. Behav. **34**, 1640–1658.
Simpson, M. J. A., Simpson, A. E., and Howe, S. (**1986**). "Changes in the rhesus mother-infant relationship through the first four months of life," Anim. Behav. **34**, 1528–1539.
Snowdon, C. T., Elowson, A. M., and Roush, R. S. (**1997**). "Social influences on vocal development in New World primates," in *Social Influences on Vocal Development*, edited by C. T. Snowdon and M. Hausberger (Cambridge University Press, New York), pp. 234–248.
Soltis, J. (**2004**). "The signal functions of early infant crying," Behav. Brain Sci. **27**, 443–490.
Stamps, J., Clark, A., Arrowood, P., and Kus, B. (**1989**). "Begging behavior in budgerigars," Ethology **81**, 177–192.
Tabachnik, B. G., and Fidell, L. S. (**2001**). *Using Multivariate Statistics*, 4th ed. (Allyn and Bacon, New York).
Todt, D. (**1988**). "Serial calling as a mediator of interaction processes: Crying in primates," in *Primate Vocal Communication*, edited by D. Todt, P. Goedeking, and D. Symmes (Springer-Verlag, Berlin), pp. 88–107.
Tokuda, I., Riede, T., Neubauer, J., Owren, M. J., and Herzel, H. (**2002**). "Nonlinear prediction of irregular animal vocalizations," J. Acoust. Soc. Am. **111**, 2908–2919.
Tomaszycki, M. L., Davis, J. E., Gouzoules, H., and Wallen, K. (**2001**). "Sex differences in infant rhesus macaque separation-rejection vocalizations and effects of prenatal androgens," Horm. Behav. **39**, 267–276.
Wallen, K. (**2005**). "Hormonal influences on sexually differentiated behavior in nonhuman primates," Front Neuroendocrinol **26**, 7–26.
Wilden, I., Herzel, H., Peters, G., and Tembrock, G. (**1998**). "Subharmonics, biphonation, and deterministic chaos in mammal vocalization," Bioacoustics **9**, 171–196.
Wood, R. M., and Gustafson, G. E. (**2001**). "Infant crying and adults' anticipated caregiving responses: Acoustic and contextual influences," Child Dev. **72**, 1287–1300.

Sound characteristics of big-snout croaker, *Johnius macrorhynus* (Sciaenidae)

Yuan C. Lin
Department of Environmental Biology and Fisheries Science, National Taiwan Ocean University, Keelung, Taiwan 202, R.O.C.

Hin K. Mok[a)]
Institute of Marine Biology, National Sun Yat-sen University, Kaohsiung, Taiwan 804, R.O.C.

Bao Q. Huang
Department of Environmental Biology and Fisheries Science, National Taiwan Ocean University, Keelung, Taiwan 202, R.O.C.

(Received 15 June 2006; revised 24 September 2006; accepted 6 October 2006)

The sounds of a big-snout croaker, *Johnius macrorhynus*, produced under hand-held and voluntary conditions (in a large aquarium and in the field) were compared. Voluntary calls included "purr" and "dual-knocks", only purrs were produced when the fish was hand-held. The purr is composed of pulses in which the first interpulse interval was six to nine times longer than the other interpulse intervals, which were approximate in duration, and is a unique sound type in the coastal water of Taiwan. Purrs emitted under these conditions did not differ significantly, suggesting that the hand-held sound can be employed to match the sound in the field. These sounds contained energy reaching 5 kHz, with two peaks at about 1 and 2 kHz—the former being the dominant frequency. First interpulse interval, main interpulse interval, repetition rate of pulse, and pulse duration may serve as the diagnostic characters for the species-specific sound (i.e., purrs). One specimen survived in a large aquarium for 6 months and its vocal activity was monitored. The fish produced fewer dual-knocks than purrs, and purrs were about 11 dB louder than dual-knocks emitted in the aquarium. The temporal sequential relationship of these types in the sound of this individual was described. [DOI: 10.1121/1.2384844]

PACS number(s): 43.80.Ka [AU] Pages: 586–593

I. INTRODUCTION

In a number of teleostean families, such as Batrachoididae, Gadididae, Holocentridae, Sciaenidae, Pomacentridae, and Gobiidae, sound production is associated with defensive and/or reproductive purposes (Hawkins and Rasmussen, 1978; Fine, 1997; Ladich, 1997). For the latter purpose, acoustic activity is limited to and peaked in the spawning season and restricted to the spawning ground (Fish and Cummings, 1972; Fine, 1978; Takemura *et al.*, 1978; Mok and Gilmore, 1983; Saucier and Baltz, 1993). Sciaenids are fisheries species commonly found in shallow coastal waters (Shao, 2005) and are well known for being soniferous (Mok and Gilmore, 1983). Specific sound types hypothesized to be made by particular sciaenid species have been employed to track and delimitate the aggregations of mature individuals and their spawning grounds (Mok and Gilmore, 1983; Saucier and Baltz, 1993). Such a passive-sonar technique has revealed information about the relative size and location of the vocalizing aggregation (Connaughton and Taylor, 1995; Luczkovich *et al.*, 1999). Information on the species-specific sound is required so that distribution of the species can be properly estimated. Unfortunately, only the sounds from the field of a few sciaenid species were verified using voluntary sounds recorded under captivity [e.g., weakfish, *Cynoscion regalis*, and Atlantic croaker, *Micropogonias undulatus* (Connaughton *et al.*, 2000; Fine *et al.*, 2004)]. As sciaenids seem very sensitive to handling, successful captivity is hard to come by and sounds of most sciaenid species remain unknown. To make thing even more difficult, most sciaenid croaking sounds are composed of a series of knocks whose interspecific differences were indistinctive. Sounds recorded from various distances from the sources may have already been subjected to different degrees of distortion and diverged from their original forms. These factors further complicate accurate matching of sciaenid sounds from the field. An easier approach to obtain the acoustic information for a species is from the sounds emitted by a captive live fish responding to disturbance. As the releasing factors for the disturbance sounds and those from the natural environment differ, one would doubt whether they will be the same. Due to the high similarity among species in their acoustic characters, reliable matching between a disturbance sound from a known producer and a particular sound type taken from the field remains problematic in most cases. The condition gets more complicated in areas where several sciaenid species co-occur.

In our past surveys on the underwater ambient biological sounds in areas adjacent to the estuaries of nine major rivers in Taiwan, eight types of sounds were recognized. The majority of these sound types are composed of trains of pulses

[a)]Electronic mail: hinkiu@mail.nsysu.edu.tw

and the interpulse intervals within each train varied to a certain degree. One of these sound types was special in having (1) the first interpulse interval obviously longer than the following intervals (or the main interpulse interval), (2) the peak-to-peak interval of a pulse being about 10 ms, and (3) the sound energy spreading to 5 kHz. We chose this sound type as a target subject. The objectives of the present study are to (1) find the producer of this target sound type; (2) compare its sounds captured from the field, under hand-held (or disturbance) and captive conditions such that the value of the disturbance sound, which is easier to obtain, for identifying the producer from the field can be evaluated; and (3) describe the factors affecting the characteristics of the sound(s) produced by the species in question. We chose it as a target subject because (1) it is common in these survey areas; (2) its characteristics are very unique and can hardly be misidentified; and (3) if the diagnostic character (a longer first interpulse interval) is present, any other changes observed in the sound can be treated as intraspecific variation. If a sciaenid species caught from areas where this target sound type is present emits voluntary or disturbance sounds carrying such diagnostic character, it will be a nice model for studying factors affecting its characteristics and distribution of the reproductively active individuals.

II. MATERIALS AND METHODS

A. Sound recording

Sound recordings were made on the coastal waters adjacent to the estuaries of nine major rivers (including the Danshuei River, Toucian River, Dajia River, Dadu River, Jhuoshuei River, Cengwun River, Gaoping River, Lanyang River, and Siouguluan River) and one islet (Gueishan Island) off the northeastern coast of Taiwan during 2002 and 2004. Recordings were made in all seasons. To find out the producer of the target sound type, a site close to the estuary of the Jhuoshuei River ($23°54'–23°56'$N, $120°11'–120°14'$E; water depth ca. 10 m) was selected where the target sound type had been recorded to capture live soniferous fish specimens by hook-and-line for the disturbance hand-held sound. Some of the live fish were brought back to the wet laboratory and kept in an aquarium. If hand-held disturbance sounds and voluntary sounds of one of these live fish emitted under captivity turns out to be the same as the target sound type heard in the field, the sound producer of this target sound type is considered identified.

Eighty-eight live specimens of six sciaenid species (5 *Johnius belengerii*, 31 *Johnius macrorhynus*, 27 *Johnius tingi*, 6 *Otolithes rubber*, 4 *Pennahia macrocephalus*, and 15 *Pennahia pawak*) were treated for sound recording. While they were still alive right after the hook was removed, they were gently held and placed into a $52\times32\times30$ cm^3 polystyrene tank in which the water depth was 20 cm. The sounds emitted when they were hand-held in water were recorded by the hydrophone placed in the tank. The hydrophone was placed at a distance of about 10 cm from the calling fish and about 15 cm under the water surface. After recording, some specimens were sacrificed to determine the sex, and the sonic muscles were also examined. The sounds were recorded using either a Sony Net MD walkman recorder (MZ-N910; sampling frequency of 44.1 kHz) or a Sony TC-D5M cassette-recorder connected to a hydrophone system (HP-A1 hydrophone series, Burns Electronics, with frequency responses of 10–25 000 Hz±3 dB, and HP-A1 Mixer-amplifier).

Thirty-one big-snout croaker, *J. macrorhynus* (138 ± 16 mm total length, range from 105 to 172 mm) were caught between 23 May and 11 November 2005. Ten were transported to the wet laboratory at the National Sun Yat-sen University and kept in a $190\times85\times42$ cm^3 glass aquarium with an external water-filter system connected to aquarium. One male big-snout croaker specimen (132-mm total length) successfully survived until November. Its sounds were noticed for the first time on 14 October when a student heard the loud sounds while staying in the wet laboratory very late in the evening. This finding led us to a close monitoring of its vocal activity. We formally began to monitor its voluntary calls from 17 October to 24 November 2005. The fish passed away afterward. Sound production might have taken place prior to 14 October. For prolonged and continuous recording of sound, the HP-A1 hydrophone system was connected to a Creative Nomad JukeBox 3 recorder with a sampling rate of 22.05 kHz.

B. Sound analysis

Data logged in the Sony Net MD Walkman recorder and Sony TC-D5M cassette-recorder were digitized using the software Avisoft-SASLab Pro (Specht, 2002) at a sampling rate of 16 kHz and a resolution of 16 bits. A known sound pressure level (SPL) of 1 kHz sound (measured by another calibrated B & K 8104 hydrophone system) generated through a DNH Aqua-30 underwater loudspeaker (made in Norway) was employed to calibrate the laboratory voluntary calls using the Creative Nomad JukeBox 3 recorder hydrophone system. Only sounds that had clear pulse structure were analyzed. The sonograms were computed using frequency resolution of 256-point fast Fourier transformation (FFT), 50% frame, hamming window, and temporal resolution by 96.87% overlap. The call duration, number of pulses per call, duration of pulse, and the interval between pulses (interpulse interval) were measured from the sonogram. The repetition rate of the pulse was calculated as the number of pulses per second. The two main frequency peaks were obtained for each call with hamming evaluation window. Sound pressure level (SPL) of the voluntary sounds recorded in the large glass aquarium was measured by root mean square. A pulse always consists of four to five cycles of acoustic energy. The first four durations of acoustic energy were measured for each pulse. The calculated acoustic frequencies were the inverse of acoustic energy cycle duration $[1/(\text{duration})\times1000]$ (Connaughton *et al.*, 2000).

C. Statistical analysis

Nonparametric tests including the Kruskal-Wallis test and Mann-Whitney test were used. The critical α level for these analyses was 0.05.

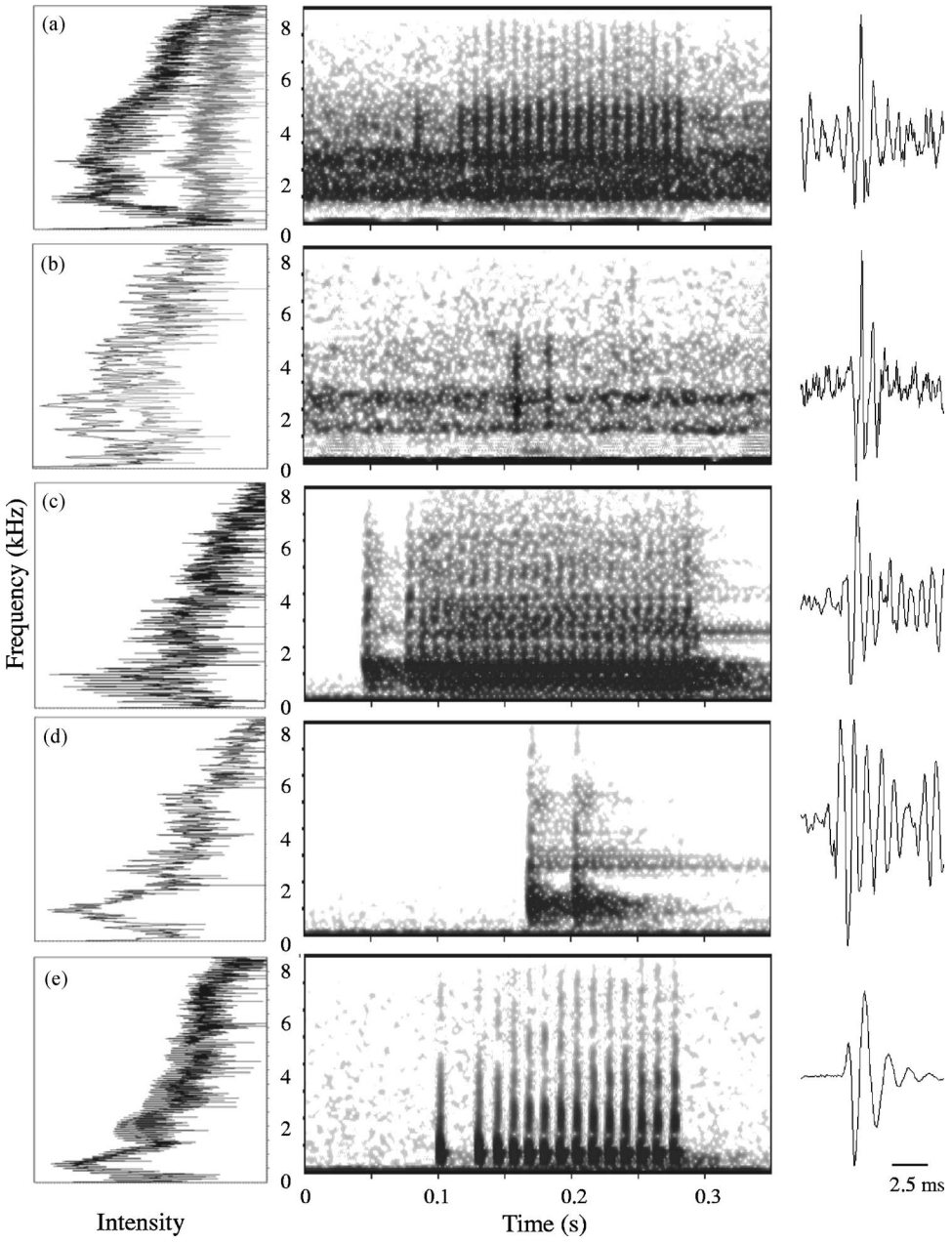


FIG. 1. Two sound types emitted by the big-snout croaker—purr (a, c, e) and dual-knocks (b,d). Power spectrum (left), sonogram (middle), and single-pulse oscillograms (right) for sounds recorded in the field (a, b), large aquarium (c, d), and small tank (e). The lower power spectrum in (a) and (b)represents the ambient noise in which the fish sound was absent.

III. RESULTS

A. Characteristics of fish sounds

Among the six species caught by hook-and-line that produced disturbance sounds, only male big-snout croakers produced the target sound while hand-held. However, the females of this species did not emit any sound under a similar condition. This species is sexually dimorphic in possession of sonic muscles; only males have these muscles. Two types of sounds were produced by this species: the purr, which is identical to the target sound type, and the dual-knocks. Purr is composed of a train of pulses in which the first interpulse interval is statistically longer than the rest of the interpulse intervals that have approximate durations (also see below). Dual-knocks are composed of only two pulses. Both sound types were emitted by the fish in the field and in the large tank. However, when the fish was hand-held, they emitted only the purr sound, which was species specific.

For the purrs recorded from the field, the first interpulse interval ranged from 17.8 to 59.4 ms (34.0±10.5 ms, N=40). Other interpulse intervals within a purr showed low variation; the average of these intervals, which was called the main interpulse interval, ranged from 2.0 to 9.9 ms (3.6±0.9 ms). Consequently, the first interpulse interval was about six to nine times the main interpulse interval.

The number of pulses in a purr ranged from 8 to 28 pulses (15.9±6.1 pulses, N=40). The call duration was 181.3±52.6 ms (range: 111.0–281.5 ms). The number of pulses per call and call duration were positively correlated. Pulses contained energy reaching 5 kHz, with two peaks at about 1 and 2 kHz (Fig. 1, Table I), the former being the dominant frequency. The duration and number of pulses in the purrs recorded from the large tank, field, and under disturbance (hand-held in a small tank), listed by decreasing order of duration and number of pulses, were significantly

TABLE I. Parameters of big-snout croaker sound emitted in two voluntary conditions (i.e., reared in a large aquarium and in the field) and hand-held disturbance condition, a, b, c, and x, y represent distinct groups in the purr and in the dual-knocks, respectively. The first six parameters were averaged for 40 purrs and 15 dual-knocks. The last parameter was measured independently for each pulse: 250 pulses in purr from the field, 60 pulses in purrs from the large aquarium, 140 pulses in the purrs from the small tank, 40 pulses in dual-knocks from the field, 50 pulses in the dual-knocks from the large aquarium were separately averaged.

		Field		Large aquarium		Hand-held
Parameters		Purr	Dual-knocks	Purr	Dual-knocks	Purr
Call duration (ms)		181.3 ± 52.6^{a}	42.8 ± 5.3^{x}	256.9 ± 29.8^{b}	45.6 ± 2.9^{x}	113.6 ± 31.8^{c}
No. of pulses per call (no.)		15.9 ± 6.1^{a}	2	22.0 ± 3.9^{b}	2	11.2 ± 4.3^{c}
Repetition rate of pulse (pulses/second)		85.9 ± 10.8^{a}	47.5 ± 6.8^{x}	85.2 ± 8.6^{a}	44.0 ± 2.8^{x}	88.5 ± 10.5^{a}
Pulse duration (ms)		6.1 ± 0.9^{a}	6.2 ± 0.8^{x}	6.3 ± 0.6^{a}	5.9 ± 1.1^{x}	6.2 ± 0.8^{a}
Interpulse interval (ms)		First: 34.0 ± 10.5^{a}	30.5 ± 4.9^{x}	First: 29.0 ± 3.6^{a}	33.5 ± 4.5^{x}	First: 16.0 ± 7.3^{b}
		Main: 3.6 ± 0.9^{a}		Main: 4.1 ± 1.2^{a}		Main: 3.3 ± 1.4^{a}
Frequency Peak (Hz)	Low (Dominant)	1146 ± 131^{a}	1133 ± 119^{x}	1052 ± 84^{a}	1135 ± 85^{x}	808 ± 142^{b}
	High	2481 ± 269^{a}	2301 ± 136^{x}	2263 ± 186^{b}	2351 ± 196^{x}	1676 ± 157^{c}
Acoustic energy	First	0.83 ± 0.12^{a}	0.82 ± 0.16^{x}	0.87 ± 0.19^{a}	0.85 ± 0.14^{x}	1.31 ± 0.17^{b}
cycle duration (ms)	Second	0.95 ± 0.11^{a}	0.80 ± 0.10^{x}	0.90 ± 0.13^{b}	0.87 ± 0.12^{y}	1.54 ± 0.31^{c}
	Third	1.00 ± 0.14^{a}	0.82 ± 0.08^{x}	0.91 ± 0.11^{b}	0.88 ± 0.18^{y}	1.44 ± 0.40^{c}
	Fourth	0.98 ± 0.15^{a}	0.88 ± 0.14^{x}	0.94 ± 0.13^{a}	0.92 ± 0.16^{x}	1.56 ± 0.28^{b}

different (call duration: 256.9±29.8 ms > 181.3±52.6 ms > 113.6±31.8 ms; number of pulses per call: 22.0±3.9 > 15.9±6.1 > 11.2±4.3; $P<0.001$). However, the repetition rate of the pulse did not differ among these three conditions ($P=0.2099$). The first interpulse duration and the low-frequency peak were higher in the purrs from the field and large tank than those from the disturbance condition (34.0±10.5 ms and 29.0±3.6 ms > 16.0±7.3 ms; 1146±131 and 1052±84 Hz > 808±142 Hz; $P<0.001$). Pulse durations and main interpulse intervals among the purrs emitted under these three conditions showed no significant difference ($P=0.3243$). The high-frequency peak of the purrs at about 2 kHz among the three recording conditions differed: purrs in the field, large-tank, and disturbance conditions were listed in decreasing frequency (2481±269 Hz > 2263±186 Hz > 1676±157 Hz; $N=40$).

Dual-knocks was composed of only two pulses, with an interpulse interval of 19.7–39.5 ms (30.5±4.9 ms) in the field and 36.0–40.2 ms (33.5±4.5 ms) in the large aquarium tank ($N=15$, Table I). Pulses contained energy also reaching 5 kHz, with two peaks at about 1 and 2 kHz (Fig. 1, Table I), the former being the dominant frequency. The call duration, repetition rate of pulse, pulse duration, interpulse interval, and frequency distribution of the dual-knocks emitted in the field and in the large tank were not significantly different (Table I).

The durations of the four cycles in the pulses of the two sound types emitted in the field and the large glass aquarium ranged from 0.83 to 1.00 ms (Fig. 1, Table I). However, the duration for the purr sound emitted when the fish was hand-held ranged form 1.31 to 1.56 ms; it was significantly longer than the other two conditions (Table I). The calculated acoustic frequencies were between 1016 and 1226 Hz in the field and large aquarium conditions and 634 to 778 Hz in the hand-held disturbance condition, respectively. The calculated acoustic frequencies were close to their dominant frequencies as exhibited in the sonograms and spectrograms (Fig. 1).

B. Voluntary vocal activity under captivity

The single male big-snout croaker kept in the large tank emitted many loud sounds that could be clearly heard about 10 m from the tank. No fish sounds had been found in the daytime and sound production was limited between 21:45 to 03:15 h. The fish made short pauses in sound production in the vocalizing evening. Duration and number of pauses varied. The longest "continuous" vocalization in one night with only short pauses lasted for about 143 min with 2322 purrs and 77 dual-knocks. Cumulated calling times per night could reach a total of about 189 min with about 2281 purrs and 337 dual-knocks. However, sound production per night could also last as short as only 2 min with 4 purrs and 9 dual-knocks. The number of purrs per night ranged from 4 to 2322 (average: 876.4±744.7) and the dual-knocks ranged from 9 to 689 (average: 284.6±218.7). In the total 1745-min sound records, 17 528 purrs and 5693 dual-knocks were detected (Fig. 2).

Under the captive condition the two types of sound did not occur randomly during the night. Dual-knocks were always the sole sound type presented uninterruptedly in the initial 3 to 5 min after the initiation of sound. The purrs then emerged and became the major composition. However, vocal activity might stop momentarily and about 10–30 dual-knocks could be emitted when vocalization restarted. As for the sound pressure levels, purrs were higher (or louder) than those of the dual-knocks (purrs: 132.1±5.6 dB versus dual-knocks: 121.0±5.1 dB; $P<0.0001$).

IV. DISCUSSION

The exceptionally small sample of the sound source in captivity (i.e., only one fish in the aquarium) in this study is probably not representing the real phenomenon and a lower variation could be expected (i.e., a lower variation in the temporal parameters of the purrs in the sounds in the aquarium emitted by the same fish). Despite the variations

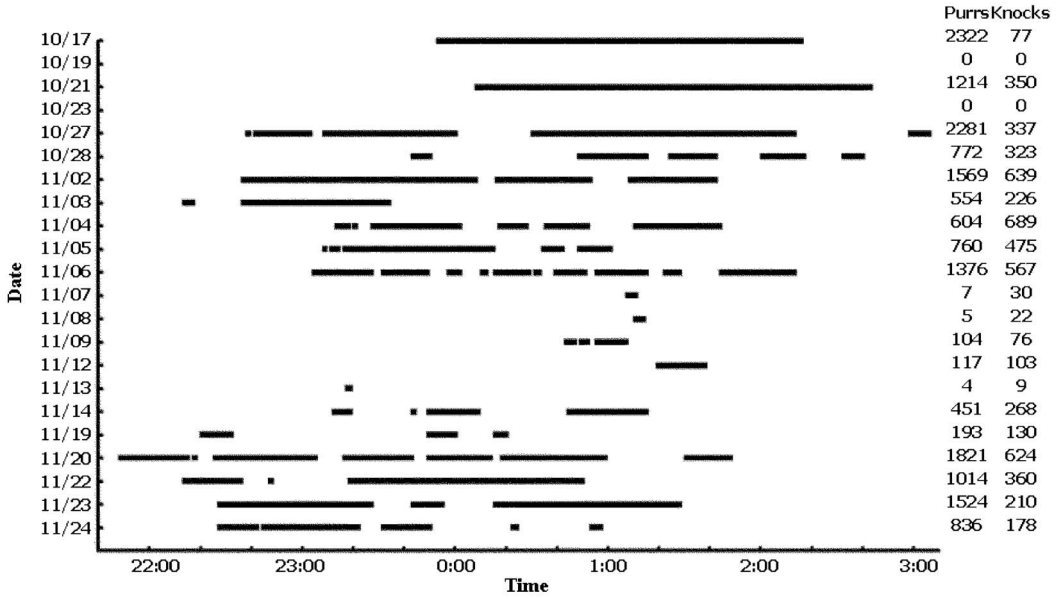


FIG. 2. The voluntarily calling time and the number of knocks and purrs of a big-snout croaker recorded in the glass aquarium from 17 October to 24 November 2005. Black bar indicates presence of sound.

found in call duration and number of pulses per call among the three recording conditions, the diagnostic characteristics of the purrs (i.e., the first interpulse interval being six to nine times longer than the main interpulse interval, and the approximate 10-ms pulse peak-to-peak interval) produced by the male big-snout croakers in the hand-held disturbance condition and produced voluntarily in the large aquarium agreed with the target sound type from the field. Consequently, the producer of the target sound type in the field is identified as *J. macrorhynus*. The main species-specific characters of this species pertained in all three conditions. It is, therefore, reliable to use the disturbance sound to represent the main sound type of the species. Implication of the hand-held disturbance sound for identifying the producer of sounds in the field may be generalized and applied to other soniferous fishes, especially to those that are not easy to hold in captivity. For the big-snout croaker, interpulse interval (both the first and the main ones), pulse repetition rate, and pulse duration are conservative and informative parameters for species-specific sound recognition.

The first interpulse interval of the purrs shortened when the fish was under stress and the reduction correlated to the degree of stress (i.e., the shortest interval was found in the hand-held condition). The duration of the calls under intensive stress (i.e., hand-held condition), which was related to the number of pulses included, was also the shortest. However, the main diagnostic character was sturdy (the exceptionally larger first interpulse interval) even under extremely stressful condition. Interestingly, stress did not affect the main interpulse interval; pulse repetition rate in the main portion of the pulse train between the second and last pulses was a persistent parameter of the purr. As such, this parameter may also be a useful one for species identification when overlap of the ranges among species is limited. Actually, pulse repetition rates in the disturbance sounds of other soniferous species we recorded differed from that of the big-snout croaker (*J. belengerii*: 65.2±8.1 pulses/second, $n=104$; *J. tingi*: 98.2±12.5 pulses/second, $n=301$; *O. rubber*: 142.9±19.1 pulses/second, $n=46$; *P. macrocephalus*: 33.8±5.3 pulses/second, $n=39$; *P. pawak*: 113.4±6.5 pulses/second, $n=63$).

The sound pressure level recorded in the aquarium could either be decreased by the reflecting sound waves, which were out of phase with the source, or be increased by the resonance of the tank. However, since both sound types were recorded under the same condition, the 11-dB difference between dual-knocks and purrs should be real. For the big-snout croaker sounds in the field and in the large glass aquarium, the number of dual-knocks was much fewer than that of purrs. In the former condition, the lower SPL of the dual-knocks may account for such a reduction as a result of sound dissipation (i.e., the hydrophone could not pick up dual-knocks from a remote distance). However, this explanation cannot be applied to the voluntary sound in the large tank. It is concluded that the fish actually produced fewer dual-knocks than purrs.

A "stressful" condition may affect vocal performance (e.g., missing certain sound type); the wild Atlantic cod, *Gadius morhua*, produced grunts, short knocks, and long series of knocks, while individuals in captivity only produced the first two sound types (Midling *et al.*, 2002). Similar result was found in the big-snout croaker; only purrs were emitted under the hand-held disturbance condition, whereas this sound type and dual-knocks were emitted under the voluntary conditions (i.e., in the large tank and in the field), despite that we only recorded less than 100 dual-knocks in our past field surveys between 2002 and 2004. In other words, in the case of the big-snout croaker, the major sound type will be kept even in a stressful condition. Due to the rarity of double-knocks, it remains possible that this sound type could be emitted by a hand-held big-snout croaker if the number of fish tested increased. Its absence could, also be due to the

fact that it can be released only by a mild stress.

For the vast majority of sciaenid species, only a sound type that may show certain quantitative variation [e.g., number of pulses and call duration (Guest and Lasswell, 1978; Luczkovich *et al.*, 1999; Connaughton *et al.*, 2002b)] is emitted, and more distinct sound types within a species have seldom been found in sciaenids. Mok and Gilmore (1983) recognized three sound types in the spotted seatrout (*Cynoscion nebulosus*) (i.e., grunt, knocks, and staccato), but no evidence of confirmation was presented. The male Atlantic croaker produced courtship sound, and both male and female produced another fright-response sound that could be elicited by casting a shadow over the surface of the holding tank or by moving a dip net through the water (Connaughton *et al.*, 2002b). The courtship sound was lower in frequency and pulse repetition rate than the fright-response sounds. These differences do not result in two very distinctive types. The two big-snout croaker sound types can be considered distinctive. The dual-knocks seem not to be a fright-response sound as obvious environmental stimuli were not expected to have taken place in the large aquarium prior to the production of the sounds. The big-snout croaker dual-knocks often being placed in the initial phase of a series of call and the purrs often being chained together suggest the communication signal of this species is actually not very diverse. A similar condition may take place in the toadfish, *Opsanus tau*, where the main boatwhistle is preceded by short grunts (Mok, pers. obs.). In the gulf toadfish, *Opsanus beta*, a long tonal boopnote of the boatwhistle (the courtship sound) can also be preceded by zero to three introductory grunts [considered as an agonstic sounds (Fine, 1982; Thorson and Fine, 2002a, b)]. The reason for a higher intraspecific diversity with more distinctive sound types is hard to imagine as there are not many sound producing mechanisms within the same species.

Acoustic energy in the purrs and dual-knocks peaked at about 1 and 2 kHz—the former peak was higher than the latter. Connaughton *et al.* (2000, 2002a) reported that the sonic muscles in weakfish, instead of the resonant frequency of the swimbladder, determine the dominant frequency of its sound. However, the swimbladder might still affect the sound characteristics (Sprague, 2000). The 2-kHz peak might possibly be the second harmonic of the fundamental frequency (i.e., 1-kHz peak).

The maximum standard body length of the big-snout croaker is 30 cm, which does not differ much from most of the other sciaenid species tested for the hand-held sound in the present study (i.e., *J. belengerii*, 30 cm; *J. tingi*, 22 cm; *O. rubber*, 90 cm; *P. macrocephalus*, 23 cm; *P. pawak*, 22 cm) (Shao, 2005; Froese and Pauly, 2006). Despite the body sizes of the live specimens used for hand-held and captive recordings being approximately equal, frequency distribution in the pulses of purrs and dual-knocks in these two conditions differed. Relative amplitude of sound energy in an acoustic signal is determined by a number of intrinsic and extrinsic factors including size of sonic muscle(s), speed of muscle contraction, volume of the swimbladder, and environmental pressure (Connaughton *et al.*, 1997, 2000; Sprague, 2000). The sounds recorded inside the tank were subjected to distortion and differed from those recorded under a far-field condition; the acoustic properties of the tanks may have distorted the original sound (Akamatsu *et al.*, 2002). According to measurements and theoretical calculations, Akamatsu *et al.* (2002) gave a rule-of-thumb for recording reliable information of fish sound in a small tank that can minimize possible distortion, which is related to the distance of the sound source from the hydrophone and the dominant frequency of the sound and the resonant frequency of the tank. The hydrophone should be placed at a distance less than the attenuation distance of the dominant frequency and the minimum resonant frequency of the tank should differ from the dominant frequency of the fish sound. In the present study, the minimum resonant frequency of the small tank ($52 \times 32 \times 30$ cm^3) for temporary holding of the fish in the field for sound recording was about 4.6 kHz, and the attenuation distances of 1 and 2 kHz are 12.8 and 13.9 cm, respectively. The disturbance sounds recorded in the small tank followed the aforementioned guide lines; the sound characteristics would be subjected to less distortion and yield reliable acoustic information. However, the minimum resonant frequency of the large aquarium was about 2.1 kHz, which was close to the high-frequency peak, and this would distort the voluntary sounds. The tank attenuation distances of 1 and 2 kHz are 30 and 101 cm, respectively. The hydrophone in the large tank was placed at a distance beyond the attenuation distance from the sound source for most of the time. Although the parameters of the sounds recorded inside the large aquarium were subjected to distortion and differed from those recorded under a far-field condition, some parameters in the time domain (e.g., pulse repetition rate, pulse duration, and interpulse duration) still remain informative for sound identification.

The repetition rate of the pulse may also serve as a criterion for identifying species-specific sound. Connaughton *et al.* (1997, 2000, 2002a) showed the influence of fish size, sonic muscle mass, and environmental temperature on the characteristics of the disturbance sound. Characteristics of the weakfish sounds recorded in the air and in the water could be considered similar (Connaughton *et al.*, 2000; Fine *et al.*, 2004). However, sound pressure level, dominant frequency, and pulse duration would change with most of the above-mentioned factors, except the repetition rate of the pulse, which remained stable in fish that differed in length and sonic muscle mass. However, the repetition rate of the pulse in the weakfish sounds almost doubled when the surrounding water temperature increased from 12 °C to 23 °C in 28 days (Connaughton *et al.*, 2000). The experiment of photoperiod and temperature on sexual recrudescence in the weakfish showed the low temperature group (13 °C) would not mature fully; low temperature would forestall or retard gametogenesis in the spring or summer spawning species (Connaughton and Taylor, 1996). The passive acoustic field surveys by Connaughton and Taylor (1995) also showed the drumming activity dropped gradually to a complete quiescence after late June. For our field surveys for big-snout croaker sounds, they were only found from March to October; we had not recorded its sounds in the winter. Winter

should not be the spawning season of the big-snout croaker. Interestingly, the repetition rate of the pulse remained similar throughout the three seasons.

Although the big-snout croaker that we kept was alone in the large aquarium, the sounds maybe an advertisement call. The average number of calls per minute emitted by the big-snout croaker was 13.3. Connaughton and Taylor (1996) measured the drumming rates in weakfish. They injected the males with human chorionic gonadotropin (hCG) and found that the rates for the males in a group were higher than the solitary male. It will be interesting to see if vocal activity expressed in number of calls emitted per minute will intensify if more big-snout croakers are kept together in the tank.

In the aquatic environments, frequencies of arthropod sounds tend to concentrate over 2 kHz [e.g., snapping shrimp: 2–5 kHz (Au and Banks, 1998)]. On the other hand, fish sounds and nonbiological ambient noise are in a lower frequency range [below 1 kHz (e.g., Guest and Lasswell, 1978; Mok and Gilmore, 1983; Connaughton *et al.*, 1997; Ladich, 1997; Bradbury and Vehrencamp, 1998)]. The result of these distributions of acoustic energy not only creates a rather quiet frequency window between 1 and 2 kHz, but also constitutes a problem of acoustic masking that affects the efficiency of the acoustical signal when most of these signals spread in a rather narrow frequency band (i.e., below 1 kHz). It becomes adaptive if a signal is located outside this crowded band. Two species of freshwater gobiids, *Padogobius martensii* and *Gobius nigricans*, were reported utilizing their social acoustic signals with frequencies between those of two noise sources [a low-frequency source attributed to the turbulent waters and a high-frequency source attributed to the bursting of water bubbles on the surface (Lugli and Fine, 2003; Lugli *et al.*, 2003)]. Whether the acoustic signals of the big-snout croaker are adaptive depends on if the most sensitive frequency of its hearing system is tuned to the relatively high dominant frequency and this coupling remains to be tested by experiment [e.g., using the auditory brainstem response method (Kenyon *et al.*, 1998)]. This speculation deserves further attention.

ACKNOWLEDGMENTS

This work was supported by four grants from the Fisheries Agency, Council of Agriculture, Executive Yuan, R. O. C. to HKM and BQH (90AS-1.4.5-FA-F1-3, 91AS-2.5.1-FA-F1-6, 92AS-9.1.1-FA-F1-17, and 93AS-9.1.1-FA-F1-6). Special thanks are given to H. Y. Yu, T. A. Pan, and P. L. Chen for their logistic assistance in sound recording in the field and K. E. Tsai for his assistance in catching the fishes. This paper represents Contribution number 4 of Asia-Pacific Ocean Research Center at National Sun Yat-sen University.

Akamatsu, T., Okumura, T., Novarini, N., and Yan, H. Y. (**2002**). "Empirical refinements applicable to the recording of fish sounds in small tanks," J. Acoust. Soc. Am. **112**, 3073–3082.

Au, W. W. L., and Banks, K. (**1998**). "The acoustics of the snapping shrimp *Synalpheus parneomeris* in Kaneohe Bay," J. Acoust. Soc. Am. **103**, 41–47.

Bradbury, J. W., and Vehrencamp, S. L. (**1998**). *Principles of Animal Communication* (Sinauer, Sunderland, MA).

Connaughton, M. A., and Taylor, M. H. (**1995**). "Seasonal and daily cycles in sound production associated with spawning in weakfish, *Cynoscion regalis*," Environ. Biol. Fish. **42**, 233–240.

Connaughton, M. A., and Taylor, M. H. (**1996**). "Drumming, courtship, and spawning behavior in captive weakfish, *Cynoscion regalis*," Copeia **1996**, 195–199.

Connaughton, M. A., Fine, M. L., and Taylor, M. H. (**1997**). "The effects of seasonal hypertrophy and atrophy on fiber morphology, metabolic substrate concentration and sound characteristics of the weakfish sonic muscle," J. Exp. Biol. **200**, 2449–2457.

Connaughton, M. A., Fine, M. L., and Taylor, M. H. (**2002a**). "Weakfish sonic muscle: influence of size, temperature and season," J. Exp. Biol. **205**, 2183–2188.

Connaughton, M. A., Lunn, M. L., Fine, M. L., and Taylor, M. H. (**2002b**). "Characterization of sounds and their use in two sciaenid species: weakfish and Atlantic croaker," in *Listening to Fish. Passive Acoustic Applications in Marine Fisheries*. Based on the International Workshop on the Applications of Passive Acoustics to Fisheries, 8–10 April 2002 (MIT Sea Grant), pp.15–19.

Connaughton, M. A., Taylor, M. H., and Fine, M. L. (**2000**). "Effects of fish size and temperature on weakfish disturbance calls: implications for the mechanism of sound generation," J. Exp. Biol. **203**, 1503–1512.

Fine, M. L. (**1978**). "Seasonal and geographical variation of the mating call of the oyster toadfish *Opsanus tau* L," Oecologia **36**, 45–57.

Fine, M. L. (**1982**). "Possible lateralization of function in toadfish sound production," Physiol. Behav. **25**, 167–169.

Fine, M. L. (**1997**). "Endocrinology of sound production in fishes," Mar. Freshwater Behav. Physiol. **29**, 23–45.

Fine, M. L., Schrinel, J., and Cameron, T. M. (**2004**). "The effect of loading on disturbance sounds of the Atlantic croaker *Micropogonius undulatus*: air vs.water," J. Acoust. Soc. Am. **116**, 1271–1275.

Fish, J. F., and Cummings, W. C. (**1972**). "A 50-dB increase in sustained ambient noise from fish (*Cynoscion xanthulus*)," J. Acoust. Soc. Am. **52**, 1266–1270.

Froese, R., and Pauly, D. (**2006**). FishBase, World Wide Web electronic publication. http://www.fishbase.org, version (06/2006).

Guest, W. C., and Lasswell, J. L. (**1978**). "A note on courtship behavior and sound production of red drum," Copeia **1978**, 337–338.

Hawkins, A. D., and Rasmussen, K. J. (**1978**). "The calls of gadoid fish," J. Mar. Biol. Assoc. U.K. **58**, 891–911.

Kenyon, T. N., Ladich, F., and Yan, H. Y. (**1998**). "A comparative study of hearing ability in fishes: the auditory brainstem response approach," J. Comp. Physiol., A **182**, 307–318.

Ladich, F. (**1997**). "Agonistic behaviour and significance of sounds in vocalizing fish," Mar. Fresh. Behav. Physiol. **29**, 87–108.

Luczkovich, J. J., Sprague, M. W., Johnson, S. E., and Pullinger, R. C. (**1999**). "Delimiting spawning areas of weakfish, *Cynoscion regalis* (Family Sciaenidae) in Pamlico Sound, North Carolina using passive hydroacoustic surveys," Bioacoustics **10**, 143–160.

Lugli, M., and Fine, M. L. (**2003**). "Acoustic communication in two freshwater gobies: Ambient noise and short-range propagation in shallow streams," J. Acoust. Soc. Am. **114**, 512–521.

Lugli, M., Yan, H. Y., and Fine, M. L. (**2003**). "Acoustic communication in two freshwater gobies: the relationship between ambient noise, hearing and sound spectrum," J. Comp. Physiol., A **189**, 309–320.

Midling, K., Soldal, A. V., Fosseidengen, J. E., and Øvredal, J. T. (**2002**). "Calls of the Atlantic cod: does captivity restrict their vocal repertoire?" Bioacoustics **12**, 233–235.

Mok, H. K., and Gilmore, R. G. (**1983**). "Analysis of sound production in estuarine aggregations of *Pogonias cromis, Bairdiella chrysoura*, and *Cynoscion nebulosus* (Sciaenidae)," Bull. Inst. Zool. Academia Sinica **22**, 157–186.

Saucier, M. H., and Baltz, D. M. (**1993**). "Spawning site selection by spotted seatrout, *Cynoscion nebulosus*, and black drum, *Pogonias cromis*, in Louisiana," Env. Biol. Fish. **36**, 257–272.

Shao, K. T. (**2005**). Taiwan Fish Database. World Wide Web electronic publication, http://fishdb.sinica.edu.tw, version (05/2005).

Specht, R. (**2002**). *Avisoft SASLab Pro - sound analysis and synthesis laboratory software for windows* (Berlin, Germany).

Sprague, M. W. (**2000**). "The single sonic muscle twitch model for the sound-production mechanism in the weakfish, *Cynoscion regalis*," J. Acoust. Soc. Am. **108**, 2430–2437.
Takemura, A., Takita, T., and Mizue, K. (**1978**). "Studies on the underwater sound. 7. Underwater calls of the Japanese marine drum fishes (Sciaenidae)," Bull. Jpn. Soc. Sci. Fish. **44**, 121–125.
Thorson, R. F., and Fine, M. L. (**2002a**). "Acoustic competition in the gulf toadfish *Opsanus beta*: Acoustic tagging," J. Acoust. Soc. Am. **111**, 2302–2307.
Thorson, R. F., and Fine, M. L. (**2002b**). "Crepuscular changes in emission rate and parameters of the boatwhistle advertisement call of the gulf toadfish, *Opsanus beta*," Environ. Biol. Fish. **63**, 321–331.

Bio-inspired wideband sonar signals based on observations of the bottlenose dolphin (Tursiops truncatus)

Chris Capus,[a] Yan Pailhas, Keith Brown, and David M. Lane
Ocean Systems Laboratory, School of Engineering and Physical Sciences, Heriot-Watt University, Riccarton, Edinburgh EH14 4AS, United Kingdom

Patrick W. Moore
Spawar Systems Center San Diego, 53560 Hull Street, San Diego, California 92152-5001

Dorian Houser
Biomimetica, 7951 Shantung Drive, Santee, California 92071

(Received 20 April 2006; revised 25 September 2006; accepted 2 October 2006)

This paper uses advanced time-frequency signal analysis techniques to generate new models for bio-inspired sonar signals. The inspiration comes from the analysis of bottlenose dolphin clicks. These pulses are very short duration, between 50 and 80 μs, but for certain examples we can delineate a double down-chirp structure using fractional Fourier methods. The majority of clicks have energy distributed between two main frequency bands with the higher frequencies delayed in time by 5–20 μs. Signal syntheses using a multiple chirp model based on these observations are able to reproduce much of the spectral variation seen in earlier studies on natural dolphin echolocation pulses. Six synthetic signals are generated and used to drive the dolphin based sonar (DBS) developed through the Biosonar Program office at the SPAWAR Systems Center, San Diego, CA. Analyses of the detailed echo structure for these pulses ensonifying two solid copper spherical targets indicate differences in discriminatory potential between the signals. It is suggested that target discrimination could be improved through the transmission of a signal packet in which the chirp structure is varied between pulses. Evidence that dolphins may use such a strategy themselves comes from observations of variations in the transmissions of dolphins carrying out target detection and identification tasks. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2382344]

PACS number(s): 43.80.Ka, 43.60.Hj, 43.30.Gv, 43.30.Vh [WWA] Pages: 594–604

I. INTRODUCTION

There is a long history of interest and research into dolphin echolocation systems related to their excellent performance, particularly in shallow water and very shallow water environments. This has led to the development of bio-inspired systems such as the dolphin based sonar (DBS), developed through the Biosonar Program office at the SPAWAR Systems Center, San Diego (SSC-SD), which has been designed to emit and receive biomimetic dolphin-like pulses.[1] Further development of these systems will benefit from the improved understanding of the signals and strategies used by marine mammals. In this paper evidence is presented for a simple chirp model which mirrors many of the spectral variations described by the existing click taxonomy for dolphins.[2]

Digital analysis of dolphin clicks is difficult because of the very short duration of the pulses, typically 50–80 μs.[3] Nevertheless a variety of pulse types have been distinguished, related to the dolphins' behavior and task completion. One category of signals has a distinctly bimodal frequency distribution, others have a strong primary peak with significant energy in a second frequency band. Application of advanced fractional Fourier processing to a range of these "bimodal" pulses has elicited evidence for a pair of down-chirp signals strongly overlapped in both the time and frequency domains. While the frequency bands and durations are vastly different, the underlying structure bears some similarities to that seen in time-frequency analyses of echolocation pulses used by the big brown bat (Eptesicus fuscus).[4,5] These analyses shed new light on the detail structure of dolphin clicks and provide a plausible model for the design of new strains of biomimetic pulses.

Six pulses have been generated using the bio-inspired double chirp model. Designed for use with the DBS system, these pulses have been tested against a variety of targets in a controlled environment. The experiments have been conducted in a fresh water test tank measuring $12 \times 10 \times 4$ m (length $\times$ width $\times$ depth). Empirical results are presented below for echo responses from calibrated spherical targets. Detailed spectral and time-frequency echo signatures are examined using both linear and bilinear methods.

A. Click data

Signals used in this research have come from the Navy Marine Mammal Program. Two data sets have been considered. All of the clicks were collected on the main response axis of the echolocation beam. The first data set comes from experiments which involved dolphins performing an identification task in Kaneohe Bay, Hawaii.[6] The dolphins were

[a]Author to whom correspondence should be addressed. Electronic mail: c.capus@hw.ac.uk

 0001-4966/2007/121(1)/594/11/$23.00

TABLE I. Click taxonomy.

Click type	Description
A	unimodal, low frequency (<70 kHz)
B	unimodal, low frequency (<70 kHz); 2° peak (>70 kHz) between −3 and −10 dB down
C	bimodal; low and high frequency peaks within −3 dB
D	unimodal, high frequency (>70 kHz)
E	unimodal, high frequency (>70 kHz); 2° peak (<70 kHz) between −3 and −10 dB down
W	wideband; single continuous bounded region within −3 dB limit (freq. bandwidth >85 kHz)
M	multimodal; three or more distinctly bounded regions within −3 dB limit

trained to recognize differences in contents of aluminum flasks suspended in an open water testing pen at a range of approximately 4.5 m. The data set comprises natural dolphin clicks and echoes from real targets in a realistic noisy environment. Animals typically emitted 12–30 clicks before establishing a decision. Signals were digitized at a sampling rate of 500 kHz with each record being 256 samples in length.

The second set of data comes from a number of free swimming echolocation and bottom-object search experiments in San Diego Bay.[7,8] Trials involved one of two dolphin subjects performing a search for a bottom-set object. Target present and target absent trials were conducted and typically lasted between 10 and 30 s with the target station situated 20–60 m from the workboat. The dolphins were trained to give a positive whistle response on identification of the target and to swim around the target station before returning to the workboat if no target was detected. The dolphin carried an instrumentation package, the biosonar measurement tool (BMT), mounted beneath a bite plate. The BMT contains attitude and motion sensors and allows for the recording of three channels of acoustic data to give simultaneous measurement of the outgoing clicks and two echo channels. For the current paper, click trains have been analyzed for six trials involving one animal. The sampling rate for these data is 312.5 kHz.

B. Spectral classification of dolphin clicks

Observations of dolphins in controlled research environments and in the open ocean have led to the identification of spectral characteristics which can be used to define click types. One click taxonomy, which is based on dominant peaks in upper and lower frequency bands,[2] defines seven distinct click variants listed in Table I. These are in broad agreement with other observers.[3]

Figure 1 gives an example of a Type B click alongside its power spectrum. This signal comes from the Kaneohe Bay data set. The spectrum has a dominant low frequency (LF) peak (<70 kHz), with a secondary peak at high frequency (HF). This is termed unimodal low frequency. Definition as a bimodal spectrum is reserved for peak maxima within 3 dB of one another.

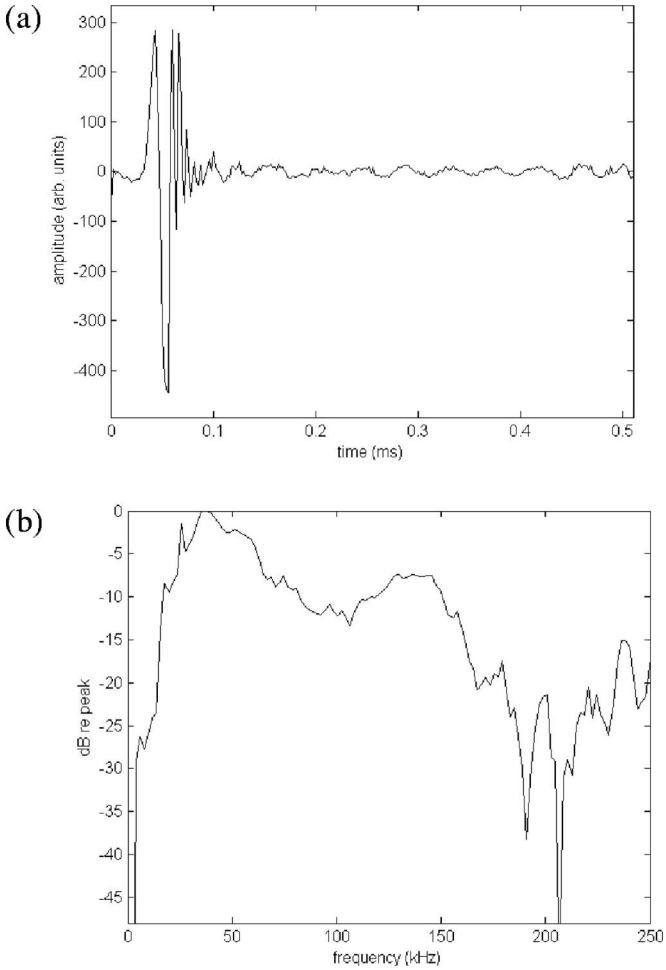


FIG. 1. Type B dolphin click: (a) Time signal; (b) power spectrum in dB re peak level.

C. Signal analysis

1. Fractional Fourier analysis

Various analyses have been performed using methods based on the fractional Fourier transform (FrFT). This is a linear transform, containing linear chirps as its basis functions, see Eq. (1)

$$F^{\alpha}f(x) = \frac{\exp\left(-j\left(\frac{1}{4}\pi\hat{\varphi} - \frac{1}{2}\varphi\right)\right)}{(2\pi|\sin\varphi|)^{1/2})} \exp\left(\frac{1}{2}jy^2\cot\varphi\right) \times \int_{-\infty}^{\infty} \exp\left(-\frac{jxy}{\sin\varphi} + \frac{1}{2}jx^2\cot\varphi\right) f(x)\mathrm{d}x, \quad (1)$$

where α: $\{0.0<\alpha<1.0\}$ defines the transform order, $\varphi = \alpha(\pi/2)$, and $\hat{\varphi}=\mathrm{sgn}\,\varphi$. There are two special cases: $F^0f(x)$ is the identity transform, returning the input signal; $F^1f(x)$ is equivalent to the ordinary Fourier transform.

The fractional Fourier transform can be used for signal separation using fractional band-pass filtering techniques.[9] For certain dolphin signals, separation of chirp components is possible in optimal fractional domains where components are overlapping in both time and frequency. Figure 2 shows

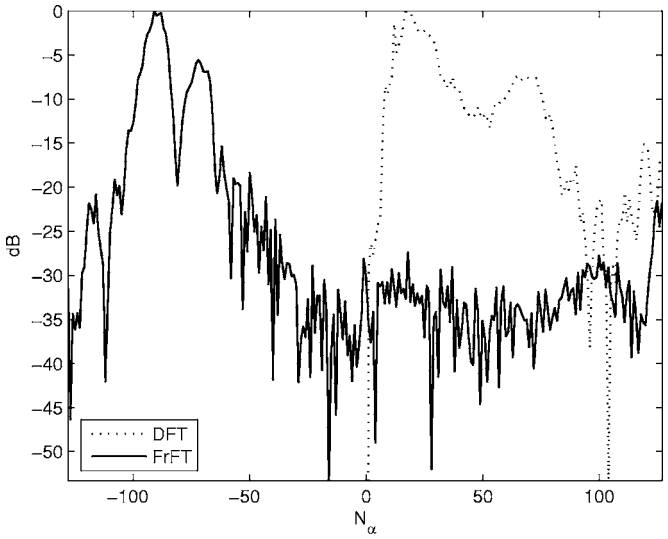


FIG. 2. Power spectra for the type B dolphin click given in Fig. 1(a): Dashed line is the frequency spectrum; solid line is the optimal fractional domain spectrum. Note that the x-axis is labeled N_α to indicate that units are different in the two domains.

the optimal fractional domain power spectrum for the type B click against the ordinary Fourier power spectrum calculated from the discrete Fourier transform (DFT). In both cases the analytic signal has been used. The x-axis is labeled N_α to indicate that the sampling spaces are different. For the DFT, the samples correspond to frequency bins. In the fractional domains sampling is in some space lying between time and frequency. Despite these differences, the two dominant peaks in the DFT and FrFT in Fig. 2 correspond to the same two major signal components for the type B click. In providing a more compact representation for the components, the optimal FrFT provides some improvement over the frequency domain for component separation and filtering.

Any analysis to find the underlying multicomponent structure in the click signals is limited by the extremely short duration of the pulses. In using the FrFT we trade-off resolution in one domain to ensure that the signals to be separated reach their minimum support in the orthogonal domain. For a linear chirp signal there is an optimal domain for representation in which the signal spread reaches its minimum. For a discrete signal, this domain α_{opt} can be found from[5,10]

$$\alpha_{\text{opt}} = -\frac{2}{\pi}\tan^{-1}\left(\frac{f_s^2/N}{2a}\right), \tag{2}$$

where a defines the linear chirp rate in Hz s^{-1}.

The FrFT of the Gaussian has Gaussian support in all fractional domains and in terms of the energy density we can write[10]

$$\sigma_\alpha^2 = \frac{\sin^2\varphi}{4\sigma_t^2} + \sigma_t^2\cos^2\varphi. \tag{3}$$

Modified uncertainty relationships have been developed to describe the uncertainty product for signal representations in two fractional domains.[11,12] For any Gaussian windowed linear chirp signal, the lower bound is achieved by the product of the signal spreads in the orthogonal domains α_{opt} and $\alpha_{\max} = \alpha_{\text{opt}} - 1$. In fact for these orthogonal domains under the FrFT, the relationship collapses to a generalized form of the conventional uncertainty product and we can write[5]

$$\sigma_{\max}\sigma_{\text{opt}} \geqslant \tfrac{1}{2}. \tag{4}$$

2. Time-frequency representations

A two-dimensional time-frequency representation (TFR) can be built using a short-time implementation of the FrFT. This STFrFT has similar properties to the short-time Fourier transform (STFT) but displays finer resolution for linear chirp signals.[5] Local optimization can be used to provide good resolution for nonlinear chirps. We compare fractional Fourier based TFRs with the Wigner-Ville distribution (WVD) and with the spectrogram. Both of these can be defined according to the general form of the Cohen class bilinear TFRs[13] (see Table II). A number of authors have investigated Radon transformation of the bilinear distributions for signal processing[14] and in particular the relationship between the FrFT and the Radon transform of the Wigner distribution.[15–17]

The interest in Radon transformation of the Wigner-Ville distribution lies in the reduction of cross-terms between signal components. Taking an alternative approach, various "smoothed" versions[18–20] have been proposed along with other distributions designed to give reduced interferences.[21] Here we also employ a reassignment method applied to the smoothed pseudo Wigner-Ville distribution, since this is expected to provide better localization in the time-frequency plane and give improved readability for TFRs generated from multicomponent inputs. The reassignment method attempts to increase the concentration of signal components in the time-frequency plane and is a generalization of the modified moving window method previously applied only to the

TABLE II. Cohen class kernel descriptions for two bilinear distributions.

Name	Kernel: $\phi(\theta,\tau)$	Distribution: $C(t,\omega)$
General Class	$\phi(\theta,\tau)$	$\frac{1}{4\pi^2}\int\int\int s*(u-\frac{1}{2}\tau)s(u+\frac{1}{2}\tau)\phi(\theta,\tau)e^{-j\theta t-j\tau\omega+j\theta u}du\,d\tau\,d\theta$
Wigner-Ville	1	$\frac{1}{2\pi}\int e^{-j\tau\omega}s*(t-\frac{1}{2}\tau)s(t+\frac{1}{2}\tau)d\tau$
Spectrogram	$\int h*(u-\frac{1}{2}\tau)h(u+\frac{1}{2}\tau)e^{-j\theta u}du$	$\lvert\frac{1}{\sqrt{2\pi}}\int e^{-j\omega\tau}s(\tau)h(\tau-t)d\tau\rvert^2$

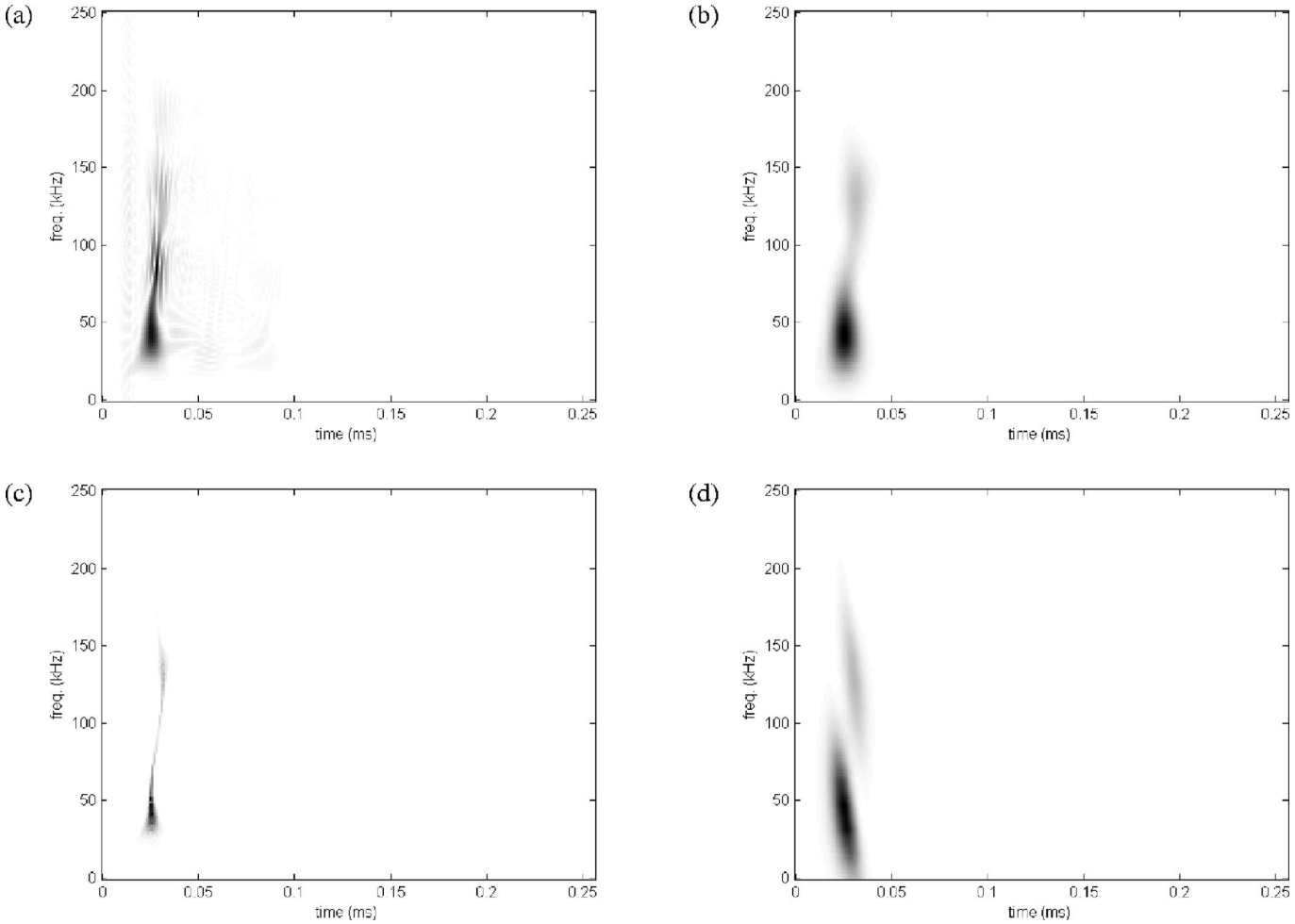


FIG. 3. TFRs for type B dolphin click: (a) Wigner-Ville distribution; (b) spectrogram; (c) reassigned smoothed pseudo Wigner-Ville distribution; (d) squared modulus of the short-time fractional Fourier transform.

spectrogram.[22] Auger and Flandrin have provided a detailed discussion and demonstrations of the properties of various TFRs and their reassigned counterparts.[23]

II. DOLPHIN CLICK ANALYSES

In this section we present outcomes of the dolphin signal analyses and develop synthetic bio-inspired signal sets suitable for experimentation with the DBS system. Figure 3 shows the selected time-frequency representations for the type B dolphin click in Fig. 1. The effect of choosing a properly matched window for a given signal has been well investigated.[24] In the simplest case for the STFT or spectrogram, this involves choosing a window of equivalent time duration to the components of interest. Our spectrograms are matched in this sense, using a 32-sample Gaussian window for the 500 kHz sampling rate, corresponding to a duration of 64 μs. The matching can be taken a stage further, by attempting to match the signal's frequency modulation. This has also been investigated for the separation of linear chirps.[25]

The WVD, Fig. 3(a), does appear to contain cross-terms indicating more than one signal component, with evidence of oscillatory behavior around 80–100 kHz. The other TFRs suggest two major components, generally with the higher frequencies delayed slightly in time. In the STFrFT, Fig. 3(d), the localization is sufficient for these to resolve into two separable chirp-like components. Over the range of signals analyzed, FrFT orders of between 0.2 and 0.4 for the 256 sample signals were most effective. The equivalent chirp rates, found by rearranging Eq. (2) to give a in terms of α, range from -0.67×10^6 to -1.5×10^6 kHz s^{-1}. This indicates downchirp behavior with approximately 54–120 kHz bandwidth over a typical click duration of 80 μs and ties in well with the measured click spectra.

Figure 3(d) also illustrates the implications of Eq. (4) on time-frequency localization under the STFrFT. The signal support is minimized in one direction in the time-frequency plane and increased in the orthogonal direction. Extending the usual notion of the uncertainty principle for time and frequency[13,26] to the fractional domains, the support of a signal cannot be arbitrarily small in two orthogonal fractional domains.

We are particularly interested in any variation in signals transmitted during a detection or identification task, since we might expect that the decision making process in the dolphin involves consideration of a sequence of echoes. We may hypothesize that the dolphin obtains additional target information by varying the transmit pulses.

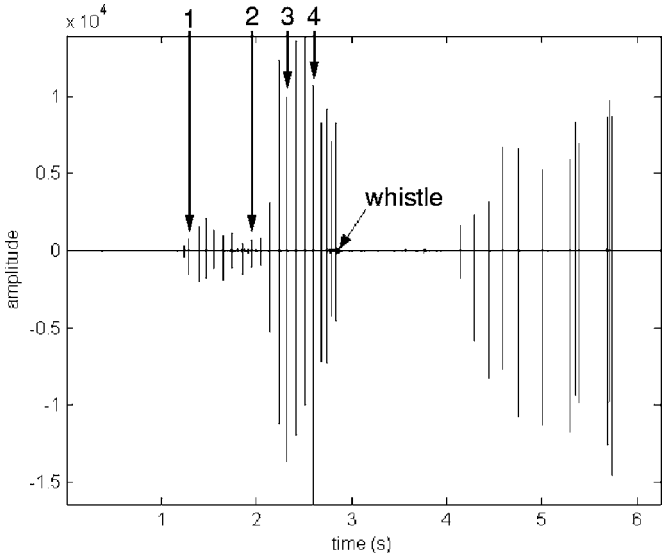


FIG. 4. Click train for a single echolocation trial. Positive (target present) whistle response is arrowed.

The two data sets both indicate systematic variations in signal structure over a trial. For one individual in the target discrimination experiments, we see patterns of cyclic variation. The lower frequency component is relatively stable, but the higher frequency component shows marked variations. At one extreme the high frequency is relatively steady, at the other extreme distinct double chirp characteristics are evident, as in Fig. 3. The higher frequencies are generally delayed in time by between 5 and 20 μs.

In the target localization experiments, for the individual identified as LUT,[8] variation is somewhat different but still involves two dominant bands and evidence of chirp and double-chirp characteristics. Figure 4 shows a typical record of the outgoing clicks produced by a dolphin during a single target detection and identification trial. A whistle, generated by the dolphin when it had detected and identified the target, is marked by an arrow. The dolphin whistle signal provides the positive identification decision point within the trial. Transmissions immediately prior to this point are typically of high power compared to an initial sequence of lower power clicks. These may indicate a shift from a general search strategy to a target acquisition strategy leading to identification. STFrFT outputs are presented in Fig. 5 for the four clicks labeled 1–4, two from each of these phases. Click 1 comes from the initial search phase and is characteristically wideband. The second click comes from the transition between search and acquire phases and has a distinct double chirp character similar to that of the type B click investigated above. Clicks 3 and 4 come from the acquire phase and demonstrate that there is a considerable variation during tar-

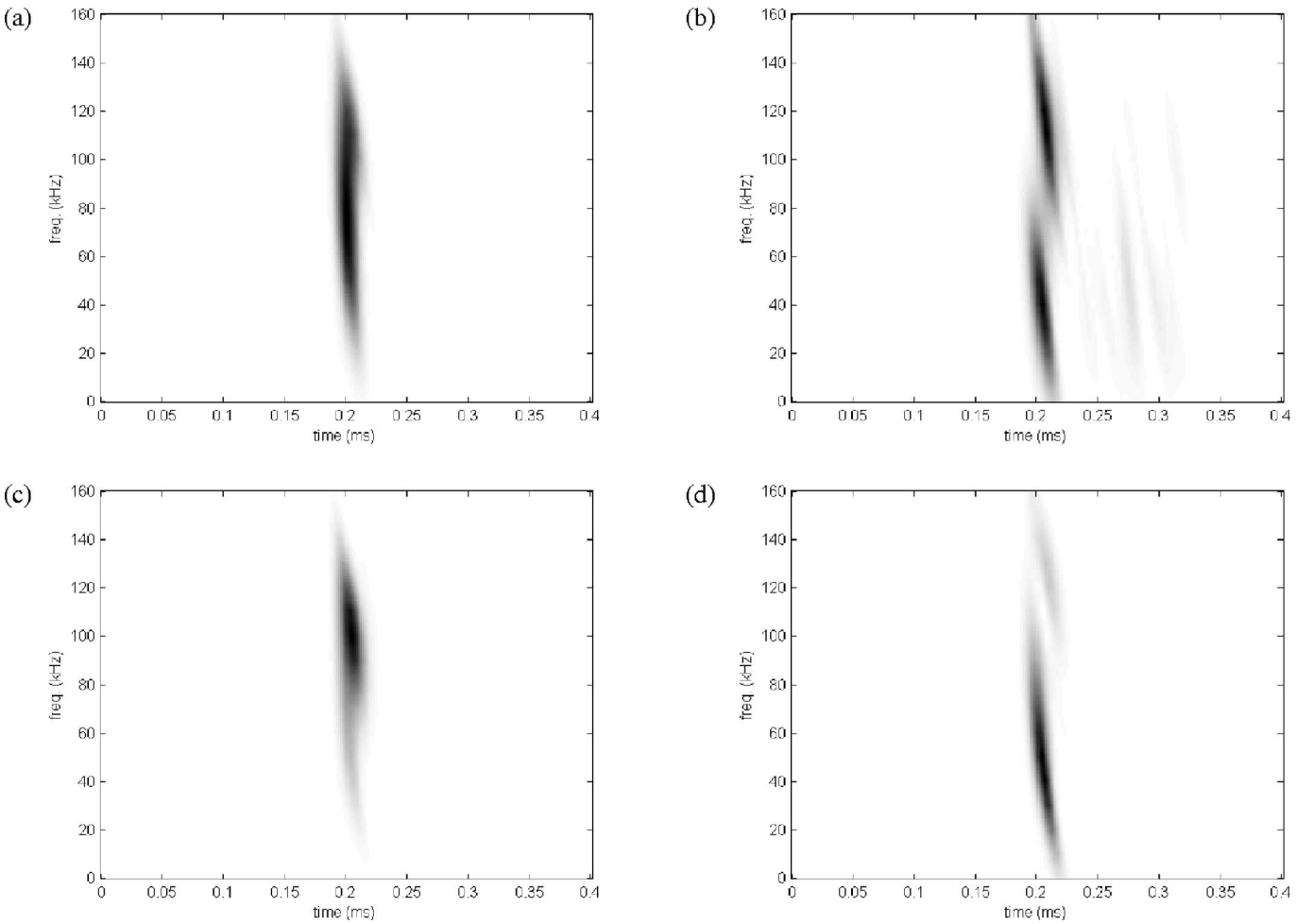


FIG. 5. STFrFTs for selected LUT clicks from Fig. 4: (a) Search phase, click 1; (b) transition, click 2; (c) acquire phase, click 3; (d) acquire phase, click 4.

TABLE III. Bio-inspired signal set.

Signal	Chirp rate, a (kHz s^{-1})	Nominal frequency bands	
		Chirp 1	Chirp 2
DC1	-0.420×10^6	30–114 kHz	46–130 kHz
DC2	-0.375×10^6	30–105 kHz	55–130 kHz
DC3	-0.330×10^6	30–96 kHz	64–130 kHz
DC4	-0.300×10^6	30–90 kHz	70–130 kHz
DC5	-0.270×10^6	30–84 kHz	76–130 kHz
DC6	-0.240×10^6	30–78 kHz	82–130 kHz

get interrogation with the dolphin using relatively high power clicks shifting between low and high frequency bands.

Following the decision point, transmitted signals typically revert to a state nearer to the general search structure. This would be consistent with a renewed search effort to relocate the work boat.

A. Bio-inspired signal models

From the observations made above, new signal models for bio-inspired synthetic pulses have been developed. The signal models are based on double downchirp structures and are designed to reflect the frequency bandwidth of bottlenose dolphin pulses. For the purposes of reducing demands on transducers and for ease of analysis and interpretation, signal durations have been increased by around 50% to 120 μs. The total duration is made up by two downchirp components having the same chirp rate but with one at a higher frequency than the other. In each case the HF component is delayed by 20 μs relative to the LF component. Each chirp is constrained by a Gaussian time window ($6\sigma=100$ μs). The increased durations give rise to somewhat lower chirp rates than those calculated from the FrFT orders for the natural dolphin signals and range from -0.24×10^6 to -0.42×10^6 kHz s^{-1} for this signal set.

The nominal frequency bands given in Table III indicate the frequencies at the start and end of the 100 μs time window for the underlying chirp in each signal component. In practice the short duration Gaussian time window applied to the chirps increases the bandwidth of the pulses beyond the nominal 30–130 kHz limits as illustrated in Fig. 6.

The construction of pulse DC6 from the Gaussian windowed underlying chirps is demonstrated in Fig. 7. The low and high frequency chirps are $\cos[2\pi(at^2+b_1t)]$ and $\cos[2\pi(at^2+b_2t)]$, respectively, with $a=-0.24\times10^6$ kHz s^{-1}, $b_1=78$ kHz and $b_2=130$ kHz. The Gaussian window is $\exp[-(t-3\sigma)^2/2\sigma^2]$, with $\sigma=100/6$ μs. For experimentation with the DBS system, the pulses are digitized at a sampling rate of 625 kHz and rescaled to utilize the 12-bit dynamic range available in the transmit buffers.

We have two immediate aims. The first is to investigate the capacity of these synthetic signals to reproduce the spectral variations characteristic of natural dolphin signals. The second aim is to investigate differential target responses to the synthetic signals to indicate their suitability for target discrimination.

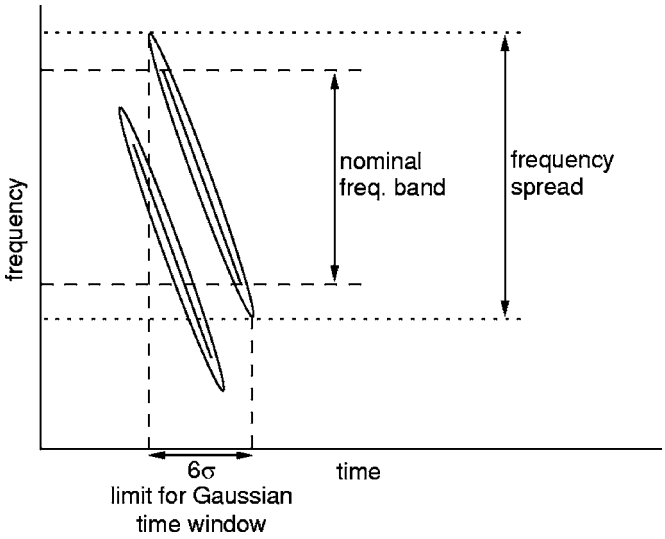


FIG. 6. Chirp components for the bio-inspired double chirp signals are designed with a fixed Gaussian window in the time domain. The relatively narrow time window increases the frequency spread beyond the nominal band limits given.

B. Spectral properties for bio-inspired signals

Figure 8 shows three of the bio-inspired synthetic signals along with their power spectra. The spectra have been produced directly from the synthesized pulses. Since transducer sensitivities have not been taken into account, it is recognized that this represents an idealized situation. Compensation would be required to achieve equal power between components through transmission into the water or at the receiver. This does highlight that peak intensities are sensitive to transducer characteristics and that these must be taken into account in the categorization of transmitted and received signals.

From the spectra for signals DC1, DC3, and DC6, these could be classified under the dolphin click taxonomy as types E, M, and C, respectively. In fact using the simple two-chirp model, in which only the rate of the constituent chirps is modified, it is possible to reproduce much of the characteristic spectral variation seen in natural dolphin signals.

Clearly, more pronounced spectral variations could be produced by allowing the relative power of the components to be altered. It is known that dolphins are able to alter the output power of their echolocation clicks and seem to do so primarily by varying the high frequency levels.[3] While environmental factors play a part in determining the signal content used, it has been noted that to produce the highest power signals both the beluga whale and bottlenose dolphin require increased output at high frequencies.[27–29]

C. Time-frequency structure

The synthesized signals have a known multicomponent structure which helps with interpretation of the various TFRs used for their representation in the time-frequency plane. In Fig. 9 four different TFRs are presented for DC1.

For components with so little separation the difficulties arising from cross-terms are clear in the WVD. The spectrogram is unable to resolve the chirp characteristics, instead

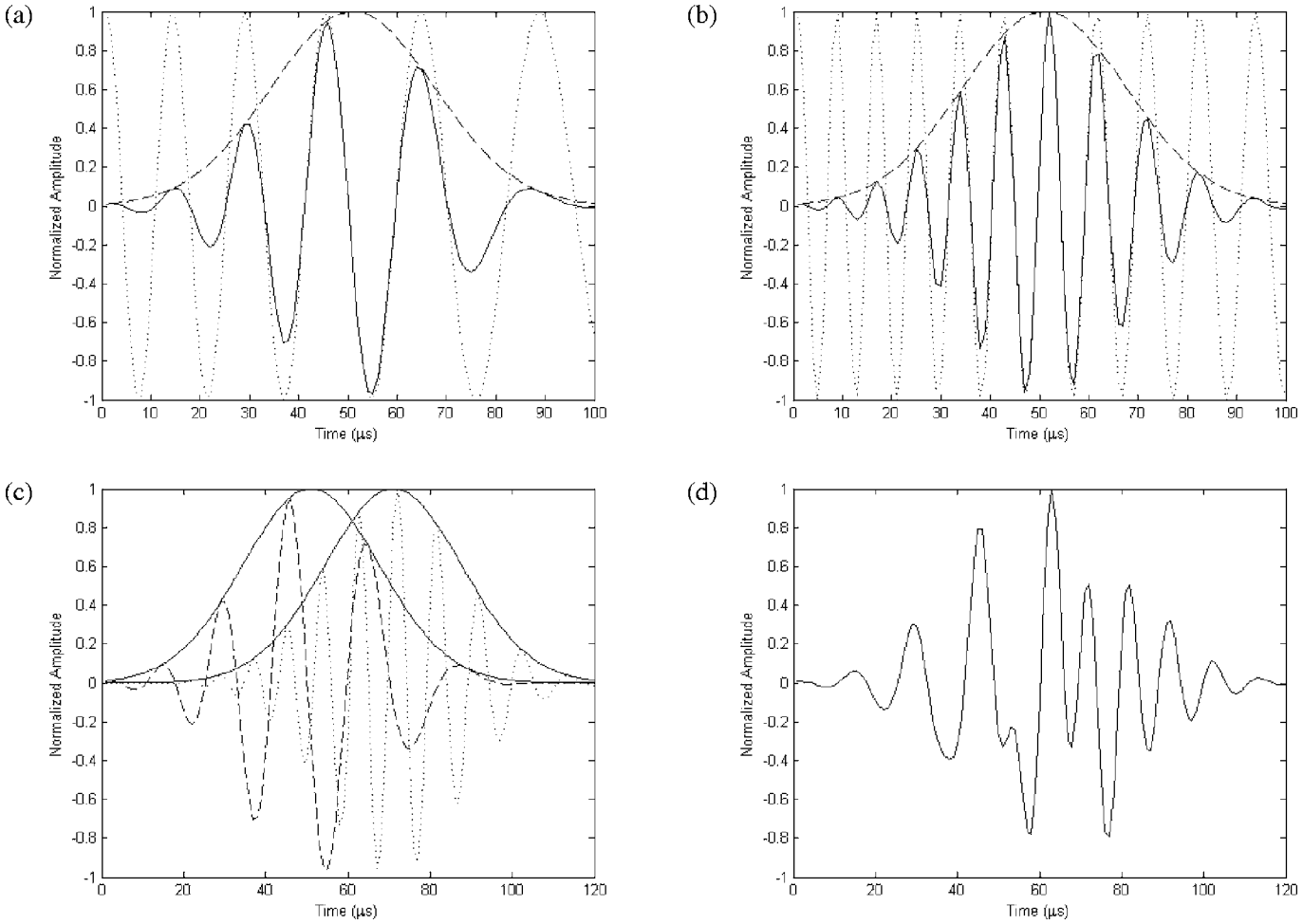


FIG. 7. Construction of signal DC6: (a) Low frequency component showing underlying linear chirp and Gaussian window; (b) high frequency component with underlying linear chirp and Gaussian window; (c) overlap with 20 μs offset; (d) complete pulse.

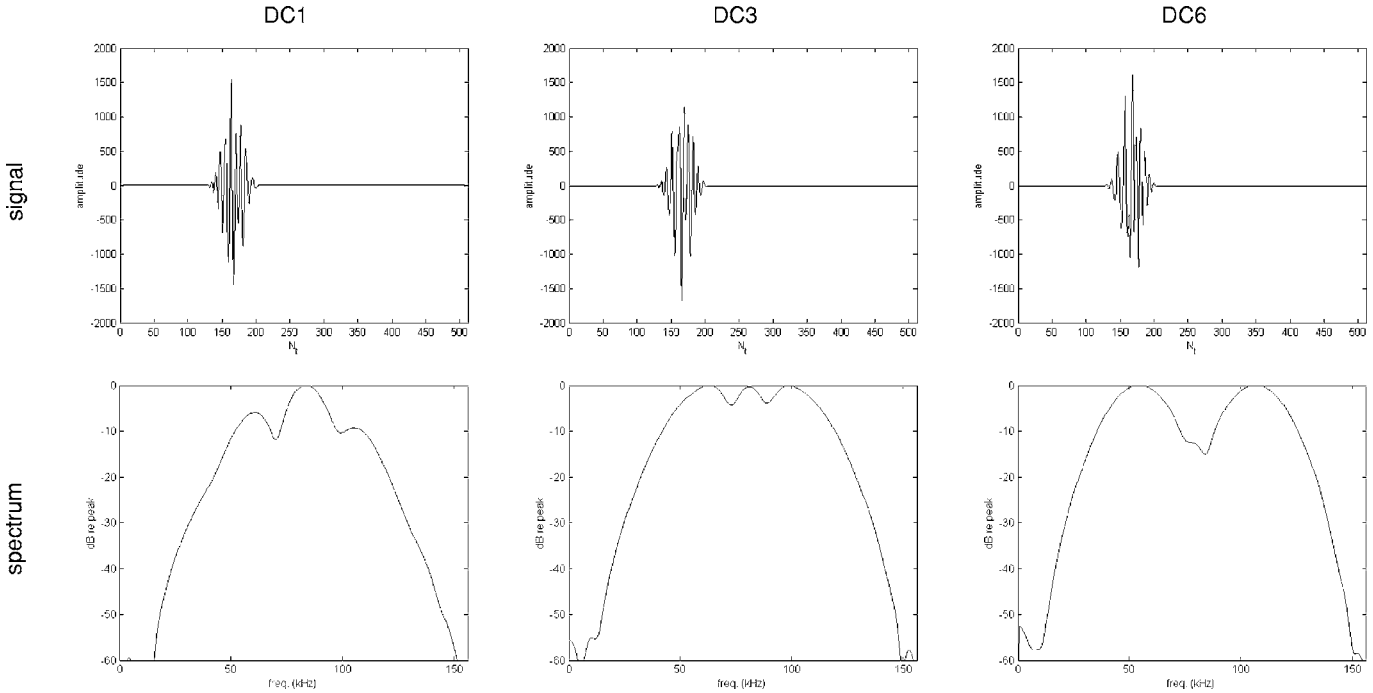


FIG. 8. Time signals and associated power spectra for bio-inspired pulses DC1, DC3, and DC6.

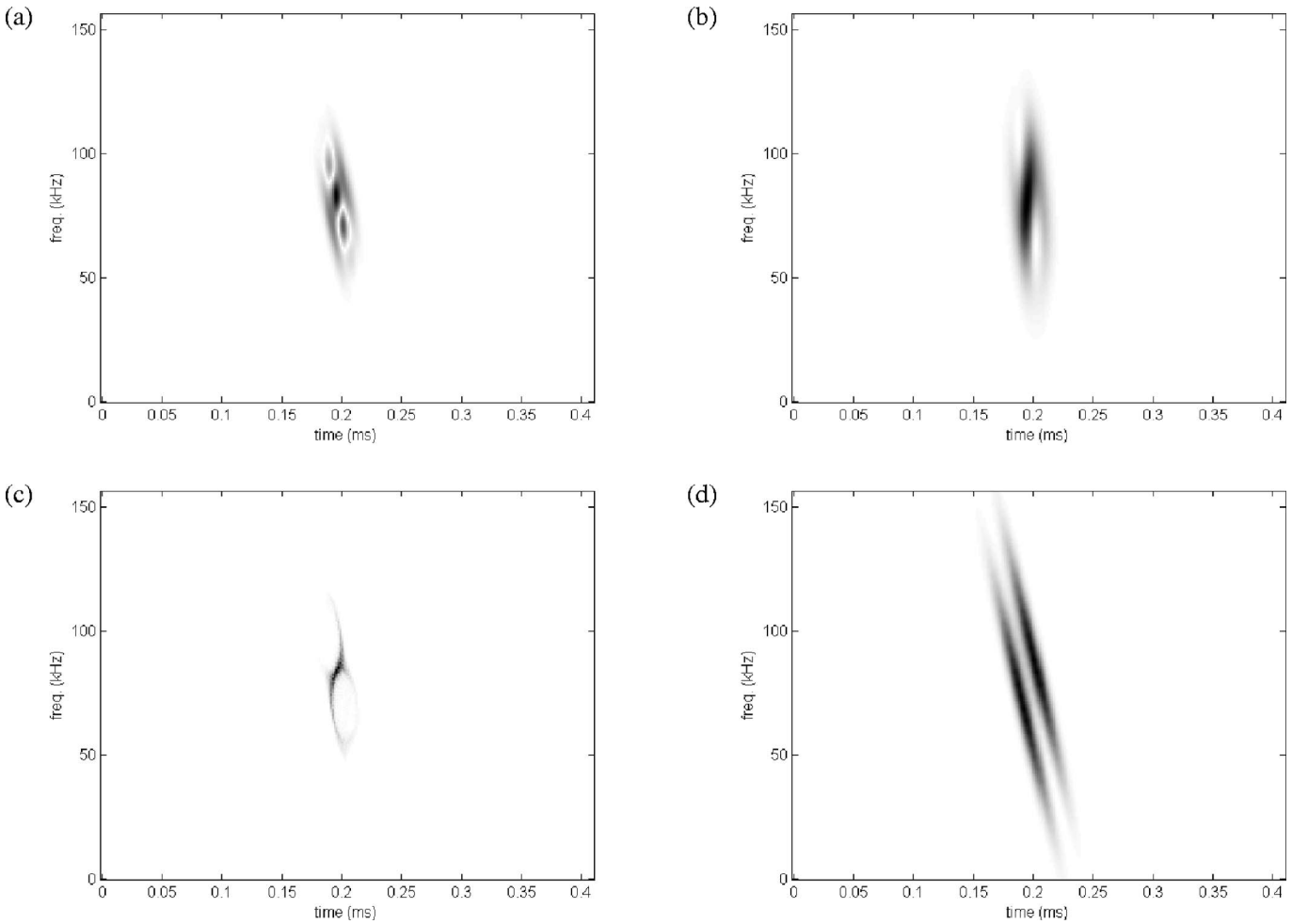


FIG. 9. TFRs for bio-inspired synthetic signal DC1: (a) Wigner-Ville distribution; (b) spectrogram; (c) reassigned smoothed pseudo Wigner-Ville distribution; (d) squared modulus of the short-time fractional Fourier transform.

giving undue weight to frequency correspondences between the components. The reassigned smoothed pseudo Wigner-Ville distribution (RSPWVD) fares no better than the WVD, with the cross terms still providing the strongest elements in the time-frequency plane. This indicates a limit on the fixed smoothing ability of the pseudo Wigner-Ville methods which are not well suited to diagonal multicomponent signals.[26] The short-time fractional Fourier method, optimized for the known chirp rate of the components of DC1, produces a TFR which provides much clearer separation between components, confirming that it does provide a good approach to analyses of these types of signals.

Figure 10 presents the same TFRs calculated for input signal DC6. The WVD is still strongly affected by cross terms, but the separation is now sufficient for the component chirps to be well represented in the RSPWVD. The spectrogram separates the components well, but does not provide a good representation of their chirp nature. The STFrFT provides the most accurate representation of the spread of the components in the time-frequency plane.

III. TARGET ECHO ANALYSES

The signal set derived in the previous section has been used with the DBS system to investigate echo responses from two spherical test targets. The targets are of solid copper construction with diameters of 60 and 23 mm and are designed for sonar calibration at 38 and 120 kHz, respectively.

In considering echo responses we concentrate on the WVD and STFrFT. For classification of targets the WVD may prove valuable because all of the structural information in the echo signal is contained within the main lobes and cross terms. Algorithms which involve smoothing mask this information and when component separations are very small, as they are in Figs. 3 and 9, understanding of the WVD's cross terms still allows an interpretation which is not readily available from the smoothed distributions. We have also seen that the RSPWVD is unable to provide any improvement in interpretation for narrowly spaced chirps, such as those in DC1. The spectrogram and STFrFT outputs have many similarities but the STFrFT provides a better interpretation for multichirp signals.

The echo responses do provide some discriminatory potential for these targets. However, certain pulse combinations are especially useful. All of the TFRs for the 120 kHz calibration sphere, Fig. 11, show a dominant HF response. This is particularly marked for pulses DC3 and DC6. The pulses with wider spacings contain more energy around 120 kHz,

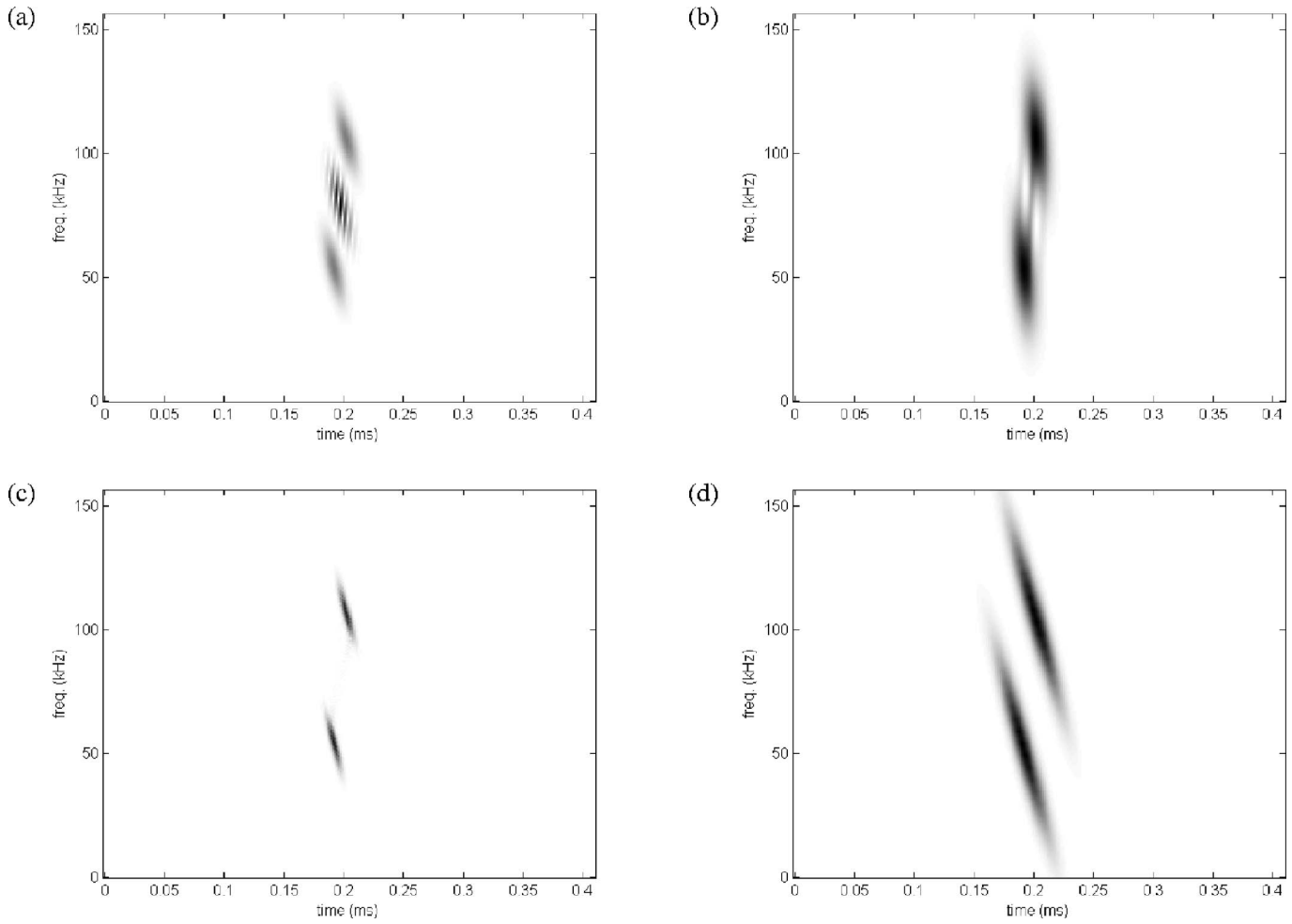


FIG. 10. TFRs for bio-inspired synthetic signal DC6: (a) Wigner-Ville distribution; (b) spectrogram; (c) reassigned smoothed pseudo Wigner-Ville distribution; (d) squared modulus of the short-time fractional Fourier transform.

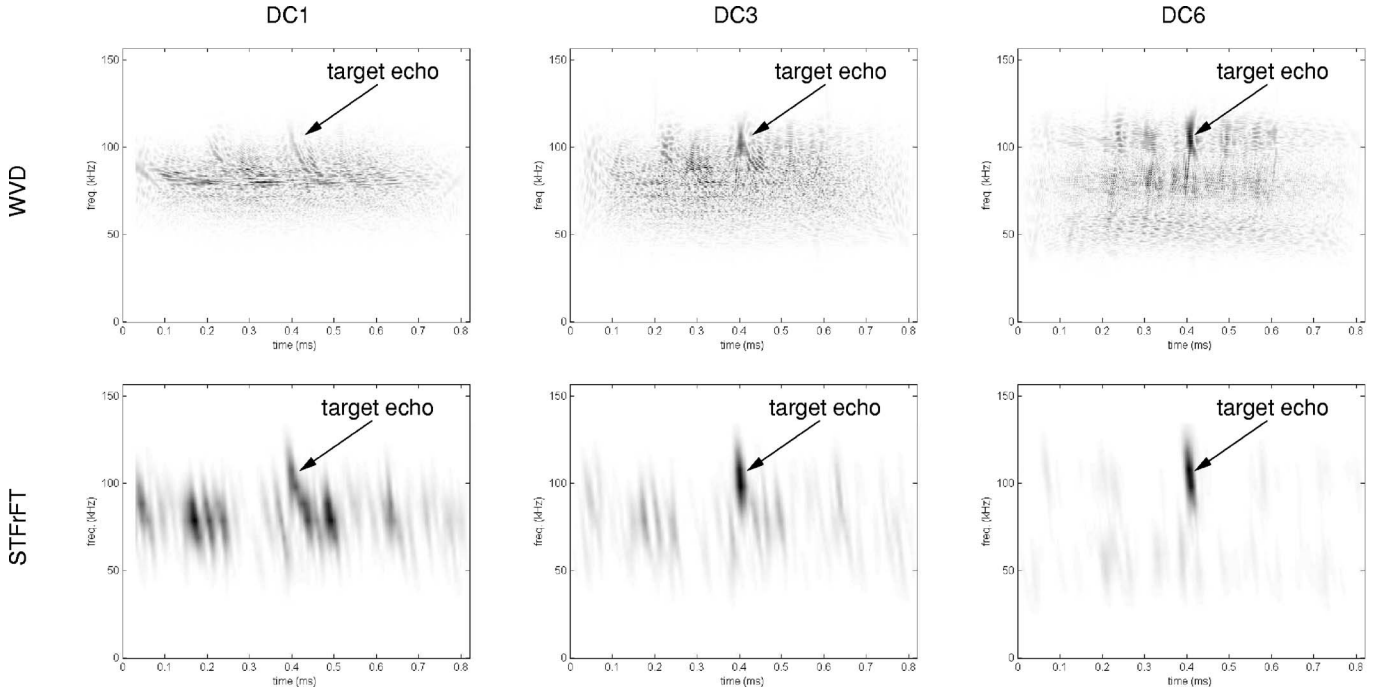


FIG. 11. TFRs for 120 kHz calibration sphere echo using signals DC1, DC3, and DC6.

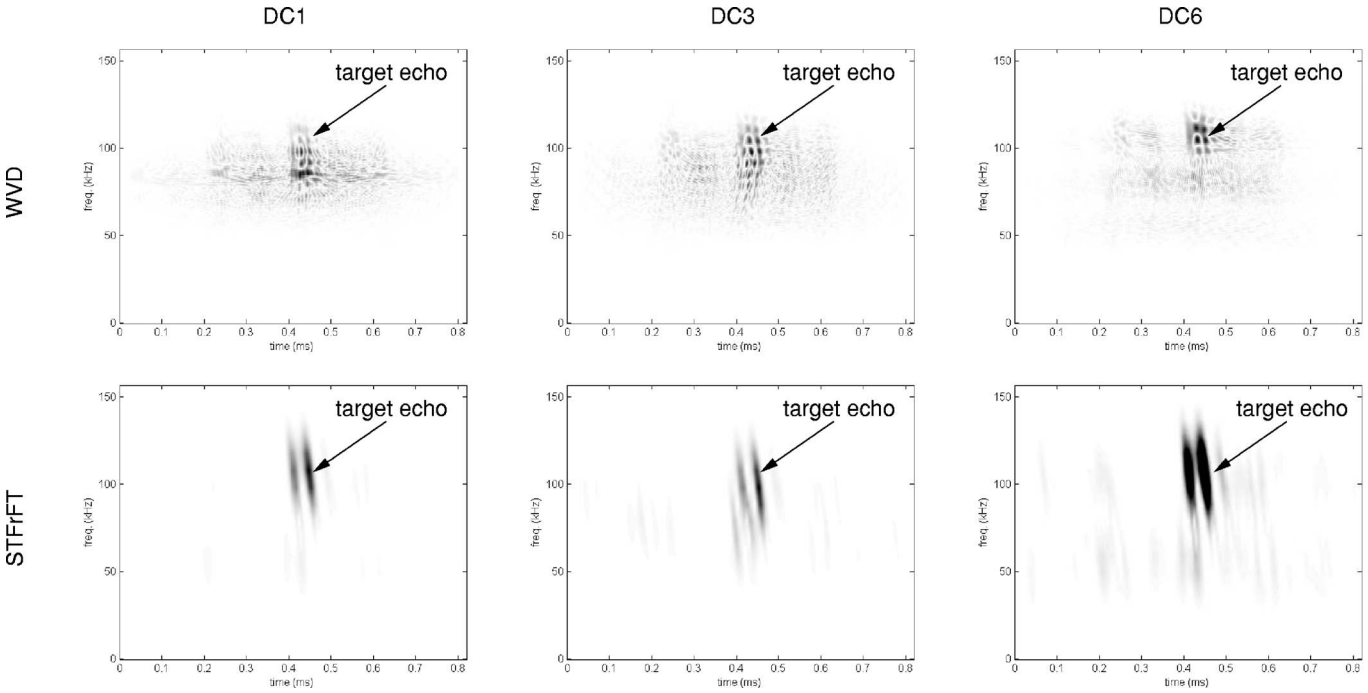


FIG. 12. TFRs for 38 kHz calibration sphere echo using signals DC1, DC3, and DC6.

corresponding to a relatively flat modal response in this spectral region for the 23 mm sphere. Despite its low strength the sphere echo, marked by an arrow, is clearly distinguished from other reverberation returns by its HF response and pulses DC3 and DC6 are seen to be more effective than DC1 in highlighting the target.

The responses for the larger sphere, Fig. 12, are also higher in the HF band, but are less concentrated around 120 kHz. Indeed pulse DC3 gives the greatest differential response for this target over the reverberation returns. A characteristic double echo is seen as the sphere is large enough for a clear separation between the first and subsequent returns. Even for a simple spherical target, the echo response is determined by many factors. Nevertheless, accurate theoretical responses can be generated using a modal decomposition,[30] and from these an estimate of the expected timings can be derived.[31]

For the 60 mm diameter solid copper sphere the expected separation between the first and second target echoes is 74.1 μs, which ties in very well with the spacings seen in Fig. 12. For the smaller sphere, the theoretical spacing is reduced to 28.2 μs. In fact it is unlikely that a strong modal response would be generated for this target at the frequencies used by the DBS. In any event, such a narrow time spacing is unlikely to be resolved given the spread of the pulses in the time-frequency plane.

IV. SUMMARY AND CONCLUSIONS

Dolphin click time-frequency analyses have revealed a common pattern of two main concentrations of energy in LF and HF bands, with the HF component slightly delayed in time. A two-chirp bio-inspired signal model has been proposed based on the fractional Fourier analysis of real dolphin clicks having a large degree of separation between spectral peaks in the frequency domain. Using the FrFT allows for the separation and independent analysis of these components and indicates a general downward frequency sweep. It is noted that such a downchirp structure is also seen in some bat echolocation signals. While components cannot be separated for many dolphin pulses even in the fractional domain space, small changes to chirp rate for the modeled components have been shown to produce significant spectral variation giving rise to spectral characteristics similar to those seen in the range of natural dolphin signals analyzed.

Six signals have been generated using the double down chirp model and these have been used to drive the DBS biomimetic system. Analyses of echoes from calibrated spherical targets have been presented and it has been shown that discriminatory potential exists in the use of these signals. The pulse variations have not been designed to highlight specific target properties but have nonetheless picked out target resonances. Significant differences in relative target and reverberation responses have been noted between pulses with only small differences in chirp characteristics.

ACKNOWLEDGMENTS

Thanks to Southampton Oceanographic Center for the loan of the calibration targets. This research has been supported by the Office of Naval Research under the NICOP program, Grant No. N00014-05-1-0725.

[1]D. Houser, S. Martin, M. Phillips, E. Bauer, T. Herrin, and P. Moore, "Signal processing applied to the dolphin-based sonar," in *Proc. MTS/IEEE OCEANS 2003* (San Diego, CA, 2003), pp. 297–303.
[2]D. Houser, D. Helweg, and P. Moore, "Classification of dolphin echolocation clicks by energy and frequency distributions," J. Acoust. Soc. Am. **106**, 1579–1585 (1999).
[3]W. Au, *The Sonar of Dolphins* (Springer-Verlag, New York, 1993).
[4]A. Bultan, "A four-parameter atomic decomposition of chirplets," IEEE Trans. Signal Process. **47**, 731–745 (1999).

[5]C. Capus and K. Brown, "Short-time fractional Fourier methods for the time-frequency representation of chirp signals," J. Acoust. Soc. Am. **113**, 3253–3263 (2003).
[6]P. Moore, H. Roitblat, R. Penner, and P. Nachtigall, "Recognizing successive dolphin echoes with an integrator gateway network," Neural Networks **4**, 701–709 (1991).
[7]S. Martin, M. Phillips, E. Bauer, P. Moore, and D. Houser, "Application of the biosonar measurement tool (BMT) and instrumented mine simulators IMS to exploration of dolphin echolocation during free-swimming, bottom-object searches," in *Proc. MTS/IEEE OCEANS 2003* (San Diego, CA, 2003), pp. 311–315.
[8]D. Houser, S. Martin, E. Bauer, M. Phillips, T. Herrin, M. Cross, A. Vidal, and P. Moore, "Echolocation characteristics of free-swimming bottlenose dolphins during object detection and identification," J. Acoust. Soc. Am. **117**, 2308–2317 (2005).
[9]H. Ozaktas, Z. Zalevsky, and M. Kutay, *The Fractional Fourier Transform: With Applications in Optics and Signal Processing* (Wiley, Chichester, U.K., 2001).
[10]C. Capus and K. Brown, "Fractional Fourier transform of the Gaussian and fractional domain signal support," IEE Proc. Vision Image Signal Process. **150**, 99–106 (2003).
[11]H. Ozaktas and O. Aytür, "Fractional Fourier domains," Signal Process. **46**, 119–124 (1995).
[12]S. Shinde and V. Gadre, "An uncertainty principle for real signals in the fractional Fourier transform domain," IEEE Trans. Signal Process. **49**, 2545–2548 (2001).
[13]L. Cohen, *Time-Frequency Analysis* (Prentice-Hall, New Jersey, 1995).
[14]J. Wood and D. Barry, "Radon transformation of time-frequency distributions for analysis of multicomoponent signals," IEEE Trans. Signal Process. **42**, 3166–3177 (1994).
[15]A. Lohmann, "Image rotation, Wigner rotation and the fractional Fourier transform," J. Opt. Soc. Am. A **10**, 2181–2186 (1993).
[16]A. Lohmann, "Relationships between the Radon-Wigner and fractional Fourier transforms," J. Opt. Soc. Am. A **11**, 1798–1801 (1994).
[17]D. Mustard, "The fractional Fourier transform and the Wigner distribution," J. Aust. Math. Soc. Ser. B, Appl. Math. **38**, 209–219 (1996).
[18]T. Claasen and W. Meclenbräuker, "The Wigner distribution—a tool for time-frequency signal analysis—Part I: continuous-time signals," Philips J. Res. **35**, 217–250 (1980).
[19]T. Claasen and W. Meclenbräuker, "The Wigner distribution—a tool for time-frequency signal analysis—Part II: discrete-time signals," Philips J. Res. **35**, 276–300 (1980).
[20]L. Stanković, "A method for time-frequency analysis," IEEE Trans. Signal Process. **42**, 225–229 (1994).
[21]J. Jeong and W. Williams, "Kernel design for reduced interference distributions," IEEE Trans. Signal Process. **40**, 402–412 (1992).
[22]K. Kodera, R. Gendrin, and C. de Villedary, "Analysis of time-varying signals with small BT values," IEEE Trans. Acoust., Speech, Signal Process. **26**, 64–76 (1978).
[23]F. Auger and P. Flandrin, "Improving the readability of time-frequency and time-scale representations by the reassignment method," IEEE Trans. Signal Process. **43**, 1068–1089 (1995).
[24]D. Jones and T. Parks, "A resolution comparison of several time-frequency representations," IEEE Trans. Signal Process. **40**, 413–420 (1992).
[25]L. Linnett, S. Morrison, and P. Nicholson, "The analysis of signals containing mixtures of linear chirps," Proc. Inst. Acoust. **23**, 55–62 (2001).
[26]P. Flandrin, *Time-Frequency Time-Scale Analysis* (Academic Press, San Diego, CA, 1999).
[27]W. Au, D. Carder, R. Penner, and B. Scronce, "Demonstration of adaptation in beluga whale echolocation signals," J. Acoust. Soc. Am. **77**, 726–730 (1985).
[28]P. Moore and D. Pawloski, "Investigations on the control of echolocation pulses in the dolphin (*Tursiops truncatus*)," in *Sensory Abilities of Cetaceans: Laboratory and Field Evidence*, edited by J. Thomas and R. Kastelein (Plenum Press, New York, 1990), pp. 305–316.
[29]C. Turl, D. Skaar, and W. Au, "The echolocation ability of the beluga (*Delphinapterus leucas*) to detect targets in clutter," J. Acoust. Soc. Am. **89**, 896–901 (1991).
[30]J. Faran, "Sound scattering by solid cylinders and spheres," J. Acoust. Soc. Am. **23**, 405–418 (1951).
[31]R. Aubauer, W. Au, P. Nachtigall, D. Pawloski, and C. DeLong, "Classification of electronically generated phantom targets by an Atlantic bottlenose dolphin," J. Acoust. Soc. Am. **107**, 2750–2754 (2000).

Echo features used by human listeners to discriminate among objects that vary in material or wall thickness: Implications for echolocating dolphins

Caroline M. DeLong[a)]
Department of Neuroscience, Brown University, Box 1953, Providence, Rhode Island 02912

Whitlow W. L. Au
Hawaii Institute of Marine Biology, P.O. Box 1106, Kailua, Hawaii 96734

Sarah A. Stamper
New College of Florida, 5700 North Tamiami Trail, Sarasota, Florida 34243

(Received 18 April 2006; revised 27 October 2006; accepted 1 November 2006)

Echolocating dolphins extract object feature information from the acoustic parameters of echoes. To gain insight into which acoustic parameters are important for object discrimination, human listeners were presented with echoes from objects used in two discrimination tasks performed by dolphins: Hollow cylinders with varying wall thicknesses (±0.2, 0.3, 0.4, and 0.8 mm), and spheres made of different materials (steel, aluminum, brass, nylon, and glass). The human listeners performed as well or better than the dolphins at the task of discriminating between the standard object and the comparison objects on both the cylinders (humans=97.1%; dolphin=82.3%) and the spheres (humans=86.6%; dolphin=88.7%). The human listeners reported using primarily pitch and duration to discriminate among the cylinders, and pitch and timbre to discriminate among the spheres. Dolphins may use some of the same echo features as the humans to discriminate among objects varying in material or structure. Human listening studies can be used to quickly identify salient combinations of echo features that permit object discrimination, which can then be used to generate hypotheses that can be tested using dolphins as subjects.
[DOI: 10.1121/1.2400848]

PACS number(s): 43.80.Ka, 43.80.Lb [FD] Pages: 605–617

I. INTRODUCTION

Echolocating dolphins emit broadband sonar sounds and perceive objects from the echoes of these sounds. They can detect and discriminate among objects using echolocation (for a review see Au, 2000). However, it is still not clear which echo acoustic features convey object properties such as size, shape, and material to an echolocating dolphin, i.e., which features are important to dolphins during object discrimination. One approach to identifying the acoustic features in echoes that dolphins may use to discriminate among objects is to present a dolphin with an echoic discrimination task, then measure the object echoes and analyze acoustic differences among the objects in conjunction with the dolphin's error patterns (see DeLong *et al.*, 2006b).

Another approach is to ask human listeners to discriminate among echoes and determine the relevant echo acoustic features. Research to date indicates that the inner ear of dolphins appears to function similarly to the human inner ear (or any other mammalian ear) except for the dolphins' ability to hear much higher frequencies (Johnson, 1967). For example, both humans and dolphins can discriminate between sounds that differ in intensity by 1 dB (Evans, 1973; Green, 1993) and the frequency discrimination abilities of humans and dolphins for tonal stimuli are comparable in the range of best hearing for each species (Herman and Arbeit, 1972; Thompson and Herman, 1975; Wier *et al.*, 1977). There are some differences between humans and dolphins that could result in differences in the perception of echo stimuli (e.g., sharp frequency tuning and short auditory integration time in dolphins; see Supin and Popov, 1995). Even so, some insight into dolphins' use of echo features may be achieved using human listeners.

The major advantage of using human listeners is that, unlike dolphins, they can report which acoustic features allowed them to discriminate among objects. Other advantages are that human listening studies are inexpensive, can be performed in a few weeks, and there are abundant research subjects. In contrast, dolphin experiments can take months to years due to extensive training time, and there are few animals at present who are available and can undertake this work. The expense of performing experiments with dolphins is very high, so the choice of experiments is scrutinized. Human listeners can quickly identify salient combinations of echo features that permit object discrimination, which can then be used to generate hypotheses that can be tested using dolphins as subjects.

Both blind and blind-folded human listeners with normal vision can judge the distance, shape, size, and surface texture of distal objects using echolocation (Kellogg, 1962; Rice, 1967). The humans in these studies investigated objects by projecting self-generated broadband signals such as tongue-

[a)]Electronic mail: caroline_delong@brown.edu

clicks, snaps, or hisses and listening to the returning echoes. When using self-generated signals, dolphins are better able to discriminate smaller differences among objects than humans. When presented with echoes that were generated using broadband dolphin signals instead of using their own signals, humans are typically capable of discriminating among objects as well or better than dolphins (Au and Martin, 1989; DeLong *et al.*, 2006a; Fish *et al.*, 1976; Helweg *et al.*, 1995). For example, DeLong *et al.* (2006a) presented human listeners with echo trains from objects used in a three-alternative match-to-sample task performed by a dolphin. The six object sets varied in size, shape, material, and surface texture. In two experiments, the human listeners ($M=84\%$) performed as well or better than the dolphin ($M=55\%$) on five of the six sets (chance performance=33%). The human listeners reported using echo features such as overall loudness, pitch, and timbre, and the pattern of changes in loudness and pitch across the echo train.

To determine whether the echo features reported by the humans were likely to have been used by the dolphin, DeLong *et al.* (2006a) compared the error patterns of the humans and the dolphin. Matching error patterns implied use of similar features whereas mismatching patterns implied use of different features. The results suggested that the dolphin attended to the pattern of changes in acoustic features as an object is scanned across a range of orientations to discriminate among the objects that varied in shape, and that it did not rely on overall amplitude differences to discriminate among the objects that varied in size. It was unclear whether the dolphin used the same cues as the humans for the objects that varied in material and texture.

In the current study, human listeners were presented with echoes from aspect-independent objects that were used in two different dolphin experiments: Hollow cylinders that varied in wall thickness (Au and Pawloski, 1992), and spheres made of different materials (Aubauer *et al.*, 2000). The hollow cylinders were presented to the human listeners in an effort to clarify the echoic cues that were used by the dolphin. The dolphin could have used time domain cues (time differences between two echo highlights), frequency domain cues (changes in frequency of prominent spectral features such as notches), or a cue derived from time domain cues called time-separation pitch (TSP). In humans, TSP is a perceived pitch of $1/T$ Hz that results from two highly correlated broadband pulses separated by time T (Thurlow and Small, 1955). The spheres were presented to the human listeners to better establish the kinds of cues that are available in objects that vary only in material. The objects used in the study by DeLong *et al.* (2006a) were intended to vary in material, but also had some variations in shape and size, so it was unclear whether the reported cues were associated with the material, size, or shape of the objects. The spheres used in the current study were machined to be exactly the same size and shape, thus ensuring that any differences in the echoes would be due to the material.

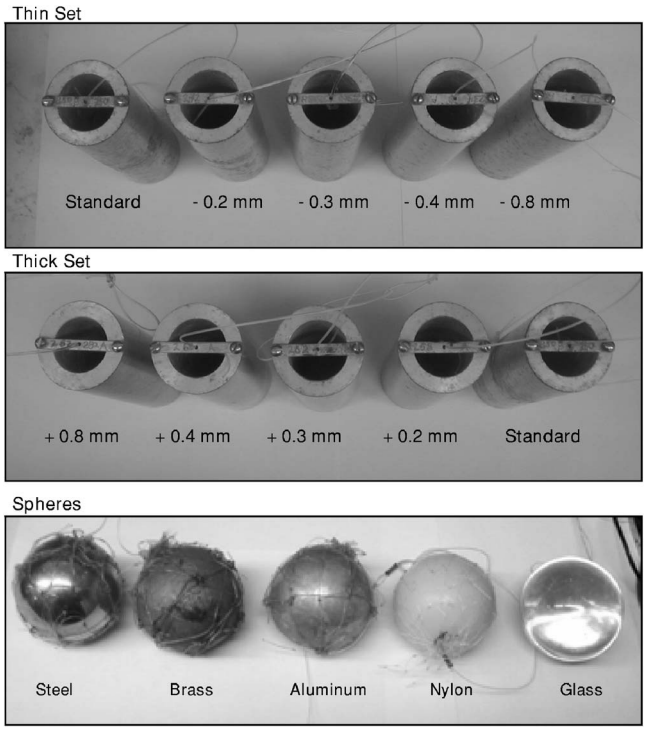


FIG. 1. Object sets. Top two panels show the thin and thick cylinder sets. The hollow aluminum cylinders were all 12.7 cm in length and 37.85 mm in outer diameter but differed in wall thickness. The standard cylinder is 6.35 mm and the comparison cylinders differ by ±0.2, 0.3, 0.4, and 0.8 mm. The bottom panel shows the sphere set (all are 7.62 cm in diameter).

II. METHOD

A. Participants

Sixteen participants (eight males and eight females) were recruited to participate in the study. Participants ranged in age from 18 to 33 years ($M=23.9$). All participants were students at Brown University or residents of Providence, Rhode Island. Before their participation, all participants were screened for normal hearing using a standard hearing test (Digital Recordings, 2002) administered via headphones on a laptop computer. At the time of testing no participants reported any hearing difficulties and all were able to detect the echo stimuli during the experiment. Participants who completed the study were paid $10.

B. Materials

1. Objects

Stimuli for this experiment consisted of pre-recorded echoes from two groups of objects: Cylinders and spheres (Fig. 1). The cylinders were previously used in a study with a bottlenose dolphin (Au and Pawloski, 1992). The hollow aluminum cylinders were all 12.7 cm in length and 37.85 mm in outer diameter but differed in wall-thickness. The cylinders had open ends, so they were filled with water when submerged. Two sets of cylinders (thin and thick), each consisting of a standard and four comparison cylinders, were used. The standard cylinder had a wall thickness of 6.35 mm. The comparison cylinders had wall thicknesses that differed from the standard by ±0.2, ±0.3, ±0.4, and ±0.8 mm. All of the cylinders were machined to a tolerance of 0.025 mm. The

thin set contained the standard cylinder and comparison cylinders that were thinner than the standard. The thick set contained the standard cylinder and comparison cylinders that were thicker than the standard. The spheres were made of stainless steel, aluminum, brass, glass, and nylon (all solid and 7.62 cm in diameter) and previously used in a study with a bottlenose dolphin (Aubauer *et al.*, 2000). The standard object was the steel sphere, and the other four spheres were the comparison objects. Participants viewed 20 cm × 28 cm photographs of each of the two cylinder sets and the spheres during the study.

2. Echo measurements and formatting

The object echoes were recorded in a cylindrical tank (1.31 m in height by 2.41 m in diameter, and contained 8.3 m^3 of seawater). A custom-made transducer consisting of a 6.35 cm diameter circular, 6.35 mm thick, 1–3 composite piezoelectric element was used to transmit the signals and receive the echoes. The transducer and targets were placed 1 m below the water surface separated by approximately 2 m. Individual targets were hung with monofilament line. The transmitted signal was a broadband dolphin echolocation signal that was 70 μs long with a peak frequency of about 120 kHz that has been used in numerous studies (see Au, 1993). This signal was recorded from a bottlenose dolphin and is considered to be a typical, average signal similar to the ones used by the dolphins in both Au and Pawloski's (1992) study and the Aubauer *et al.* (2000) study. The dolphin signal was generated by a Qua-Tec WK10 function generator housed in a "lunch-box" PC and amplified by a Hafler P3000 Transnova power amplifier. The received signal was gated, filtered and amplified with a custom made gated-amplifier, before being digitized at 1 MHz using a Rapid System R1200 data acquisition system. Two hundred echoes were collected for each object. The cylinder and sphere echoes were recorded on different days.

The frequency span of the stimuli was slowed down to shift the spectra of the echoes into the human hearing range using Audition version 1.0 (Adobe, 2003). The original echoes were digitalized at 1 MHz and had center frequencies around 120 kHz. The echoes were all time stretched by a factor of 167 by converting the echoes to a sampling rate of 6 kHz. The time-stretched echoes had center frequencies around 719 Hz. This stretch factor of 167 was chosen so that the echoes presented to the human listeners would fall near their range of best sensitivity (Green, 1976). The original amplitude relations among echoes was retained but each subject was allowed to adjust the overall amplitude level of the echoes once before beginning the experiment. In previous experiments, the echoes presented to human listeners were at low rates (e.g., 4/s.) compared to the much higher rates that dolphins emit sounds and receive echoes (e.g., 10–50/s; Au, 1993). To better mimic the trains of signals emitted by dolphins, echoes were grouped into clusters of six that lasted 100 ms (echo repetition rate of 60/s.).

The echoes used for training and testing the participants were randomly selected from the set of 200 recorded echoes per object and appeared in random order (each echo was used no more than once for either training or testing). There

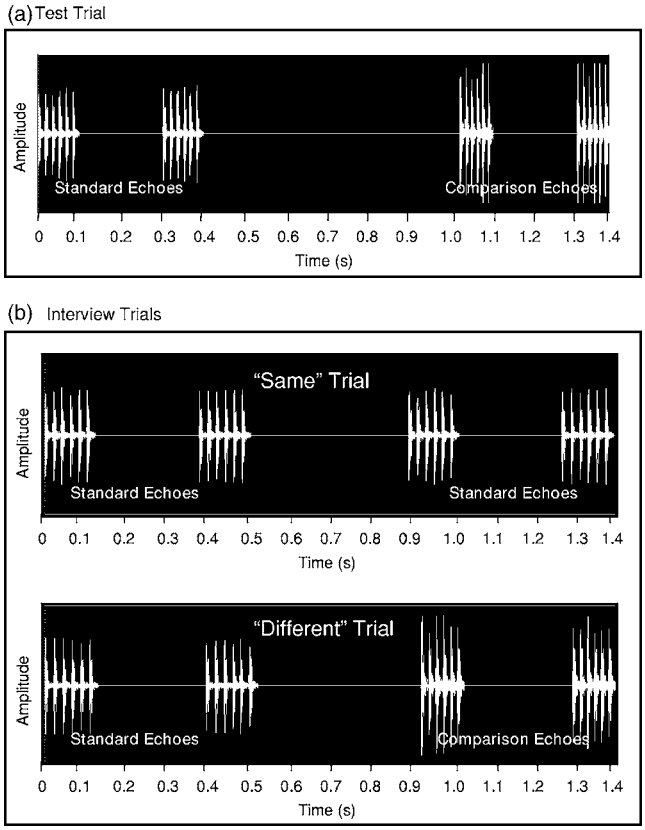


FIG. 2. (a) Example test trial. Two groups of six echoes from the standard object and two groups of six echoes from the comparison object were presented. The standard echoes were presented first on 50% of the trials. In this example, the comparison echoes are from the +0.8 mm sphere. (b) The two types of trials used in the interview phase. "Same" trials contained only echoes from the standard cylinder while "different" trials contained echoes from the standard and one comparison cylinder.

was one training trial created for each object. Each training trial consisted of six echoes followed by 0.6 s silence and then another six echoes (all 12 echoes came from the same object, but each individual echo was different). The echoes used in training trials were not used in the test trials to avoid pre-exposure effects.

Figure 2(a) shows an example of a test trial. Each test trial consisted of two sets of six echoes from the standard object and two sets of six echoes from one of the comparisons (there was no time separation between individual echoes within a set). In half of the test trials the standard echoes came first, and in the other half the comparison echoes came first. The duration of each test trial was approximately 1.35 s. Each set of echoes was approximately 0.1 s in duration and sounded like a single sound (i.e., the four groups of echoes sounded like four separate sounds, not 24 separate sounds).

Each cylinder test session consisted of two blocks, each containing eight trials, for a total of 16 trials per participant. In each of the two eight-trial blocks, each of the four comparison cylinders was presented twice. The sphere test session consisted of two blocks, each containing 16 trials, for a total of 32 trials per participant. In each of the two 16-trial blocks, each of the four comparison spheres was presented four times. For all test sessions the order of the test trials was

randomized separately for each participant. Participants completed the test sessions with no pause between the blocks.

C. Procedure

The dolphins in the previous studies were required to discriminate between a standard target and a set of comparison objects in either a left/right two-alternative forced choice procedure by pressing a paddle on the same side as the standard (Au and Pawloski, 1992) or a go/no-go procedure by pressing a paddle for the standard and remaining still for the comparisons (Aubauer *et al.*, 2000). To simulate the dolphins' tasks, the human listeners in the current study were required to discriminate between the standard target and the comparison targets in a two-alternative forced choice procedure in which they were played sequences of recorded object echoes and had to verbally identify which set of echoes came from the standard. In addition, the human listeners were also asked to identify which of the comparison stimuli was presented on each trial (not a task done by the dolphins) so that they would be encouraged to listen carefully to each stimulus in order to be better able to report the echo features that allowed them to discriminate among the objects.

The human listening experimental set-up consisted of a single experimenter with a laptop computer that channeled the echo stimuli via two Koss UR29 headphones to both the participant and the experimenter (for the purpose of monitoring the echoes). To control for inadvertent cueing, the participant sat faced away from the experimenter during testing. The experimenter ran a randomized playlist for each participants using Panther OS X iTunes software (Apple, 2004) which controlled the sequence of trials and object sets.

Participants were tested individually in a quiet, sound-attenuating room. The participants heard a set of instructions and read a vocabulary sheet with terms to describe the echoes. On the vocabulary sheet were four terms and their operational definitions: "Loudness" (how loud or soft the sound is, i.e., sensation derived from sound intensity), "pitch" (how high or low the sound is, i.e., sensation derived from sound frequency), "duration" (the length of the sound), and "timbre" (the property in musical tones that makes it possible to distinguish one instrument from another or to distinguish voices from instruments, i.e., tone color/sound quality). This procedure ensured that all the participants had the same minimum set of descriptive tools with which they could describe echoic cues (although it was emphasized that they could also use terms not on the list). This vocabulary sheet was available for the participant to reference throughout the experiment. Participants were played pure tone sounds demonstrating each of the terms as they viewed the vocabulary sheet (these sounds were not the echoes used in the experiment). They were presented with a set of sounds that differed in loudness only (same pitch and timbre), a set of sounds that differed in pitch only (880 Hz tone vs 220 Hz tone; same volume and timbre), and a set of sounds that differed in timbre only (middle C played by a french horn, muted trumpet, and soprano sax at the same volume).

All participants were tested on three object sets. Participants were randomly assigned to one of four object set orders: A (spheres, thick cylinders, thin cylinders), B (spheres, thin, thick), C (thick, thin, spheres), and D (thin, thick, spheres). For each object set session there were three phases: training, testing, and interview. The participants completed all three phases for a single object set session before moving on to the next object set session.

1. Training

At the beginning of the object set session participants were given the photograph of the object set (see Fig. 1). Training stimuli were played twice in the same order. For the thin cylinder object set the order was standard, −0.2, −0.3, −0.4, and −0.8 mm; for the thick cylinder object set the order was standard, +0.2, +0.3, +0.4, and +0.8 mm; and for the sphere object set the order was standard, aluminum, brass, nylon, and glass. Participants then listened to a sample of a single test trial and were given instructions regarding the testing procedure (see below). The sample test trial was the same for each participant and contained echoes that were not used in any of the actual test trials. After listening to the sample test trial, then presentation of the training stimuli was repeated again in the same order as before. Finally, participants were allowed to have the training stimuli repeated again as many times as they thought necessary, and in any order, before beginning testing. Participants were not allowed to review the training stimuli once they began the test phase.

2. Testing

The test phase immediately followed the training phase. Each test trial began with a presentation of a test trial stimulus [see Fig. 2(a)]. Participants were allowed to listen to the test trial stimulus as many times as they liked before responding. They had to respond to two questions for each trial: (1) Was the standard object the first or second group of echoes? and (2) which comparison object did the nonstandard echoes come from? For each trial, following the participant's response, the experimenter told the participants whether they were correct or incorrect, and then indicated the correct choice if the participants were incorrect.

3. Interview

Participants immediately began the interview phase when the test phase was completed. The interview phase had two parts: (1) The open-ended interview followed by (2) the close-ended interview trials. The closed-ended interview trials presented an opportunity to test whether the participants would describe the same cues as they did in the open-ended interview (when they freely recalled the cues) compared to when they were blind to the identity of the object in the close-ended trials. In the open-ended interview which lasted approximately 5–10 min, participants' responses to the question "What cues did you use to discriminate among the objects?" were tape-recorded. Then participants listened to a series of closed-ended interview trials. There were eight trials for each of the three object set sessions. Four of the interview trials contained echoes only from the standard object ("same" trials) and the other four trials contained echoes

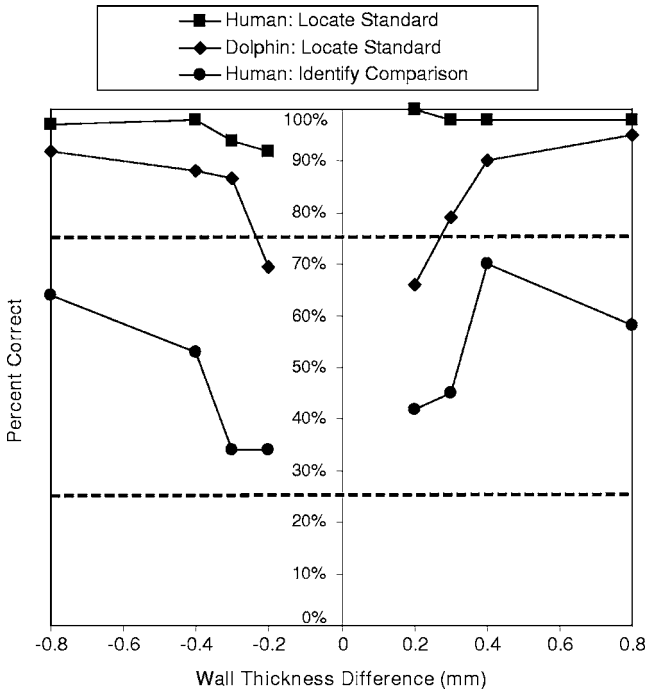


FIG. 3. Performance of the human listeners and the dolphin on the cylinder sets. The top two solid lines show the performance of the human listeners on the task of locating the standard cylinder. The middle two solid lines show the performance of dolphin on the same task (from Au and Pawloski, 1992). The first dotted line shows the 75% correct threshold for the "locate standard" task. The bottom two solid lines show the performance of the human listeners on identifying the comparison cylinder. The bottom dotted line shows chance choice performance (25% correct) on the "identify comparison" task.

from the standard and one comparison object ["different" trials; see Fig. 2(b)]. The participants were required to indicate (1) whether they perceived a difference between the first group of echoes and the second group of echoes and (2) if yes, they were asked to briefly explain which echo features sounded different. Participants were able to complete the training, testing and the interview phase for a single object set session in approximately 20 min. All participants completed the entire experiment in approximately 1 to 2 h (M =58 min).

III. RESULTS

A. Cylinders

1. Performance accuracy

a. Locate standard. The human listeners were able to determine whether the standard echoes were in the first or second group of echoes with nearly perfect accuracy (overall M=97.1%). Figure 3 shows the human listeners' choice accuracy for each of the cylinders (top two lines). The performance of the dolphin is also shown in Fig. 3 (middle two lines, taken from Au and Pawloski, 1992). Au and Pawloski (1992) used a response level of 75% correct to determine the wall thickness discrimination threshold. Interpolation of the dolphin performance data indicated a discrimination threshold of −0.23 mm and +0.27 mm (i.e., the dolphin could not discriminate the ±0.2 mm cylinder from the standard but it could discriminate the ±0.3 mm cylinder from the standard). The same 75% response level was used to determine the discrimination threshold for the human listeners.

To determine whether the human listeners' performance was above 75% on each cylinder, a t-test was performed separately for each cylinder using one score for each participant (average performance on the cylinder) and comparing the participants' scores to a value of 0.75. The participants' performance was significantly above 75% for all eight cylinders (−0.2 mm, $t(15)=4.6$, $p<0.001$; −0.3 mm, $t(15)=3.9$, $p<0.01$; −0.4 mm, $t(15)=15.0$, $p<0.001$; −0.8 mm, $t(15)=10.3$, $p<0.001$; +0.3 mm, $t(15)=15.0$, $p<0.001$; +0.4 mm, $t(15)=15.0$, $p<0.001$; +0.8 mm, $t(15)=15.0$, $p<0.001$; t-test could not be performed for the +0.2 cylinder because all participants achieved 100% correct). A one-way analysis of variance (ANOVA) revealed that the performance of the human listeners did not vary significantly across all eight cylinders, $F(7,120)=1.12, p>0.05$ [preliminary analyses indicated no effect of gender, object set order, block, or position of standard echoes (first or second) so these factors were not included in subsequent analyses].

b. Identify comparison. The human listeners were able to identify six of the eight comparison cylinders (see Fig. 3). Chance choice accuracy is 25%, because the participants could choose from among four alternatives (four cylinders in each set). To determine whether the participants' performance was above chance for each cylinder, a t-test was performed separately for each cylinder using one score for each participant (average performance on the cylinder) and comparing the participants' scores to a value of 0.25. The participants' performance was above chance for all of the cylinders in the thick set (+0.2 mm, $t(15)=3.2$, $p<0.01$; +0.3 mm, $t(15)=3.9$, $p<0.01$; +0.4 mm, $t(15)=8.7$, $p<0.001$; +0.8 mm, $t(15)=4.4$, $p<0.01$) and for two cylinders in the thin set (−0.4 mm, $t(15)=5.6$, $p<0.001$; −0.8 mm, $t(15)=6.5$, $p<0.001$). The participants' performance was not significantly different from chance on the −0.2 mm cylinder [$t(15)=1.7$, $p>0.05$] or the −0.3 mm cylinder [$t(15)=1.5$, $p>0.05$].

A one-way analysis of variance (ANOVA) revealed that the performance of the human listeners varied significantly among the eight cylinders, $F(7,120)=5.24$, $p<0.001$. Simple effects tests indicated that performance on the −0.8 mm cylinder was significantly better than performance on the −0.2 and −0.3 mm cylinders, and performance on the +0.4 cylinder was significantly better than performance on the ±0.2 and ±0.3 mm cylinders.

Table I shows the choices made by the participants in response to the question "What is the identity of the comparison cylinder?" For the thin set, the participants often confused the −0.2, −0.3, and −0.4 mm cylinders (e.g., when the test stimulus was the −0.3 mm cylinder, participants reported that it was the −0.4 mm cylinder 44% of the time). For the thick set, the participants often confused the +0.2 mm cylinder with the +0.3 mm cylinder. These confusions suggest that the participants thought the −0.2, −0.3, and −0.4 mm cylinders sounded similar, as did the +0.2 and +0.3 mm cylinders.

2. Reported use of acoustic cues

Table II shows the echoic cues reported by the human listeners during the open-ended interview and Table III shows the cues reported during the closed-ended interview

TABLE I. Participants' choices on "identify comparison" task for cylinders and spheres.

		Choice			
Cylinders: Thin Set		−0.2 mm	−0.3 mm	−0.4 mm	−0.8 mm
Sample	−0.2 mm	**34%**	31%	16%	19%
	−0.3 mm	8%	**34%**	44%	14%
	−0.4 mm	11%	30%	**53%**	6%
	−0.8 mm	14%	20%	2%	**64%**
		Choice			
Cylinders: Thick Set		+0.2 mm	+0.3 mm	+0.4 mm	+0.8 mm
Sample	+0.2 mm	**42%**	36%	13%	9%
	+0.3 mm	30%	**45%**	17%	8%
	+0.4 mm	9%	20%	**70%**	0%
	+0.8 mm	23%	14%	5%	**58%**
		Choice			
Spheres		Aluminum	Brass	Nylon	Glass
Sample	Aluminum	**55%**	17%	11%	16%
	Brass	13%	**60%**	4%	23%
	Nylon	14%	6%	**74%**	5%
	Glass	27%	16%	18%	**39%**

Note. Each cell represents the percentage of choices made by all 16 participants for each object. Correct answers are shown in bold print (e.g., when the −0.2 mm cylinder was the sample, the participants correctly chose the −0.2 mm cylinder 34% of the time).

trials. The results of the open-ended interview and the closed-ended interview trials were consistent, showing that the participants reported the same cues whether they were asked to recall the cues for each object from their memory (open-ended) or instantly analyze the difference between two sets of echoes without knowing the identity of the objects (close-ended). Thus, the results of both the open- and closed-ended interviews will be discussed together.

The primary cue reported by the participants to discriminate between the standard and the comparison cylinders was pitch and the secondary cue was duration. The participants tended to report using more than one cue (e.g., both pitch and duration). For both the thin and the thick set, participants reported that the pitch of the ±0.2, 0.3, and 0.4 mm increased incrementally relative to the standard (e.g., ±0.3 mm was higher than ±0.2 mm, and ±0.4 mm was higher than ±0.3 mm). They also reported that the ±0.8 mm cylinders were an exception to this trend. Instead of being higher in pitch than the ±0.4 mm cylinder, it was either higher than the standard but lower than the ±0.2 mm cylinder, equal in pitch to the ±0.2 mm cylinder, or slightly lower than the standard. Participants reported that the duration of the ±0.2, ±0.3, and ±0.4 mm cylinders was shorter than the standard. A minority of participants reported using timbre or loudness as cues (e.g., four thought that the ±0.8 cylinder was louder and had a different timbre than the standard).

TABLE II. Echoic cues reported by human participants during the open-ended interviews.

	Object Set		
	Cylinders		
Echoic Cues	Thin set	Thick set	Spheres
Pitch	16	16	13
Loudness	2	3	6
Timbre	4	4	14
Duration	5	10	1
Average number of cues reported per person	1.7	2.1	2.1

Note. Each cell contains the number of participants ($N=16$) who reported using each cue for each of the object sets. Participants could report multiple cues for each set.

3. Cylinder echoes

Figures 4(a) and 4(b) show the cylinder echoes. Two features were measured for each of the 200 echoes per cylinder: (1) Overall duration and (2) the time separation T between the peaks of the first and second highlights of the echo. The time separation T was used to derive a third feature: Time-separation pitch TSP ($1/T$=TSP). Duration was defined as the time between the start of the echo and the end of the last major highlight. Highlights are local maxima in echo amplitude and an example of how major highlights were counted is shown in Fig. 4(d). These features were measured from the original dolphin echoes, then scaled by a factor of 167 since the human listeners listened to slowed-down versions of the original echoes. These "slowed-down" echo features are shown in Table IV.

There were minor differences in the overall durations of the individual cylinder echoes, such that the ±0.2, ±0.3, and ±0.4 cylinders were about 0.001–0.002 s shorter than the

TABLE III. Echoic cues reported by human participants for the cylinders during closed-ended interview trials.

	Thin set				Thick set			
Echoic Cues	−0.8	−0.4	−0.3	−0.2	+0.2	+0.3	+0.4	+0.8
Pitch	15	16	14	14	16	15	16	12
Loudness	1	1	1	0	1	0	1	4
Timbre	2	2	4	3	2	1	3	2
Duration	4	10	9	6	10	9	11	7
Average number of cues reported per person	1.5	1.8	1.8	1.4	1.8	1.6	2.0	1.6

Note. Each cell contains the number of participants ($N=16$) who reported using each cue for each of the object sets. Participants could report multiple cues for each set.

standard. Since the human listeners listened to groups of 6 cylinders, it may be better to compare durations of echo groups (see Table IV), which shows that the echo groups for the ±0.2, ±0.3, and ±0.4 cylinders were about 0.006–0.012 s shorter than the standard. The human listeners reported that they thought the ±0.2, 0.3, and 0.4 mm cylinders sounded shorter than the standard, which accurately reflects the actual measured duration of the echoes.

The time separation between the first and second highlights of the echoes and time-separation pitch also varied between cylinders, but the TSP did not increase incrementally from the thinnest to the thickest cylinders. For the thin cylinders, the lowest TSP was generated by the −0.8 mm (176.8 Hz) and then the cylinders increased in TSP from the standard to the −0.4 mm (standard=179.8 Hz, −0.2 mm =187.8 Hz, −0.3=192.7 Hz, −0.4=196.8 Hz). For the thick cylinders, the lowest TSP was generated by the standard (179.8 Hz) and the comparison cylinders were all higher than the standard. From lowest to highest they were as follows: +0.8 mm (186.4 Hz), +0.3 mm (191.4 Hz), +0.4 mm (193.7 Hz), and +0.2 mm (203.7 Hz). The participants' reports reflected these changes in TSP. They reported that the pitch of the ±0.2, 0.3, and 0.4 mm cylinders was higher than the standard and that the ±0.8 mm cylinders were an exception to this trend (±0.8 mm were not higher than ±0.4 mm, but slightly higher or lower than the standard). [Note that humans can distinguish between two tone bursts under 200 Hz that differ by less than 1 Hz so all the cylinders should be discriminable using TSP (Wier *et al.*, 1977).]

Figure 5 shows histograms representing the number of echoes (out of 200 per cylinder) at each TSP for each cylinder in both the thin and thick sets. These graphs show that (1) each cylinder covered a range of TSP values, and (2) there was overlap between the cylinders such that there could be more than one cylinder with any given TSP. For example, an echo with a TSP of 193.16 Hz is almost equally as likely to have come from the −0.3 or −0.4 mm cylinder (there is also a small chance it could have come from the −0.2 mm cylinder; see Fig. 5). The same is true for the thick cylinders: An echo with a TSP of 196.3 Hz could have come from the +0.2 mm, +0.3 mm, or +0.4 mm cylinder. Since the echoes presented to the human listeners were chosen randomly from the 200 echoes recorded for each cylinder, it is likely they heard these variations in the pitch of the echoes from trial to

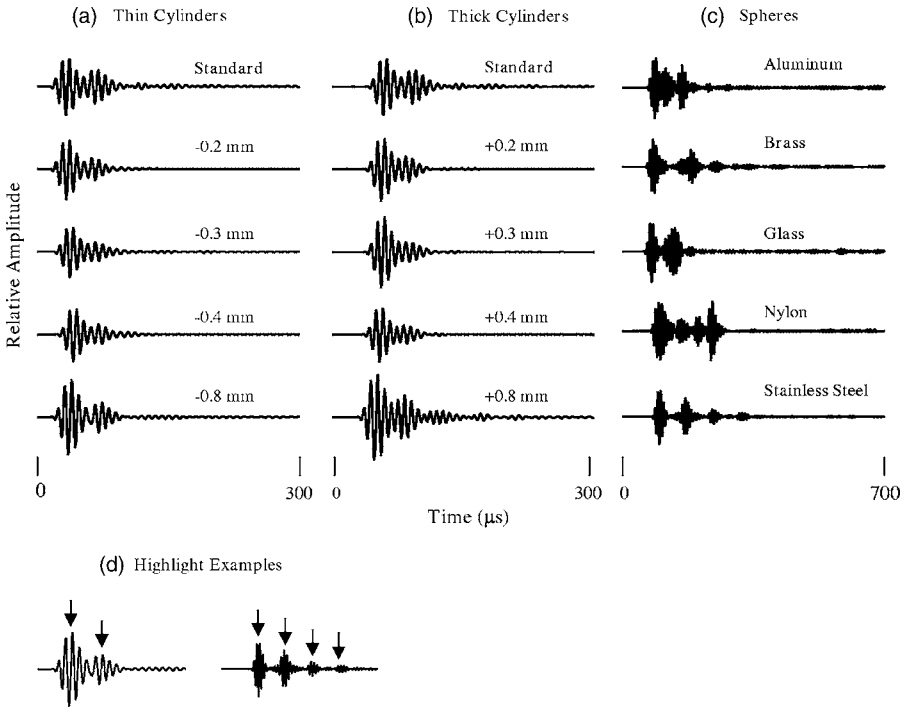


FIG. 4. Object echoes from the thin cylinder set (a), the thick cylinder set (b), and the spheres (c). Figure (d) shows the location of echo highlights in the −0.8 mm cylinder (left) and the steel sphere (right).

TABLE IV. Cylinder and sphere echo measurements.

	Echo Feature					
	Time Separation T (ms)		TSP (Hz)		Duration (s)	
Cylinders	*M*	*SD*	*M*	*SD*	Indiv.	Group
−0.8 mm	5.66	0.06	176.83	1.93	0.006	0.036
−0.4 mm	5.08	0.04	196.84	1.60	0.004	0.024
−0.3 mm	5.19	0.04	192.68	1.41	0.005	0.030
−0.2 mm	5.32	0.09	187.86	3.27	0.005	0.030
Standard	5.56	0.07	179.85	2.50	0.006	0.036
+0.2 mm	4.91	0.06	203.76	2.50	0.005	0.030
+0.3 mm	5.22	0.08	191.48	3.00	0.005	0.030
+0.4 mm	5.16	0.09	193.76	3.23	0.005	0.030
+0.8 mm	5.36	0.05	186.46	1.62	0.006	0.036
Spheres						
Steel	11.54	0.05	86.66	0.36	0.021	0.126
Aluminum	5.22	0.79	195.58	27.77	0.011	0.066
Brass	16.94	0.13	59.02	0.47	0.017	0.102
Nylon	11.27	1.12	89.70	9.91	0.017	0.102
Glass	8.36	0.57	120.07	7.39	0.011	0.066

Note. All measurements are given for the slowed down echoes that were presented to the human listeners, not the original echoes. Means (M) and standard deviations (SD) are calculated for 200 echoes per object. T = time separation between first and second highlight in the echo. Time separation pitch (TSP) is calculated from $1/T$. Duration is given for individual echoes and groups of six echoes.

trial. The participants' confusions between the comparison cylinders reflect that they may have heard these overlaps in TSPs (see Table I). For example, the participants' confusions suggest that they thought the −0.2, −0.3, and −0.4 mm cylinders sounded similar, and Fig. 5 shows that those three cylinders had overlapping values between 187 and 193 Hz.

B. Spheres

1. Performance accuracy

a. Locate standard. The human listeners were able to determine whether the standard sphere was the first or second group of echoes (overall $M=88.3\%$). Choice accuracy for each comparison sphere is as follows: Aluminum (89.1%), brass (80.5%), nylon (90.1%), and glass (93.0%). To determine whether the participants' performance was above chance (50%) for each cylinder, a t-test was performed separately for each cylinder using one score for each participant (average performance on the cylinder) and comparing the participants' scores to a value of 0.5. The participants' performance was significantly above chance for all four spheres (aluminum, $t(15)=10.4$, $p<0.001$; brass, $t(15)=6.0$, $p<0.001$; nylon, $t(15)=9.7$, $p<0.001$; glass, $t(15)=13.3$, $p<0.001$). A one-way analysis of variance (ANOVA) revealed that the performance of the human listeners did not vary significantly across the four spheres, $F(3,60)=1.74$, $p>0.05$.

The dolphin was also able to discriminate between the standard sphere and the comparison spheres (Aubauer *et al.*, 2000). In the dolphin's go/no-go task, a single target (either a standard or comparison) was presented on each trial, unlike the humans' two-alternative forced choice task in which a set of standard and comparison echoes were presented on each trial. Thus, there is a score for the standard steel sphere (93%), as well as each of the comparisons: Aluminum (77%), brass (89%), and nylon (100%; the glass sphere was not presented to the dolphin). Also, the dolphin received more brass trials ($n=299$) than aluminum ($n=13$) or nylon trials ($n=13$). To determine whether the human listeners' performance was significantly different from the dolphin's, a t-test was performed separately for each sphere using one score for each participant (average performance on the sphere) and comparing the participants' scores to a value matching the dolphin's score (aluminum=0.77, brass=0.89, nylon=1.0). The human listeners' performance was not significantly different from the dolphin on the brass sphere [$t(15)=-1.7$, $p>0.05$]. The human listeners' performance on the aluminum sphere was significantly better than the dolphin's (Ms=89% vs 77%; $t(15)=3.2$, $p<0.01$), but the dolphin's performance on the nylon sphere was significantly better than the humans' (Ms=100% vs 90%; $t(15)=-2.2$, $p<0.05$).

b. Identify comparisons. The human listeners were able to identify three of the four comparison spheres. Choice accuracy for each comparison sphere is as follows: Aluminum (55.5%), brass (60.2%), nylon (74.2%), and glass (39.1%). Chance choice accuracy is 25%, because the participants could choose from among four alternatives. To determine whether the participants' performance was above chance for each sphere, a t-test was performed separately for each sphere using one score for each participant (average performance on the sphere) and comparing the participants' scores to a value of 0.25. The participants' performance was above chance for three of the spheres [aluminum, $t(15)=5.3$, $p<0.001$; brass, $t(15)=5.2$, $p<0.001$; nylon, $t(15)=8.0$, $p<0.001$] but not the glass sphere [$t(15)=2.1$, $p=0.05$]. A one-way analysis of variance (ANOVA) revealed

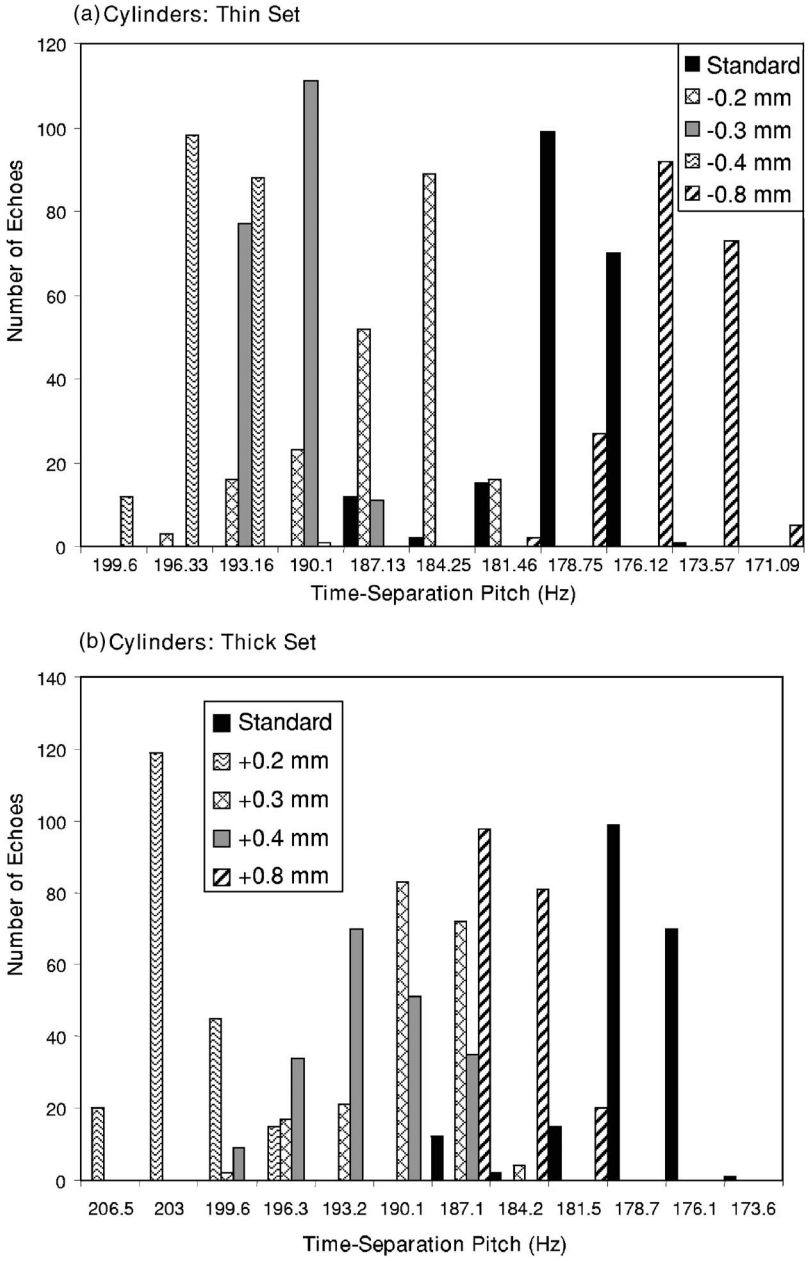


FIG. 5. Histograms representing the number of echoes (out of 200 per cylinder) at each time-separation pitch (TSP) for each cylinder in the thin set (a) and the thick set (b).

that the performance of the human listeners varied significantly among the spheres, $F(3,60)=5.18$, $p<0.01$. Simple effects tests indicated that performance was best on the nylon sphere and worst on the glass sphere.

Table I shows the choices made by the participants in response to the question "What is the identity of the comparison sphere?" When aluminum was the sample, the participants confused it with brass or glass on 33% of the trials. When brass was the sample, the participants confused it with aluminum or glass on 36% of the trials. When glass was the sample, they confused it with aluminum on 27% of the trials. These confusions suggest that they thought the aluminum, brass, and glass spheres sounded similar. When nylon was the sample, they usually correctly answered "nylon" (74% of trials) and did not confuse it with the other spheres, which suggests that the nylon sphere sounded distinctive compared to the other spheres.

2. Reported use of acoustic cues

As with the cylinders, the results of the open-ended interview and the closed-ended interview trials were consistent for the spheres, so the results of both will be discussed together.

The two main cues reported by the participants to discriminate between the standard and the comparison spheres were pitch and timbre (see Tables II and V). The participants tended to report using more than one cue (e.g., both pitch and timbre). The majority of the participants reported that the timbre of the spheres varied, but their descriptions were diverse (e.g, the aluminum sphere was "percussive," or "flat;" the brass sphere was "hollow and distant" or "ringy;" the nylon sphere was "fuzzy" or "damp;" and the glass sphere was "sharp" or "crisp"). The majority of the participants reported that pitch was a cue, but opinions varied as to whether each comparison was higher, lower, or the same pitch as the

TABLE V. Echoic cues reported by human participants for the spheres during closed-ended interview trials.

	Object			
Echoic Cues	Aluminum	Brass	Nylon	Glass
Pitch	15	11	12	12
Loudness	0	4	5	3
Timbre	7	11	9	9
Duration	0	1	0	0
Average number of cues reported per person	1.5	1.8	1.8	1.5

Note. Each cell contains the number of participants ($N=16$) who reported using each cue for each of the objects. Participants could report multiple cues for each set.

standard (e.g., in the open-ended interview, seven reported that aluminum was higher than the standard, four reported that it was lower). The majority opinion was that the nylon, brass, and glass spheres had a lower pitch than the standard, and the aluminum sphere had a higher pitch than the standard. A minority of participants reported using loudness and duration as cues, but there was no consensus as to how these features varied (e.g., two reported that glass was softer and one reported that glass was louder than the standard).

3. Sphere echoes

Figure 4 shows the sphere echoes and Table IV shows the echo measurements. The aluminum and glass echoes were the shortest in duration (0.011 s), followed by brass and nylon (0.017 s), and the steel echoes were longest (0.021 s). Only one participant reported using differences in echo duration to discriminate among the echoes, which suggests either the participants did not pick up on the duration differences in the echoes, or that duration was simply not as salient as the other two cues they reported (pitch and timbre) so they ignored duration differences.

For the spheres, even though there were more than two highlights in all the echoes, the time separation T was still defined as the time interval between the first and second highlight and TSP was calculated using that value. The TSP of the spheres from lowest to highest was as follows: Brass (59.0 Hz), steel (86.7 Hz), nylon (89.7 Hz), glass (120.1 Hz), and aluminum (195.6 Hz). Three of the spheres had large standard deviations, indicating that the values obtained for the 200 echoes were spread over a wide range (aluminum: $M=195.6$, $SD=27.8$; nylon: $M=89.7$, $SD=9.9$; glass: $M=120.1$, $SD=7.4$).

The participants reported differences in pitch among the sphere echoes, but the differences they reported between the standard steel echo and the comparison echoes accurately matched differences in the calculated TSP values for the aluminum, brass, and nylon spheres but not the glass sphere (see Table IV). The majority of participants reported that aluminum was higher in pitch than steel, and that brass was lower in pitch than steel, and both opinions match the actual TSP values. The majority of the participants reported that the nylon sphere was lower in pitch than the steel sphere, which in fact was the case for 55% (110/200) of the echoes since TSPs for nylon ranged from 78.8 to 115.1 Hz. The majority of participants reported that the glass sphere was lower in pitch than the steel sphere, but the calculated TSP of all the glass echoes was higher than steel.

IV. DISCUSSION

A. Cylinders

Both the human listeners and the dolphin were able to discriminate between the standard cylinder and the ±0.8 mm, ±0.4 mm, and ±0.3 mm comparison cylinders. However, the human listeners were able to discriminate between the standard and the ±0.2 mm comparisons at the 75% correct response level whereas the dolphin could not. The human listeners' performance appeared to be significantly better than the dolphin's primarily on the ±0.2 mm cylinders and the +0.3 mm cylinder (see Fig. 3).

The human listeners' performance could have been higher than the dolphin's for several reasons. First, the echoes for the human listeners were collected in a test tank and presented in a controlled laboratory situation. The echoes had a high signal to noise ratio and they did not contain any extraneous echoes (e.g., echoes from passing fish) whereas the dolphin's echoes may have contained background noise (e.g., from snapping shrimp) and extraneous echoes. A second reason the human listeners performed differently than the dolphin could be the way in which the echoes were presented to the human listeners in two groups of six echoes, whereas the dolphin made its choice based on variable numbers of echoes (e.g., it could emit 10 clicks on one trial and 50 clicks on the next). In the future, it would be useful to collect the dolphin's actual echoes during the task, and present those echoes to the human participants so that both the dolphins and the humans are basing their decisions on the same amount and quality of echo information.

The human listeners' performance could have been different from the dolphin's due to the rate at which the echoes were slowed down to bring them into the human hearing range. In this study the echoes were time-stretched by a factor of 167 so that their center frequencies were around 719 Hz. Other studies have utilized a time-stretch factor of 50, corresponding to an echo center frequency around 2.4 kHz (Au and Martin, 1989; Helweg *et al.*, 1995), or a time-stretch factor of 125 corresponding to an echo center frequency around 1 kHz (DeLong *et al.*, 2006a). It is possible that playing back the echoes at different frequencies (via different slow-down rates) could change the results.

Finally, the human listeners' performance could have been different than the dolphin's because of the exclusion strategy adopted by the dolphin. In the dolphin's task, one target was presented to the left and one target to the right of the dolphin (the standard and one comparison) and it was required to press a paddle on the same side as the standard. The human listeners had an analogous task of listening to echoes from the standard and one comparison and indicating whether the standard echoes came first or second. Thus, both the humans and the dolphins were presented with both the standard and comparison within each trial. However, measurement of the dolphin's echolocation signals indicated that

it emitted 96% of its signals toward the target located on the right, meaning that it did not typically listen to echoes from both the standard and the comparison within a single trial (Au and Pawloski, 1992). The dolphin had adopted an exclusion strategy in which it examined a single target and determined whether that target was the standard or not, i.e., the dolphin memorized the echo characteristics of the standard target and compared the received echoes from the right target to that template of the standard. In contrast, the human listeners always had access to echoes from both targets and had to listen to both stimuli before reporting a choice. Although the human listeners undoubtedly memorized some characteristics of the standard target, they may have performed better than the dolphin because they could have used a different strategy than the dolphin (listen to both targets, not just one).

There are some discrepancies between this study and Au and Pawloski (1992) in the echo measurements of the same cylinders. For example, Au and Pawloski (1992) measured echoes in 1992 and estimated that the TSP of the standard, −0.2 mm, and −0.3 mm cylinder were 28.3, 27.9, and 27.8 kHz, respectively. The current study, using a different set of echoes measured in 2005 estimated the average TSP of the same three cylinders as 30.0, 31.4, and 32.2 kHz.

There are at least two factors that may explain why the echo measurements of the same cylinders taken in 1992 and 2005 were not the same. First, the speed of sound of water in a factor since the wave travels through the water that is contained in the inner diameter of the target. The speed of sound varies in saltwater (1,513 m/s) vs fresh water (1,493 m/s), and the measurement tank may have had different quantities of salt water in the two years due to evaporation and rainfall. Second, the echoes from these cylinders are very sensitive to vertical orientation and just a minor shift in tilt cause by wind induced waves can cause large changes. Au and Pawloski (1992) suspended the target with a rig that probably provided good stability and vertical alignment (20 echoes per target were measured). In the current study, the cylinders were measured by suspending a line through the holding bar across the cylinders without perfectly stabilizing the cylinders (200 echoes per target were measured). The present measurements probably represent more closely what the dolphin encountered since a special rig was not used to present the targets to the dolphin. The current echoes presented to the human listeners probably represented the dolphin's situation more closely (i.e., having to make decisions based on echoes from the same target that vary slightly from trial to trial).

Au and Pawloski (1992) proposed three cues that could be used to discriminate among the cylinders: Time domain cues (time differences between two echo highlights), frequency domain cues (changes in frequency of prominent spectral features such as notches), and time-separation pitch (TSP). The first is linked to the envelope of the time waveform and the second is linked to the spectral envelope, and both are implicated in the perception of timbre (ANSI, 1960). A minority of the participants (up to 25%) reported using timbre, so it is possible that these first two cues played a role in discrimination. However, the primary echo feature reported by nearly all of the human listeners as a discriminatory cue was pitch, and the pitch they heard could be the TSP created by the two main highlights in the cylinder echoes, since this type of signal would produce TSP in the human auditory system (Thurlow, 1957). The dolphin may also have used TSP cues.

Other researchers have proposed that dolphins could make use of TSP cues. Au and Martin (1988) suggested that a dolphin used TSP to discriminate among metallic plates that varied in material composition and thickness. Hammer and Au (1980) suggested that TSP cues were used by a dolphin to discriminate certain hollow aluminum standard cylinders from comparison cylinders of different material composition, or of the same material composition but with different wall thickness and internal structure (hollow or solid). Au and Pawloski (1989) have shown that dolphins can discriminate noise having a rippled power spectrum from noise having a nonrippled spectrum. Noise having a rippled spectrum produces TSP in humans, which suggests that a dolphin may also have the capability to perceive TSP. However, there is no direct evidence that a dolphin can perceive TSP. This study suggests that further research is warranted to determine whether dolphins perceive TSP cues.

B. Spheres

Both the human listeners and the dolphin in the Au *et al.* (2000) study were able to discriminate between the standard steel sphere and the comparisons (brass, aluminum, nylon) with high accuracy (overall *M*s=86% vs 89%, respectively). Both the humans and the dolphin performed best on the nylon sphere (*M*s=100% vs 90%) in the "locate standard" task. In addition, in the "identify comparison" task the humans performed relatively well with the nylon sphere (74%) compared with the other spheres (aluminum=56%, brass=60%). They often confused the aluminum and brass spheres but did not often confuse the nylon spheres with the others. These results indicate that the nylon echoes may have been distinctive to both the dolphin and the humans compared with the other spheres' echoes. The human listeners' performance was also similar to the dolphin's on the brass sphere (*M*s=80% vs 89%), although their performance on the aluminum sphere was significantly better than the dolphin's (*M*s=89% vs 77%). Thus, the dolphin's and humans' performance was comparable on two (nylon, brass) of the three spheres. It is possible that the dolphin and the humans may have been using at least some of the same echo features to discriminate among the spheres.

The human listeners in the current study reported using primarily pitch and timbre to discriminate among the spheres of varying materials. The participants' reports of differences in pitch between the spheres sometimes but did not always match the TSP values that were calculated for the echoes. This could mean that the participants' perception of pitch was not always accurate (i.e., not reflecting the actual TSP of the echoes). One reason for this could be the large standard deviation in TSP values for three of the spheres that made TSP values inconsistent. Since they heard a random selection of echoes from the 200 recorded echoes per object, the TSP values for any given sphere were not always the same. Al-

ternately, the "pitch" reported by the participants could have been something other than time-separation pitch, like the "click pitch" reported by human listeners in Au and Martin (1989) and defined by the authors as the pitch associated with the peak frequency of the echo.

The results of this study compare favorably with the results of other human listening studies in overall accuracy rates and use of pitch and possibly also timbre as material discrimination cues (Au and Martin, 1989; DeLong *et al.*, 2006a). Au and Martin (1989) presented human listeners with echoes from cylinders varying in size and material that had been used in dolphin experiments (Hammer and Au, 1980; Schusterman *et al.*, 1980). The participants in Au and Martin's (1989) study reported using pitch to discriminate between the aluminum, bronze, and steel cylinders (e.g., bronze had a lower pitch than aluminum). The authors suggest the participants were using time-separation pitch (bronze TSP=385 Hz, aluminum TSP=444 Hz). For the aluminum vs glass discrimination, the time between the first and second highlights was virtually the same, which eliminates TSP as a cue. The participants reported using echo duration to discriminate between the aluminum and glass echoes, but when Au and Martin (1989) eliminated duration as a cue, they reported a secondary cue they called "click pitch," defined by the authors as the pitch associated with the peak frequency of the echo. This click pitch could be similar to what the participants in the current experiment called either the "pitch" or the "timbre" of the echo. DeLong *et al.* (2006a) presented human listeners with objects that varied primarily in material but also had some slight variations in shape and size and they reported using pitch and timbre. Taken together, all three studies (current study; Au and Martin, 1989; DeLong *et al.*, 2006a) show that human listeners report pitch and timbre and to a lesser extent duration and amplitude as material discrimination cues.

Echo features that contribute to the perception of pitch and timbre should be explored as material discrimination cues for dolphins. Yet it is difficult to determine exactly which features of echoes would give rise to the perception of timbre. Timbre is a complex sound quality that is a combination of acoustic parameters. According to the American National Standards Institute (1960), timbre is defined as "that attribute of sensation in terms of which a listener can judge two sounds similarly presented and having the same loudness and pitch as dissimilar" and that "timbre depends primarily upon the spectrum of the stimulus, but it also depends upon the waveform, the sound pressure, the frequency location of the spectrum, and the temporal characteristics of the stimulus." Timbre is not independent from other acoustic parameters, but appears to arise from an interaction of pitch, amplitude, and time features. This may make it difficult to investigate whether dolphins perceive or use timbre.

C. Conclusions

The human listeners in this study typically performed as well or better than the dolphin and they were able to report the echo features that allowed them to discriminate among the objects: Pitch and duration for wall thickness discrimination, and pitch and timbre for material discrimination. The echo features used by the humans could also have been used by the dolphins. However, the overall similar discrimination abilities in humans and dolphins found in this study are not conclusive evidence that they are using the same echo features. A more in-depth analysis of performance, such as the analysis of errors (object confusions), provides one way to determine whether humans and dolphins share the use of certain features. Two objects are confused to the degree that they share similar features. If the error patterns of the humans and the dolphin match (i.e., they confuse the same objects), it would imply that they may have used the same features. Conversely, if the error patterns do not match, it would imply they may have used different features. This error analysis was used successfully in a different study (DeLong *et al.*, 2006a). It was not possible in the current study because a two-alternative forced choice or go/no-go method was used with the dolphin instead of a method that yields object confusions (e.g., match-to-sample).

Similarities and differences between human and dolphin auditory perception should be considered when deciding whether echo features reported by humans could be used by dolphins. Since both dolphins and humans can discriminate sounds that vary in intensity by about 1 dB (Evans, 1973; Green, 1993) and the frequency discrimination abilities of dolphins and humans for tonal stimuli are comparable in the range of best hearing for each species (Herman and Arbeit, 1972; Thompson and Herman, 1975; Weir *et al.*, 1976), differences in echo intensity and frequency should be available as cues to both humans and dolphins. This means that if humans report amplitude and pitch cues in the echoes it is reasonable to assume that dolphins also have access to these cues. However, dolphins have sharper frequency tuning curves than humans (Supin and Popov, 1995) so they may be better able to interpret frequency information in echo stimuli than the human listeners.

Temporal resolution is the ability to rapidly resolve stimuli that are close together in time (e.g., to identify two stimuli as separate events rather than a single stimulus). Popov and Supin (1997) recorded the dolphin auditory brainstem response evoked by a short gap in noise and found that the dolphin's hearing is sensitive to extremely short gap durations as compared to other animals and humans. The dolphin gap detection threshold was 0.1 ms (100 μs) whereas gap detection thresholds in humans were an order of magnitude longer than dolphins (about 2.2 ms; Snell *et al.*, 1994). At first this would appear to give dolphins an advantage over humans and it might be assumed that dolphins could access finer temporal cues. However, it is important to note that when the echoes are slowed down to bring them into the human hearing range, the duration of the echoes is lengthened relative to the echoes the dolphin hears. For the dolphins, the echoes in this study were approximately 24–125 μs in duration. When the echoes were slowed down for the human listeners, they were each approximately 4–21 ms in duration. Because dolphins must be able to resolve temporal differences in signals that last for μs, whereas humans must be able to resolve temporal differences in signals that last for ms, the process of slowing the echoes down

may in effect allow the temporal processing abilities of humans to "catch up" to the dolphins' abilities to some extent.

Human listening studies are just one way of elucidating how dolphins use information in echoes to discriminate among objects. Although caution must be used in interpreting the results, since there are some ways in which dolphin and human auditory detection and perception are different, it is still a worthwhile endeavor because of the relative challenge of performing experiments with dolphins (extensive training time, few research animals, expensive animal care and maintenance). In human listening studies, many experiments can be performed in a short time and subjects can verbally report salient acoustic cues. Human listeners can quickly identify salient echo features that permit object discrimination, which can be used to generate hypotheses that can be tested using dolphins as subjects. For example, this study suggests that time-separation pitch (TSP) is a cue potentially used for both material and wall thickness discrimination so it is worthwhile to conduct an experiment with dolphins to test their ability to perceive TSP, and then explore whether TSP is a salient echo feature for dolphins.

ACKNOWLEDGMENTS

We would like to thank James Simmons for his assistance in formatting the echoes and Andrea Simmons for the use of a sound-treated experimental chamber. We thank James Simmons, Kelly Benoit-Bird, Nathaniel Gibbs, and three anonymous reviewers for their constructive suggestions on earlier versions of this manuscript.

ANSI (**1960**). *American Standard Acoustical Terminology*, American National Standards Institute, New York.

Au, W. W. L. (**1993**). *The sonar of dolphins* (Springer, New York).

Au, W. W. L. (**2000**). "Echolocation in dolphins," in *Hearing by whales and dolphins*, edited by W. W. L. Au, A. N. Popper, and R. R. Fay (Springer, New York), pp. 364–408.

Au, W. W. L., and Martin, D. W. (**1988**). "Sonar discrimination of metallic plates," in *Animal sonar: Processes and performance*, edited by P. E. Nachtigall and P. W. B. Moore (Plenum, New York), pp. 809–813.

Au, W. W. L., and Martin, D. W. (**1989**). "Insights into dolphin sonar discrimination capabilities from human listening experiments," J. Acoust. Soc. Am. **86**, 1662–1670.

Au, W. W. L., and Pawloski, J. L. (**1989**). "Detection of ripple noise by an Atlantic bottlenose dolphin," J. Acoust. Soc. Am. **86**, 591–596.

Au, W. W. L., and Pawloski, D. A. (**1992**). "Cylinder wall thickness difference discrimination by an echolocating Atlantic bottlenose dolphin," J. Comp. Physiol., A **170**, 41–47.

Aubauer, R., Au, W. W. L., Nachtigall, P. E., Pawloski, D. A., and DeLong, C. M. (**2000**). "Classification of electronically generated phantom targets by an Atlantic bottlenose dolphin (*Tursiops truncatus*)," J. Acoust. Soc. Am. **107**(5), 2750–2754.

DeLong, C. M., Au, W. W. L., Harley, H. E., Roitblat, H. L., and Pytka, L. (**2006a**). "Human listeners provide insights into echo features used by dolphins to discriminate among objects," (in preparation).

DeLong, C. M., Au, W. W. L., Lemonds, D. W., Harley, H. E., and Roitblat, H. L. (**2006b**). "Acoustic features of objects matched by a bottlenose dolphin," J. Acoust. Soc. Am. **119**(3), 1867–1879.

Evans, W. E. (**1973**). "Echolocation by marine delphinids and one species of fresh-water dolphin," J. Acoust. Soc. Am. **54**, 191–199.

Fish, J. F., Johnson, C. S., and Ljungblad, D. K. (**1976**). "Sonar target discrimination by instrumented human divers," J. Acoust. Soc. Am. **59**, 602–606.

Hammer, C. E., Jr., and Au, W. W. L. (**1980**). "Porpoise echo-recognition: An analysis of controlling target characteristics," J. Acoust. Soc. Am. **68**, 1285–1293.

Green, D. M. (**1976**). *An introduction to hearing* (Lawrence Erlbaum Associates, Hillsdale, NJ).

Green, D. M. (**1993**). "Auditory intensity discrimination," in *Human psychophysics*, edited by W. A. Yost, A. N. Popper, and R. R. Fay (Springer, New York), pp. 13–55.

Helweg, D. A., Roitblat, H. L., Nachtigall, P. E., Au, W. W. L., and Irwin, R. J. (**1995**). "Discrimination of echoes from aspect-dependent targets by a bottlenose dolphin and human listeners," in *Sensory systems of aquatic mammals*, edited by R. A. Kastelein, J. A. Thomas, and P. E. Nachtigall (De Spil Publishers, Woerden, The Netherlands), pp. 129–136.

Herman, L. M., and Arbeit, W. R. (**1972**). "Frequency discrimination limens in the bottlenose dolphin: 1–70 Ks/c," J. Aud Res. **2**, 109–120.

Johnson, C. S. (**1967**). "Sound detection thresholds in marine mammals," in *Marine bio-acoustics*, edited by W. N. Tavolga (Pergamon, New York), pp. 247–260.

Kellogg, W. N. (**1962**). "Sonar system of the blind," Science **137**, 399–404.

Popov, V. V., and Supin, A. Ya. (**1997**). "Detection of temporal gaps in noise in dolphins: Evoked potential study," J. Acoust. Soc. Am. **102**(2), 1169–1176.

Rice, C. E. (**1967**). "Human echo perception," Science **155**, 656–664.

Schusterman, R. J., Kersting, D. A., and Au, W. W. L. (**1980**). "Response bias and attention in discriminative echolocation by *Tursiops truncatus*," in *Animal sonar systems*, edited by R. G. Busnel and J. F. Fish (Plenum, New York), pp. 983–986.

Snell, K. B., Ison, J. R., and Frisina, D. R. (**1994**). "The effects of signal frequency and absolute bandwidth on gap detection in noise," J. Acoust. Soc. Am. **96**, 1458–1464.

Supin, A. Ya., and Popov, V. V. (**1995**). "Frequency tuning and temporal resolution in dolphins," in *Sensory systems of aquatic mammals*, edited by R. A. Kastelein, J. A. Thomas, and P. E. Nachtigall (De Spil Publishers, Woerden, The Netherlands), pp. 95–110.

Thompson, R. K. R., and Herman, L. M. (**1975**). "Underwater frequency discrimination in the bottlenose dolphin (1–140 kHz) and the human (1–8 kHz)," J. Acoust. Soc. Am. **57**, 943–948.

Thurlow, W. R. (**1957**). "Further observations on pitch associated with a time difference between two pulse trains," J. Acoust. Soc. Am. **29**, 1310–1311.

Thurlow, W. R., and Small, A. M., Jr. (**1955**). "Pitch perception for certain periodic auditory stimuli," J. Acoust. Soc. Am. **27**, 132–137.

Wier, C., Jesteadt, W., and Green, D. (**1977**). "Frequency discrimination as a function of frequency and sensation level," J. Acoust. Soc. Am. **61**, 178–184.

Evoked-potential recovery during double click stimulation in a whale: A possibility of biosonar automatic gain control

Alexander Ya. Supin[a)]
Institute of Ecology and Evolution of the Russian Academy of Sciences, 33 Leninsky Prospect, 119071 Moscow, Russia

Paul E. Nachtigall[b)] and Marlee Breese
Marine Mammal Research Program, Hawaii Institute of Marine Biology, University of Hawaii, P.O. Box 1106, Kailua, Hawaii 96734

(Received 7 June 2006; revised 15 September 2006; accepted 2 October 2006)

False killer whale *Pseudorca crassidens* auditory brainstem responses (ABR) were recorded using a double-click stimulation paradigm specifically measuring the recovery of the second response (to the test click) as a function of the inter-click interval (ICI) at various levels of the conditioning and test click. At all click intensities, the slopes of recovery functions were almost constant: 0.6–0.8 μV per ICI decade. Therefore, even when the conditioning-to-test-click level ratio was kept constant, the duration of recovery was intensity-dependent: The higher intensity the longer the recovery. The conditioning-to-test-click level ratio strongly influenced the recovery time: The higher the ratio, the longer the recovery. The dependence was almost linear using a logarithmic ICI scale with a rate of 25–30 dB per ICI decade. These data were used for modeling the interaction between the emitted click and the echo during echolocation, assuming that the two clicks simulated the transmitted and echo clicks. This simulation showed that partial masking of the echo by the preceding emitted click may explain the independence of echo-response amplitude of target distance. However, the distance range where this mechanism is effective depends on the emitted click level: The higher the level, the greater the range. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2382343]

PACS number(s): 43.80.Lb [WWA] Pages: 618–625

I. INTRODUCTION

Many of the basic mechanisms underlying the echolocation of odontocetes (dolphins, porpoises, and toothed whales) remain uncertain in spite of the general investigation over the past few decades (Nachtigall and Moore, 1988; Au, 1993; Thomas *et al.*, 2003). In particular, the problem of automatic gain control within the auditory system requires additional investigation and explanation. Echo levels vary because of differences in target strength and distance from the animal to the target. Therefore, the performance of the biosonar requires that its receiving part (the auditory system) be capable of analyzing echoes under these widely varying conditions. It is not known yet, whether this capability is achieved by a kind of automatic gain control which keeps the response of the auditory system to the echo rather constant, or if the auditory system is equally capable of analyzing the echo-signals at a wide variety of neuronal response levels.

One way to decrease the variation of the echo level is by adjusting the level of the transmitted pulse. Experiments in the wild have shown that dolphins and whales vary the intensity of their echolocation pulses according to the distance to the target, roughly at a rate of 20 dB per distance decade (Rasmussen *et al.*, 2002; Au and Benoit-Bird, 2003; Au and Herzing, 2003; Au and Würsig, 2004; Au *et al.*, 1988). To some extent, this may compensate for the echo attenuation with distance, although the compensation may be not complete since, depending on the target size, the actual echo attenuation varies from 20 to 40 dB per distance decade. However, this mechanism cannot entirely solve the problem of automatic gain control because varying the transmitted pulse intensity does not solve the problem of the transmission-to-echo ratio. If this ratio is very high (because of the low target strength and/or large distance), the echo may be masked by the much more intense preceding transmitted pulse.

Another mechanism of the automatic gain control in the odontocete's sonar involves a variation of sensitivity of the auditory system to the echo. Recording of brain stem evoked potentials associated with natural echolocation has shown that the response to the echo remained almost constant with variation of target distance, although the intensity of transmitted echolocation pulses varied only slightly (Supin *et al.*, 2004, 2005). It was assumed that one explanation for the constancy of the echo response was its delay-dependent release from forward masking by the previous louder transmitted pulse as had been previously reported in two pulse experiments (Popov and Supin, 1990). Indeed, with the target distance increase, the echo delay increases proportionally, which results in release from masking. The rate of this release in dolphins is 30–40 dB per delay decade. Interestingly, the echo intensity also decreases with a rate from 20 to 40 dB per distance decade, depending on the target size to sound wavelength ratio. These two opposite processes may compensate one another to a large extent: With target dis-

[a)]Electronic mail: alex_supin@sevin.ru
[b)]Electronic mail: nachtiga@hawaii.edu

 0001-4966/2007/121(1)/618/8/$23.00

tance increase, the echo intensity decreases, making the response smaller, but at the same time the echo delay increases and the response releases from masking, thus becoming larger. As a result, the auditory response to the echo remains almost constant.

It should be stressed, however, that the measurements of the evoked-response recovery in a double-pulse stimulation paradigm (Popov and Supin, 1990) have been done with the bottlenose dolphin (*Tursiops truncatus*) while the evoked-response investigations of echolocation have been completed with the false killer whale *Pseudorca crassidens* (Supin *et al.*, 2004, 2005). The temporal resolution of hearing of these different species may differ; therefore, the rate of recovery from the forward masking may also differ. In order to either validate or disprove the hypothesis of the forward masking as a mechanism of automatic gain control, it is useful to measure the evoked-response recovery within double-pulse stimulation passive listening conditions in the same species that was used to measure echolocation-related evoked responses. Therefore, the goal of the present study was a quantitative investigation of recovery of auditory brainstem responses in a double-pulse stimulation paradigm in a false killer whale *Pseudorca crassidens*.

II. MATERIALS AND METHODS

A. Subject and experimental conditions

The experiments were carried out in facilities of the Hawaii Institute of Marine Biology, Marine Mammal Research Program. The subject was a false killer whale *Pseudorca crassidens*, an approximately 30 year-old female kept in a wire-net enclosure in Kaneohe Bay, Hawaii. It was the same animal used in previous evoked-potential studies of echolocation (Supin *et al.*, 2003, 2004, 2005, 2006). Audiometric investigations of the subject have shown some high-frequency hearing loss: The best sensitivity around 20 kHz and upper frequency limit above 45 kHz (Yuen *et al.*, 2005). The animal was trained to accept latex suction cups containing EEG electrodes to pick up the evoked potentials and to keep a position in a hoop station while listening to a sound source.

The experimental facilities were laid out as follows (Fig. 1). The experimental enclosure was constructed of a floating pen frame (1), 8×10 m in size, supported by floats and bearing an enclosing wire net. The enclosure was supplied with a hoop station (2). At a distance of 2 m in front of the hoop station, there was a sound-transmitting transducer (3). The training experimenter kept a position (4) to give instructions to the animal and to reward it with fish for correct performance. The electronic equipment and the operating experimenter were housed in a shack (5).

B. Experimental procedure

Each session began with the experimenter attaching suction-cup electrodes to record the AEPs. In each trial, the animal was sent to a hoop station. As soon as the animal held the proper position, stimuli (sound pulses) began to be played through the transducer, and the AEPs to these stimuli were collected. The animal was required to stay in the hoop for approximately 1 min while the AEP collection was completed. After that, the animal was called back to the training experimenter to receive fish rewards.

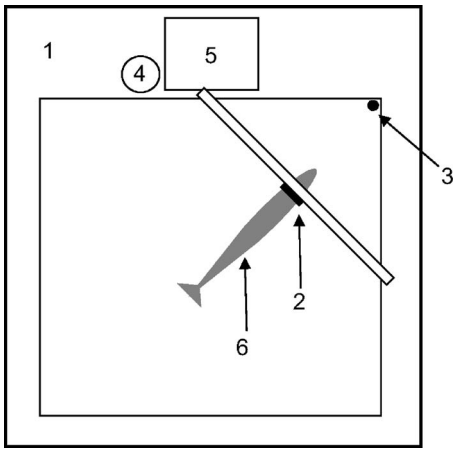


FIG. 1. Experimental design. 1—experimental enclosure, 2—hoop station, 3—transducer, 4—trainer's position, 5—operator's shack, 6—animal.

C. Instrumentation and data collection

The recording equipment was designed as shown in Fig. 2. Brain potentials were picked up from the subject (1) by EEG electrodes (2) which were gold-plated disks 10 mm in diameter mounted within rubber suction cups 60 mm in diameter. The active electrode was attached with conductive gel at the dorsal head surface, at the midline, 5–7 cm behind the blowhole. The reference electrode was also attached along with conductive gel on the animal's back near the dorsal fin. Brain potentials were carried by shielded cables to a balanced EEG amplifier (3) and amplified by 2.5×10^4 within a frequency range from 200 to 5000 Hz. The amplified signal was monitored by an oscilloscope Tektronix TDS1002 (4) and entered into a 12-bit analog-to-digital con-

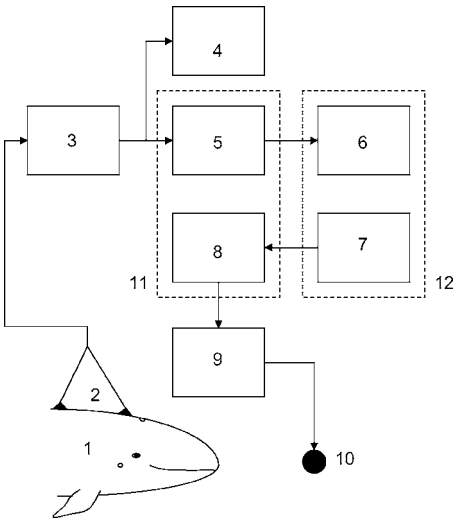


FIG. 2. Instrumentation. 1—animal, 2—suction-cup electrodes, 3—EEG amplifier, 4—oscilloscope, 5—A/D converter, 6—averager and data storage, 7—digital stimulus generator, 8—D/A converter, 9—sound power amplifier and attenuator, 10—transducer, 11—NI DAQ-6062E card, 12—laptop computer.

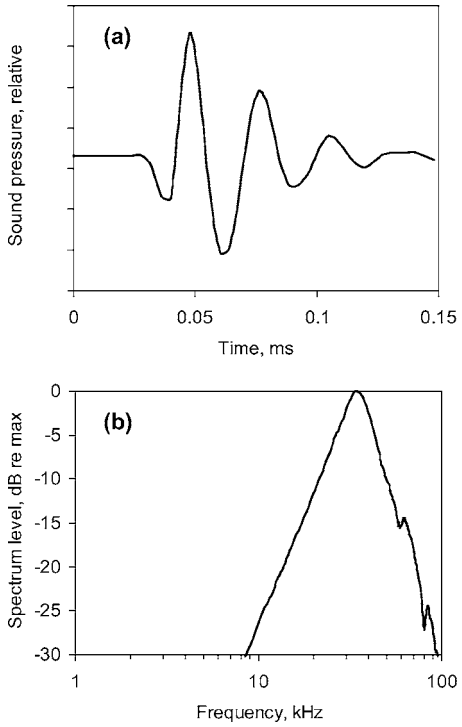


FIG. 3. Stimulating click waveform (a) and frequency spectrum (b).

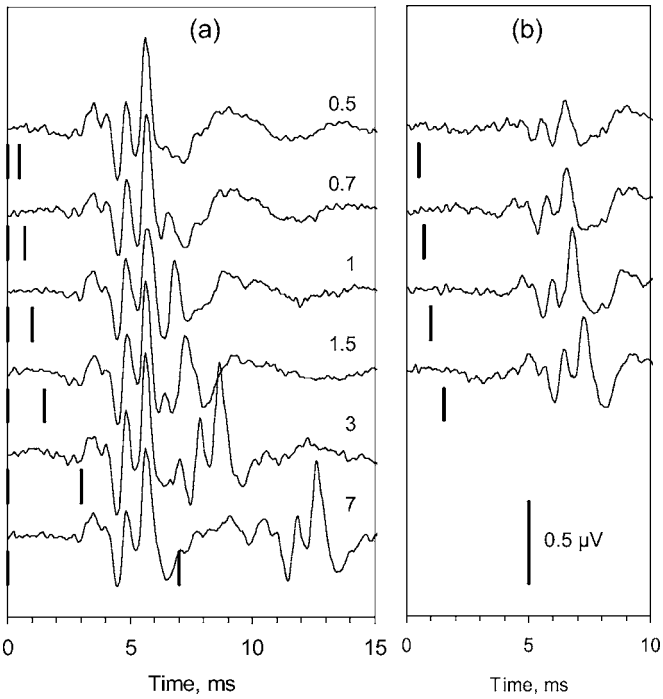


FIG. 4. (a) ABRs to double clicks of equal levels (170 dB re 1 μPa). ICI is indicated near records. (b) ABR to the test response extracted from corresponding records in (a) by subtraction procedure. Vertical ticks mark click instants, both clicks in (a) and the test click in (b).

verter (5) of a data acquisition card DAQ-6062E, National Instruments (11) installed in a standard laptop computer.

Data acquisition was controlled by a custom-made program based on LabVIEW software (National Instruments). To extract low-amplitude AEPs from background brain-wave noise, on-line averaging was used. The averaging was triggered by the external stimuli presented at a rate of 20/s. The acquisition window varied from 10 to 60 ms, depending on the inter-pulse interval; the sampling rate was 25 kHz. AEPs were collected by averaging 1000 individual records and stored in memory (6) of the computer (12).

Sound stimuli were digitally generated (7) by the same card and played through a 12-bit digital-to-analog converter (8), custom-made power amplifier-attenuator with a passband up to 1 MHz (9), and ITC-1032 (International Transducer Corporation) spherical transducer (10). The stimuli were single or double clicks produced by activation of the transducer by 7 μs rectangular pulses. The activation of this particular transducer produced acoustic pulses shown in Fig. 3. The acoustic pulse waveform and corresponding spectrum was similar to the biosonar pulses of the experimental subject as described by Supin *et al.* (2006), and the spectrum matched well the hearing range as described by Yuen *et al.* (2005). The inter-pulse interval varied from 0.5 to 50 ms between pulses when double pulse stimuli were presented. The ratio of amplitudes of the two pulses in a pair varied from 0 dB (equal amplitudes of the pulses) to 40 dB (the first pulse stronger than the second one). With the use of the hardware attenuator, the amplitude of each of the pulses could vary independently from 180 dB re 1 μPa peak-to-peak to values such that the response disappeared.

III. RESULTS

A. ABR waveform and features

Similar to the ABRs described in many other odontocetes (review by Supin *et al.*, 2001), the ABR in the false killer whale was a complex of a few positive-negative waves lasting 0.5–1 ms each. The onset latency of the response was around 2.5 ms, however, the most prominent waves occupied a time range from around 4 to 7 ms after each click (Fig. 4). At high stimulus intensities, their peak-to-peak amplitudes approached 1 μV.

When two clicks were presented with an interval as short as a few ms (or fractions of ms), the response to the second click was markedly reduced as compared to the first one or a single click presented alone. We can safely assume that the reduction of the second response was caused by the influence of the preceding stimulus. This is a standard phenomenon in evoked response investigations and the first stimulus in a pair is usually termed the *conditioning* stimulus and the second one is called the *test* stimulus; respectively, the two responses are named as *conditioning* and *test* responses.

Figure 4(a) exemplifies responses to paired clicks of equal intensities of 170 dB. Since each response was a few ms long, two responses partially overlapped at intervals shorter than 3 to 4 ms. To extract the real waveform of the test response, the waveform of the response to the single click was subtracted from that to the pair of clicks; the results present the extracted response to the test click [Fig. 4(b)]. The absence of a response to the conditioning click in the difference curves confirmed validity of the subtraction procedure. Waveform and latency of the obtained response to

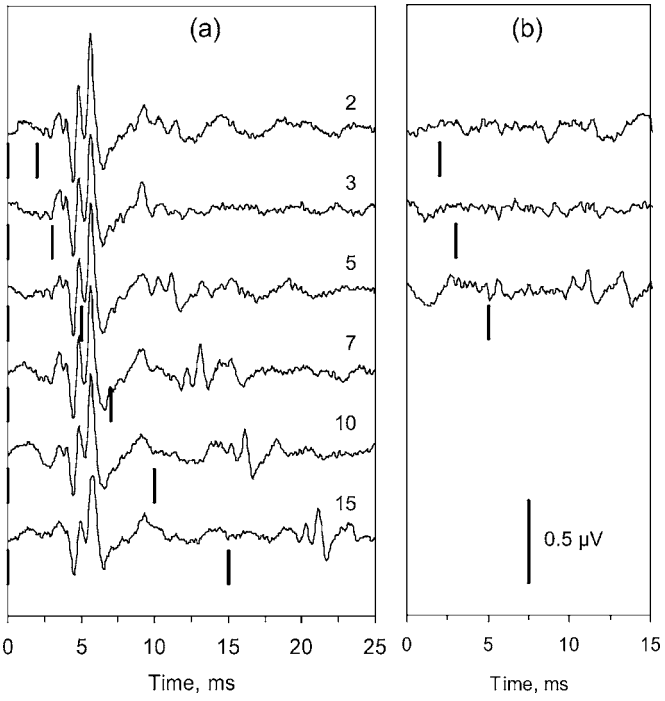


FIG. 5. ABRs to double clicks of different levels: Conditioning click level 170 dB re 1 μPa, test click level 150 dB re 1 μPa. Designations are the same as in Fig. 4.

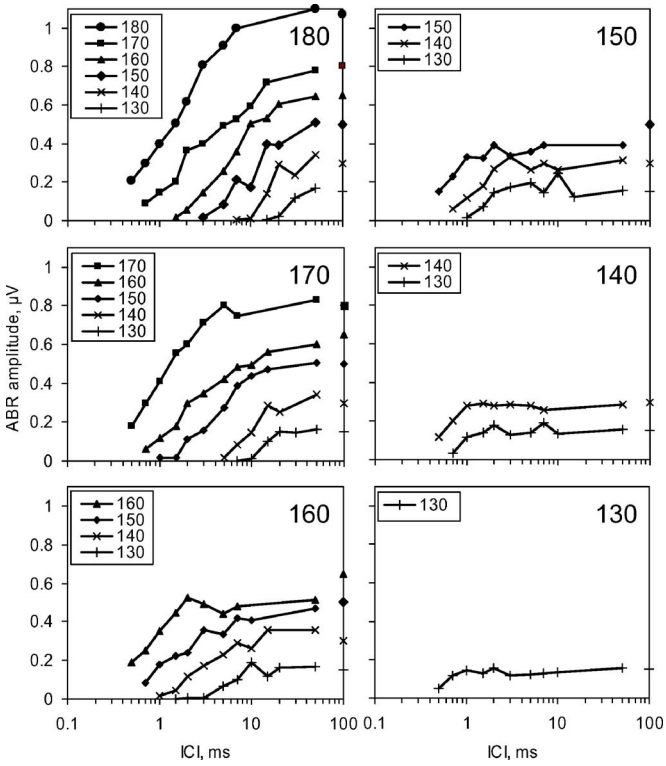


FIG. 6. ABR recovery functions (amplitude-vs-ICI) at various conditioning and test click levels. In each panel, the test click level (dB re 1 μPa) is taken as a parameter; conditioning click level (also dB re 1 μPa) is specified in each panel. Symbols at the right edge of the panel represent 100% recovery (ABR amplitude to a single test click of corresponding levels).

the test click were almost the same as those to a single click, but the response amplitude was reduced: The shorter the interval the smaller the amplitude of the test response.

The rate of recovery of the test response from suppression by the conditioning click strongly depended on the ratio of the intensities of the two clicks. Figure 5 exemplifies responses to paired clicks when the conditioning click was markedly more intensive (170 dB) then the test click (150 dB). A comparison shows that not only was the amplitude of the test response smaller but also its recovery time was much longer at unequal than at equal click levels. With equal click levels, the test response was well detectable at an interval as short as 0.5 ms and reached almost maximum amplitude (i.e., the compete recovery) at intervals longer than 3 ms (Fig. 4). With clicks of unequal levels, the test response became detectable at intervals longer than 3 ms and reached the maximum amplitude (the complete recovery) at 10–15 ms (Fig. 5).

B. ABR recovery functions at different intensities of the two clicks

Using records similar to those exemplified in Figs. 4 and 5, the peak-to-peak amplitude of the test response was measured as a function of the inter-stimulus interval at various intensities of both conditioning and test clicks. Measurements were made with conditioning click intensities of 180, 170, 160, 150, 140, and 130 dB; at 120 dB the response was undetectable. At each of the conditioning click intensities, the test click intensity varied by 10 dB steps below the conditioning click intensity. Thus, overall 20 combinations of conditioning and test click intensities were tested. At each of the combinations of conditioning and test click intensities, measurements were made at inter-click intervals varied by 1.4 to 1.5-fold steps as follows: 0.5, 0.7, 1, 1.5, 2, 3, 5, 7, 10, 15, 20, 30, 50 ms.

The results of the measurements are summarized in Figs. 6 and 7 where test response amplitude is plotted as a function of inter-click interval (ICI). For better illustration, the recovery (amplitude-vs-interval) functions are presented in two ways. In Fig. 6, the intensity of the test click is taken as a parameter specified in the legends whereas different intensities of the conditioning click (180–130 dB) are presented in different panels. In Fig. 7, the intensity of the conditioning click is taken as a parameter; different intensities of the test click (also 180–130 dB) are presented in different panels. All the recovery functions shared some common properties:

(i) All the functions consisted of two branches: An oblique and a flat one; the flat branch represented complete recovery, i.e., the response amplitude approached that of a single test click in the absence of a conditioning click.

(ii) The level of the flat branch, i.e., the amplitude of completely recovered response, was intensity dependent: The higher intensity of the test click, the higher amplitude. The rate of this dependence was around 0.015 μV/dB.

(iii) The slope of the oblique branch of all the functions was similar and equal to 0.6–0.8 μV per ICI decade. Therefore, the length of the oblique branch was intensity-dependent: The higher intensity of the test click, the longer the oblique branch.

(iv) At equal intensities of the conditioning and test clicks (presented by the leftmost plot of each panel in both

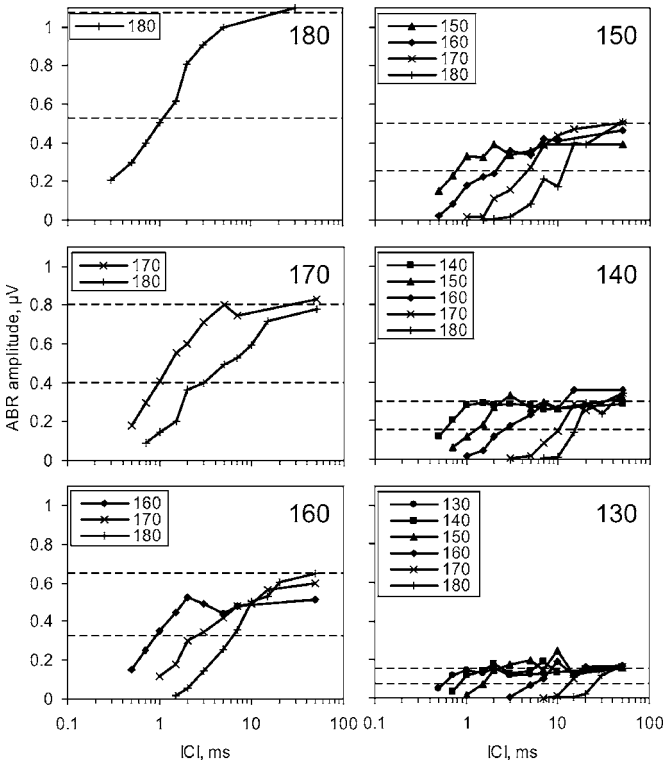


FIG. 7. The same ABR recovery functions as in Fig. 6, but conditioning click level is taken as a parameter and test click level is specified in each panel. Dashed straight lines show 100% and 50% recovery.

Figs. 6 and 7), the recovery of the test response was very quick: A detectable test response appeared at ICI less than 0.5 ms. With ICI increase, AEP amplitude increased with a rate of 0.6–0.8 μV/decade, thus reached its maximum at IVI from around 1 ms (at low conditioning and test intensities) to 10 ms (at high intensities).

(v) The position of recovery functions on ICI axis depended on the ratio of conditioning-to-test click intensities: The larger the ratio, the longer the recovery time. This dependence was observed both when test-click intensity was varied keeping the conditioning-click intensity constant (Fig. 6) and when conditioning-click intensity was varied keeping the test intensity constant (Fig. 7). In both cases, the rate of amplitude-vs-ISI dependence was 25–30 dB per ICI decade.

The dependence of recovery time on conditioning-to-test intensity ratio is presented quantitatively in Fig. 8. The plots show which ICI resulted in test responses recovery to an amplitude of 0.1 μV (a), 0.2 μV (b), and 0.3 μV (c). The plotted ICI values were calculated by linear interpolation between ICIs resulting in ABR amplitude below and above the specified amplitude, i.e., 0.1, 0.2, or 0.3 μV; if ABR amplitude was higher than 0.1 μV at ITC of 0.5 ms, the 0.1 μV recovery time was calculated by linear extrapolation of the 0.5–0.7 ms segment. All the plots show that (i) all the functions can be well approximated by straight regression lines when click intensity is presented on a dB scale and recovery time on log scale; (ii) all the plots feature rather similar slopes: 0.039–0.041 time decade/dB. Being expressed in a more commonly used inversed measure (dB/decade), the time-vs-intensity trade was 25 dB/decade.

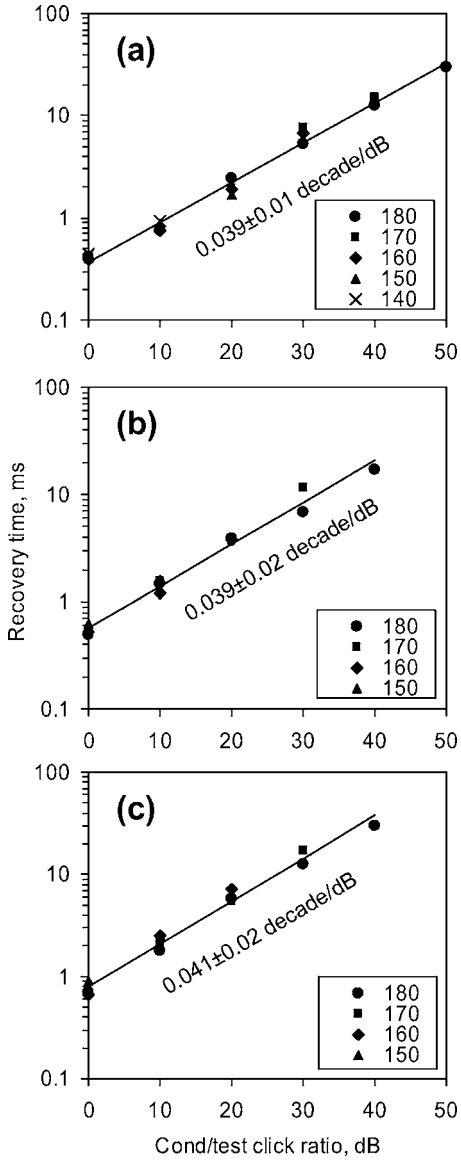


FIG. 8. ABR recovery time to a specified amplitude: 0.1 μV (a), 0.2 μV (b), and 0.3 μV (c) as a function of the conditioning/test click level (dB), keeping the conditioning click level (dB re 1 μPa) as a parameter specified in the legends. The slope (decade/dB) ± standard error is indicated near each plot.

IV. DISCUSSION

A. Temporal resolution of hearing revealed by the double-click test

The data presented above confirmed a high temporal resolution of the auditory system of odontocetes. This high temporal resolution of hearing in odontocetes has been demonstrated in many behavioral (Moore *et al.*, 1984; Au *et al.*, 1988; Au, 1990; Dubrovskiy, 1990) and electrophysiological (Supin and Popov, 1985, 1995a, b; Popov and Supin, 1990, 1997, 1998) investigations. One of manifestations of the high temporal resolution is a quick recovery of the test response in the double-click test (Popov and Supin, 1990; Supin and Popov, 1995b). The same was observed in the experiments described here with the false killer whale. When both the conditioning and test click were of equal intensities, complete recovery of the test response required a time as short as

a few milliseconds depending on intensity. When the conditioning click was more intense than the test click, the recovery time increased but still remained rather short, not more than a few tens of milliseconds with a conditioning-to-test intensity ratio as high as 50 dB.

However, the temporal resolution in the investigated false killer whale subject, although rather high, seems to be a little lower than that in previously investigated in bottlenose dolphins. Indeed, in a bottlenose dolphin, the 50% recovery time was around 2 ms at a conditioning-to-test ratio of 20 dB, around 7 ms at 40 dB, and up to 20 ms at 60 dB (Popov and Supin, 1990). In the false killer whale, the recovery was a little slower: 50% recovery time at 40 dB ratio was around 15 ms (see Figs. 6 and 7, 170/130 and 180/140 dB conditioning-test combination), i.e., about twice longer than in the bottlenose dolphin within similar stimulation conditions. Since only one subject was available for the present study, it remains to be shown whether this difference is interspecies or inter-individual.

B. Implication to biosonar automatic gain control

The paradigm of double-click stimulation may be considered as a simulation of events appearing during echolocation. Outgoing echolocation pulses and their returning echoes may be considered as clicks. The animal hears its own emitted click and shortly after that an echo click. This situation may be simulated by double-click stimulation. The intensive conditioning (transmitted) click is expected to produce a forward masking of the test (echo) response. However, instead of disrupting biosonar performance, this masking may make the responses to a target invariable to the target distance. As distance increases, the decrease of echo intensity and release from masking compensate one another, thus keeping the echo-response magnitude almost constant. Previous investigations in bottlenose dolphins (Popov and Supin, 1990) demonstrated that this sort of compensation is at its best when the echo intensity decreases with distance at a rate around 35 dB/decade. For the false killer whale investigated in this study, the best compensation is possible at a rate of around 25 dB/decade. Both of these values are within a range of possible rates, from 20 to 40 dB/decade, depending on the properties of the targets.

In the present study, the recovery of responses was measured with a wide range of intensities of both the conditioning and test clicks. These data allow us to compute the range of target strengths and distances in which the forward-masking mechanism might be capable of maintaining a relatively constant response amplitude independent of distance. For modeling the echolocation based on double-pulse stimulation data, we have to introduce two concepts: "*self-hearable level*" and "*internal attenuation*" of the emitted click. We use the term "self-hearable level" to specify how effectively the emitted click stimulates the ears of the animal itself. We define it as the level of an external (in the free field) click producing the same auditory response magnitude. The term "internal attenuation" is used herein for a difference between the source level and self-hearable level of the emitted click. This shows how effectively the acoustic energy is concentrated in the sonar beam and isolated from the animal's ears.

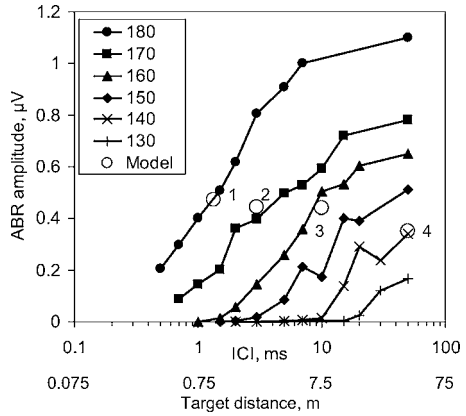


FIG. 9. An example of modeling of self-masking in sonar using double-click stimulation data. Double scale of the abscissa present ICI and target distance corresponding to this ICI. The plots represent ABR dependence on ICI at a conditioning click level of 180 dB re 1 μPa and test click levels from 180 to 130 dB, as specified in the legend [see Fig. 6(a)]. Points 1–4 (open circles) represent ABR amplitudes at a few ICIs of 1.33, 3, 10, and 50 ms (corresponding to target distances of 1, 2.25, 7.5, and 37.5 m) calculated on assumption of echo level decrease with distance at a rate of 25 dB/decade.

With these definitions, the parameters of double-click stimuli were used to model sonar parameters as follows:

- The self-hearable level of the emitted click was modeled by conditioning-click level;
- the source level of the emitted click was modeled as a sum of conditioning-click level and a constant of internal attenuation;
- the echo level was modeled by the test-click level;
- the target distance (in meters) was modeled by ICI (in milliseconds) multiplied by a factor of 0.75 m/ms (two-way sound spread at the 1500 m/s speed of sound in water).

Using these definitions, and taking various combinations of conditioning click level (imitating the self-hearable level of the emitted sonar click) and conditioning-to-test click level difference at a delay of 1.33 ms (imitating the echo level at a target distance of 1 m, i.e., the target strength) we calculated the test-click intensity (imitating the echo intensity) and found the test (echo) ABR amplitude as a function of ICI (imitating the target distance).

The calculation procedure is exemplified in Fig. 9. We assumed a rate of echo attenuation with distance of 25 dB/decade, which is close to the slope of majority of the plots in Fig. 8. For this particular example, we assume that the source level of the emitted click is 180 dB of a self-hearable level which is imitated by 180 dB conditioning click. Then we assume that the target strength is 0 dB re internal attenuation which is imitated by equal levels of the conditioning and test click (180 dB) at a target distance of 1 m which is the standard to specify target strength. The distance of 1 m is imitated by ICI of 1.33 ms. Interpolation between points of 1 and 1.5 ms of the plot for 180 dB test click level gives ABR amplitude of 0.47 μV (point 1 in Fig.

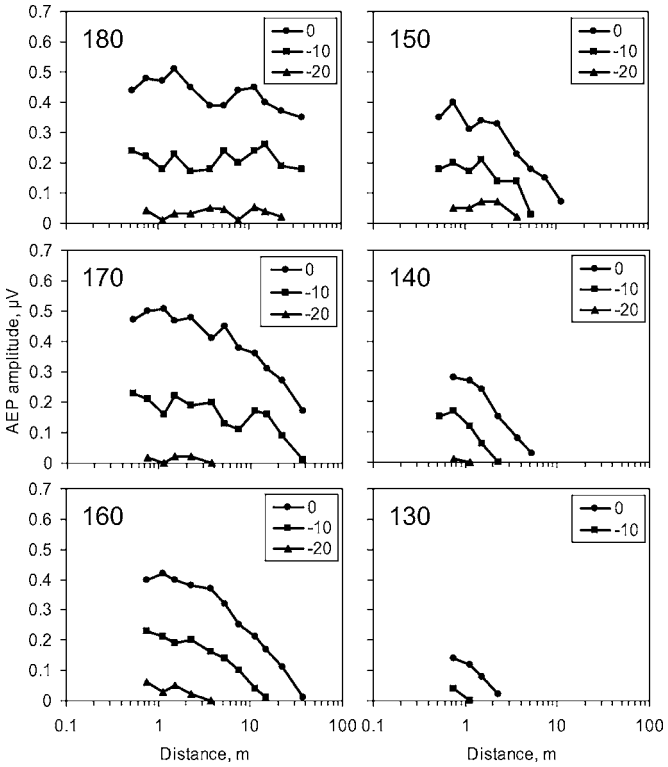


FIG. 10. Results of modeling of self-masking in sonar using double-click stimulation data. ABR amplitude to an echo modeled by test stimulus is plotted as a function of target distance modeled by ICI. In each panel, target strength (0 to −20 dB relative internal attenuation) as modeled by test-to-conditioning click level ratio is taken as a parameter. Emitted click level (180–130 dB self-hearable level) as modeled by conditioning click level is specified for each of the panels.

9). Then we take any other target distance, say 2.25 m which is imitated by ICI of 3 ms. With the rate of 25 dB/decade, the attenuation at this distance is −8.8 dB, thus the echo is imitated by a test-click level of 171.2 dB. An interpolation between plots for 170 and 180 dB at ICI of 3 ms results in ABR amplitude of 0.45 μV (point 2). Similarly, a target distance of 7.5 m (ICI of 10 ms, attenuation of 18.1 dB) results in ABR amplitude of 0.44 μV (point 3), a target distance of 37.5 m (ICI of 50 ms, attenuation of 39.4 dB) results in ABR amplitude of 0.35 μV (point 4), etc.

The test-response amplitudes found by the described procedure are presented in Fig. 10 as functions of simulated target distance at various combinations of simulated levels of the emitted pulse and target strengths. The plots show that at rather high intensity of emitted clicks (with a self-hearable level of 180 dB), neglecting a certain data scatter, the forward masking provides an independence of response amplitude with distance within a wide range, from 0.5 to 37.5 m (delays from 0.7 to 50 ms) in spite of 46 dB distant-dependent variation of simulated echo intensity. The response amplitude depends on the target strength: It diminished when the target strength decreased from 0 to −20 dB re internal attenuation. However, at all target strengths, the amplitude remained slightly dependent on the distance.

Decreasing the level of the simulated emitted pulse (self-hearable levels 170 dB, etc., down to 140 dB) little influenced the maximum response amplitude: It remained 0.4–0.5 μV at a target strength of 0 dB re internal attenuation, around 0.2 μV at −10 dB, and less than 0.1 μV at −20 dB. The independence of test-response amplitude on the intensity of the emitted clicks seems paradoxical at a first glance but this result is a direct consequence of the action of forward masking: Being intensity-dependent, the masking compensates for the intensity variation.

However, it was the range of constant test-response amplitude that really depended on the level of the emitted pulse. Being more than 38 m (50 ms) at a 180 dB self-hearable level, the constant-amplitude range decreases as the self-hearable level decreased from 170 to 150 dB, was almost absent at 140 dB, and definitely absent at 130 dB. This result is also easily understandable. To be a subject of regulation by masking, the response must be large enough in nonmasked conditions; therefore, when both the emitted click and echo are weak, the echo-response plainly decreases with distance because of decreasing echo intensity. This limitation agrees very well with the fact that whales and dolphins increase the level of their emitted clicks with increasing distance to a target as has been demonstrated in echolocating wild dolphins (Rasmussen *et al.*, 2002; Au and Benoit-Bird, 2003; Au and Herzing, 2003; Au and Würsig, 2004; Au *et al.*, 2004). If they did not increase their outgoing click intensities the echoes would be too weak to perceive them even though they were completely released from masking.

Another limitation revealed by the modeling described above concerns the target strength. As Fig. 9 shows, the test response approached zero when the target strength was 20 dB below the internal attenuation. Does this prediction agree with the real abilities of the odontocete sonar? It has been shown using the same subject as that in the present study that the internal attenuation may be as effective as −35 dB (Supin *et al.*, 2006). Thus, based on the data presented herein, it may be predicted that masking events allow the perception of echoes from targets with a strength as low as −55 dB. At lower target strengths, the echo responses should be masked at any source level of the emitted pulse and any distance to the target. This requirement fits the real conditions of echolocation. For example, the target strength of moderate-size prey fish was estimated at −20 to −35 dB (Benoit-Bird and Au, 2003).

Thus, the model of interaction of emitted and echo signal based on the double-click data (i) confirms the possibility of automatic gain control based on forward-masking in the odontocete sonar and (ii) allows us to make some predictions for the odontocete biosonar abilities. Further investigation should allow the opportunity to examine the validity of these predictions.

ACKNOWLEDGMENTS

This study was supported by the Office of Naval Research Grant Nos. N00014-98-1-0687 and N00014-05-1-0738, for which the authors thank Robert Gisiner. Support was also provided by the Russian Ministry of Science and Education Grant No. NSh-7117.2006.4. Work was conducted

under U.S. Marine Mammal Permit 978-1567 issued to Paul E. Nachtigall. This is contribution number 1250 of the Hawaii Institute of Marine Biology.

Au, W. W. L. (**1990**). "Target detection in noise by echolocating dolphins," in *Sensory Abilities of Cetaceans. Laboratory and Field Evidence*, edited by J. A. Thomas and R. A. Kastelein (Plenum, New York), pp. 203–216.

Au, W. W. L. (**1993**). *The Sonar of Dolphins* (Springer-Verlag New York).

Au, W. W. L., and Benoit-Bird, K. J. (**2003**). "Automatic gain control in the echolocation system of dolphins," Nature (London), **423**, 861–863.

Au, W. W. L., and Herzing, D. L. (**2003**). "Echolocation signals of wild Atlantic spotted dolphin (Stenella frontalis)," J. Acoust. Soc. Am. **113**, 598–604.

Au, W. W. L., Moore, P. W. B., and Pawloski, D. A. (**1988**). "Detection of complex echoes in noise by an echolocating dolphin," J. Acoust. Soc. Am. **83**, 662–668.

Au, W. W. L., and Würsig, B. (**2004**). "Echolocation signals of dusky dolphins (Lagenorhynchus obscurus) in Kaikoura, New Zealand," J. Acoust. Soc. Am. **115**, 2307–2313.

Benoit-Bird, K. J., and Au, W. W. L. (**2003**). "Acoustic backscattering by Hawaiian latjanid snappers. I. Target strength and swimbladder characteristics," J. Acoust. Soc. Am. **114**, 2757–2766.

Dubrovskiy, N. A. (**1990**). "On the two auditory systems in dolphins," in *Sensory Abilities of Cetaceans. Laboratory and Field Evidence*, edited by J. A. Thomas and R. A. Kastelein (Plenum, New York), pp. 233–254.

Moore, P. W. B., Hall, R. W., Friedl, W. A., and Nachtigall, P. E. (**1984**). "The critical interval in dolphin echolocation: What is it?" J. Acoust. Soc. Am. **76**, 314–317.

Nachtigall, P. E., and Moore, P. W. B. (**1988**). *Animal Sonar: Processes and Performance* (Plenum, New York).

Popov, V. V., and Supin, A. Ya. (**1990**). "Auditory brain stem responses in characterization of dolphin hearing," J. Comp. Physiol., A **166**, 385–393.

Popov, V. V., and Supin, A. Ya. (**1997**). "Detection of temporal gaps in noise in dolphins: Evoked-potential study," J. Acoust. Soc. Am. **102**, 1169–1176.

Popov, V. V., and Supin, A. Ya. (**1998**). "Auditory evoked responses to rhythmic sound pulses in dolphins," J. Comp. Physiol., A **183**, 519–524.

Rasmussen, M. H., Miller, L. A., and Au, W. W. L. (**2002**). "Source levels of clicks from free-ranging white beaked dolphins (Lagenorhinchus albirostris Gray 1846) recorded in Icelandic waters," J. Acoust. Soc. Am. **111**, 1122–1125.

Supin, A. Ya., Nachtigall, P. E., Au, W. W. L., and Breese, M. (**2003**). "Evoked potential recording during echolocation in a false killer whale Pseudorca crassidens (L)," J. Acoust. Soc. Am. **113**, 2408–2411.

Supin, A. Ya., Nachtigall, P. E., Au, W. W. L., and Breese, M. (**2004**). "The interaction of outgoing echolocation pulses and echoes in the false killer whale's auditory system: Evoked-potential study," J. Acoust. Soc. Am. **115**, 3218–3225.

Supin, A. Ya., Nachtigall, P. E., Au, W. W. L., and Breese, M. (**2005**). "Invariance of evoked-potential echo-responses to target strength and distance in an echolocating false killer whale," J. Acoust. Soc. Am. **117**, 3928–3935.

Supin, A. Ya., Nachtigall, P. E., and Breese, M. (**2006**). "Source-to-sensation level ratio of transmitted biosonar pulses in an echolocating false killer whale," J. Acoust. Soc. Am. **120**, 518–526.

Supin, A. Ya., and Popov, V. V. (**1985**). "Recovery cycles of the dolphin brainstem evoked potentials for paired acoustic stimuli," Dokl. Biol. Sci. **283**, 535–537.

Supin, A. Ya., and Popov, V. V. (**1995a**). "Envelope-following response and modulation transfer function in the dolphin's auditory system," Hear. Res. **92**, 38–46.

Supin, A. Ya., and Popov, V. V. (**1995b**). "Temporal resolution in the dolphin's auditory system revealed by double-click evoked potential study," J. Acoust. Soc. Am. **97**, 2586–2593.

Supin, A. Ya., Popov, V. V., and Mass, A. M. (**2001**). *The Sensory Physiology of Aquatic Mammals* (Kluwer, Dordrecht).

Thomas, J. A., Moss, C. F., and Vater, M. (**2003**). "*Echolocation in Bats and Dolphins*" (University of Chicago Press, Chicago-London).

Yuen, M. M. L., Nachtigall, P. E., Breese, M., and Supin, A. Ya. (**2005**). "Behavioral and auditory evoked potential audiograms of a false killer whale (Pseudorca crassidens)," J. Acoust. Soc. Am. **118**, 2688–2695.

Dolphin (*Tursiops truncatus*) echoic angular discrimination: Effects of object separation and complexity

Brian K. Branstetter[a)]
Psychology Department, University of Hawaii, Manoa, Honolulu, Hawaii, 96822-2294

Sonja J. Mevissen
The Dolphin Institute, 420 Ward Ave., Suite 212, Honolulu, Hawaii

Adam A. Pack and Louis M. Herman
Psychology Department, University of Hawaii, Manoa, Honolulu, Hawaii, 96822-2294 and The Dolphin Institute, 420 Ward Ave., Suite 212, Honolulu, Hawaii

Scott R. Roberts and Lea K. Carsrud
The Dolphin Institute, 420 Ward Ave., Suite 212, Honolulu, Hawaii

(Received 26 July 2006; revised 21 October 2006; accepted 24 October 2006)

A bottlenose dolphin was tested on its ability to echoically discriminate horizontal angular differences between arrays of vertically oriented air-filled PVC rods. The blindfolded dolphin was required to station in a submerged hoop 2 radial m from the stimuli and indicate if an array with two rods ($S+$) was to the right or the left of a single rod ($S-$). The angular separation between the two rods (θ_w) was held constant within each experiment while the angle between the S+ and the S− stimuli (θ_b) varied to produce angular differences ($\Delta\theta = \theta_b - \theta_w$) ranging from 0.25 to 4°. In experiment I, θ_w was maintained at 2° and in experiment II, θ_w was maintained at 4°. Resulting 75% correct thresholds (method of constant stimuli) were 1.5 and 0.7°, respectively. The two main findings of this study are: (1) decreasing the number of targets does not aid in localization, and (2) increasing the space between the rods enhances localization. Taken as a whole, the experiments suggest dolphins have a well-developed ability to resolve spatial information through sonar. [DOI: 10.1121/1.2400664]

PACS number(s): 43.80.Lb, 43.66.Qp, 43.66.Pn, 43.66.Ba [WWA] Pages: 626–635

I. INTRODUCTION

The dolphin's ability to echoically resolve fine spatial details from an object will be dependent at least upon the physical characteristics of the object, the characteristics of the dolphin's sonar signal, the dolphin's auditory signal processing of the returned echoes, and the dolphin's ability to localize echoes from the various reflecting object features. The focus of the present study is to determine the limit of the dolphin's echolocation spatial acuity in the horizontal plane.

A. Background

The incident sonar signal from a bottlenose dolphin[1] can be described as a broadband transient with peak frequencies between 40 and 140 kHz, durations as short as 40 μs, and peak-to-peak sound pressure levels often exceeding 220 dB re: 1 μPa (Au, 1993). The concave shape of the dolphin's skull, coupled with the focusing properties of the lipid melon, focus the signal forward in a tight beam, elevated about 5° above the rostrum, with a 3 dB beam width of approximately 10° (Au, 1980). Consequently, targets directly forward of the melon will be highly ensonified, increasing the signal-to-noise ratio relative to peripheral targets. The high frequency, short duration properties of individual clicks contribute to the dolphin's ability to resolve fine spatial details. Higher frequencies allow for echo returns from small targets, provided the target's spatial extent is greater than the wavelength of the incident signal. High frequency components also allow dolphins to detect small structural features within a target that may facilitate target identification. Short duration clicks can improve range resolution by limiting overlapping echoes from closely spaced objects in the same line of acoustic propagation. In addition, fine temporal resolution, demonstrated by an auditory critical interval of 264 μs (Moore *et al.*, 1984) allows the dolphin to resolve echoes occurring in close temporal proximity.

Despite the challenges imposed by an aquatic environment (e.g., sound traveling approximately 4.5 times faster in water than in air), several of the standard terrestrial mammalian localization mechanisms also appear to be well developed in dolphins. When a sound source originates off the midsagital plane, a differential distance from each ear to the sound source is produced. The differential distance results in arrival time differences between the two ears known as interaural temporal differences (ITDs). A differential distance between the ears, coupled with sound shadowing by the head, will also produce an intensity difference between the ears known as interaural intensity differences (IIDs). Moore *et al.* (1995) investigated the dolphin's ability to utilize ITDs and IIDs. Jaw phones (i.e., hydrophones embedded in rubber suction cups and attached to the right and left lower jaws)

[a)]Author to whom correspondence should be addressed. Electronic mail: branstet@hawaii.edu

were used to provide precise control over binaural stimulus presentation to the dolphin. ITD thresholds were as small as 7 μs for a 30 kHz pulse. For sonar-type pulses (with higher peak frequencies above 60 kHz) ITDs were between 17 and 18 μs. IID thresholds were below 1.0 dB for all frequencies tested. Thus, both ITDs and IIDs appear to be very salient cues a dolphin could employ to localize sound sources.

Recently, Branstetter *et al.* (2003) investigated a dolphin's ability to echoically resolve angles between multiple targets in the horizontal plane. A blindfolded dolphin positioned its head within a stationary, vertically oriented hoop, 2 radial m from the stimuli, and then indicated whether an array with four rods ($S+$) was to the left or the right of an array with two rods ($S-$), by pressing a corresponding paddle. The angular separation between the rods within each array (θ_w) was maintained at 2° but the angular separation between the two arrays (θ_b) was varied to produce angular differences ($\Delta\theta = \theta_b - \theta_w$) ranging between 0.25 and 4°. A modified method of constant stimuli, used to test for angular discrimination ability, yielded a psychometric function with a 75% correct threshold of 1.6° (see Fig. 8). The results were consistent with passive hearing studies that determined the smallest possible angle (minimum audible angle or MAA) between two sound sources that would allow the two sources to be perceived by the dolphin as discrete. Renaud and Popper (1975) reported the horizontal MAA for 30, 60, and 90 kHz pure tones as 2.5, 3.0, and 3.0°, respectively. Horizontal and vertical MAAs for click stimuli (35 μs duration, 64.35 kHz peak frequency) were 0.9 and 0.7°, respectively.

The results from Branstetter *et al.* (2003), suggested that the dolphin's echoic spatial acuity was sufficient to support cross-modal matching of objects between vision and echolocation as shown by Pack and Herman (1995) and Herman *et al.* (1998). Because the dolphin in these studies interrogated relatively large objects (approximately 43 cm^2) at relatively short distances (<1 m), the angle the objects subtended, and the angles between object components, were much greater than the angular differences threshold of 1.6°. Hence, the objects and object features used in the cross-modal studies should be resolvable through the dolphin's echolocation sense.

The current study builds on the findings of Branstetter *et al.* (2003). One goal of Branstetter *et al.* was to generalize the findings to the cross-modal paradigm (Pack *et al.*, 2002). If the dolphin can echoically resolve fine spatial details from multiple echoes in the angular discrimination experiment (Branstetter *et al.* (2003)), the dolphin can probably echoically resolve fine details in the cross-modal experiments. Because the objects in the cross-modal experiments (Pack *et al.*, 2004) were three dimensional with multiple features that produced multiple echoes, the stimuli in Branstetter *et al.* (2003) were composed of multiple targets (four rods vs two rods) thus producing multiple echoes as well. However, the resulting complex stimuli may not have been optimal to produce the smallest echoic angular discrimination threshold. In humans, reducing the number of distracting sound sources can aid in localization (Wightman and Kistler, 1997). We employed this tactic in Experiment 1 of the current study by reducing the $S+$ and $S-$ stimuli to two rods and one rod, respectively.

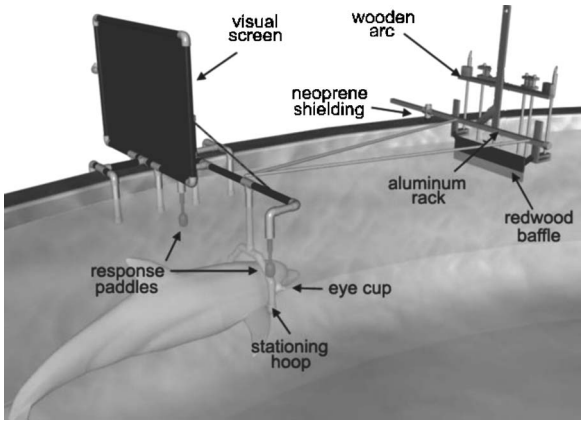


FIG. 1. Experimental apparatus with dolphin in underwater stationing hoop.

In the Branstetter *et al.* (2003), experiment, the angular difference between the rods within each array (θ_w) was always 2°. The proportion correct at an angular difference of 2° was 83% compared to 93% at 4°. This suggests when $\theta_w = 2°$, the perceived positions of the rods were subject to spatial blurring. The echoes from each rod could have effectively masked each other. In humans, an improvement in the ability to localize sound sources occurs due to increased spatial separation (Langendijk *et al.*, 2001). We adopted this strategy in Experiment II by increasing θ_w to 4°. If all the rods were more fully resolvable, perhaps the dolphin could make better decisions about the rod positions relative to one another.

II. GENERAL METHODS

A. Subject

The subject for the study was a 16-year-old, male, Atlantic bottlenose dolphin (*Tursiops truncatus*) named Hiapo. Hiapo was the same dolphin used in the Branstetter *et al.* (2003) study. Hiapo was housed in two interconnected seawater tanks (each circular tank was 1.8 m deep with a diameter of 15.2 m) with three female dolphins at the Kewalo Basin Marine Mammal Laboratory in Honolulu, Hawaii. During the experimental procedures, Hiapo was maintained alone or with one other dolphin. Approximately 11.8 kg of herring, capelin, smelt and squid were fed to Hiapo daily, a portion of which (2.9 kg) was fed during the experiment.

B. Apparatus

The same experimental apparatus (see Fig. 1) employed from Branstetter *et al.* (2003) was employed in all of the experiments presented here. A complete description can be found in Branstetter *et al.* (2003). A detailed summary is presented here.

The dolphin was trained to wear latex rubber suction cups (eye cups), over its eyes to prevent the use of vision during the experiment. A submerged, vertically oriented, hollow, PVC stationing hoop was attached to the tank wall. The hoop was submerged approximately 55 cm below the sur-

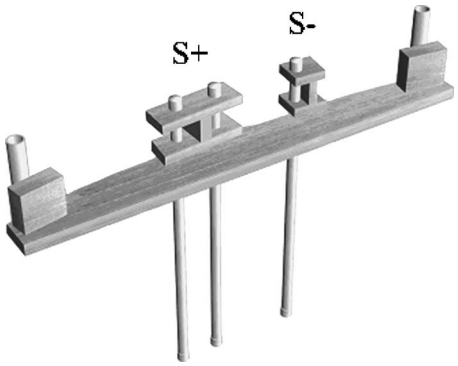


FIG. 2. PVC rods on wooden rack showing *S*+ and *S*− stimuli.

face. Two submerged response paddles were connected to the periphery of the stationing hoop allowing the dolphin to exit the hoop and respond to either the left or right paddle. The stimuli were presented using a customized aluminum rack, which was positioned 2 m from the tip of the dolphin's rostrum. A redwood baffle at the bottom of the rack occluded inspection (visually or echoically) of the tips of the stimuli when they were positioned on the rack prior to full immersion in the water.

The experimental stimuli consisted of vertically oriented, schedule 40, PVC rods (Fig. 2). Each rod was air filled to provide a high impedance mismatch to water making them highly reflective to dolphin sonar signals. In all experiments, the positive stimulus to respond to (*S*+ stimulus) consisted of two rods and the negative stimulus not to respond to (*S*− stimulus) consisted of one rod. Each PVC rod was 1.9 cm in outer diameter, 1.7 cm in inner diameter and 75 cm in length. PVC caps were placed on each end of each rod (2.7 cm outer diameter, 2.3 cm length) to ensure they were airtight. A total of 12 PVC rods with caps were used to construct three *S*+ stimuli and three *S*− stimuli. Each rod and cap was constructed from the same type of PVC with the same batch number ensuring the material composition was nearly identical. However, as a precaution, before each session, each rod and each cap were randomly assigned to an *S*+ or an *S*− array to control for the possibility that Hiapo could learn to identify a particular rod artifact associated with a particular array.

The rods in *S*+ stimulus were separated from each other by θ_w° from a 220 cm distance. This distance represented the radial distance from the arrays to the midpoint between Hiapo's mandibular windows on his lower jaw. An *S*+ and an *S*− stimulus were positioned on wooden arcs with the angular separation (θ_b) between the stimuli predetermined from a counterbalanced schedule. The angular difference can be defined as

$$\Delta_\theta = \theta_b - \theta_w,$$

where Δ_θ is the angular difference, θ_b is the separation between the *S*+ and *S*− arrays (separation between the closest rods) and θ_w is the angle between the rods within *S*+ array (Fig. 3). The dolphin can only identify the correct position of the *S*+ stimulus by comparing the angles between the rods of the *S*+ stimulus, and the angle between both stimuli. Thus, $\Delta\theta$ is a measure of the dolphin's angular discrimination ability. The arrays were centered on the wooden arc so the far end of the *S*+ array and the far end of the *S*− array were an equal distance to the center of the arc and, hence, the dolphin's median plane (see the Appendix).

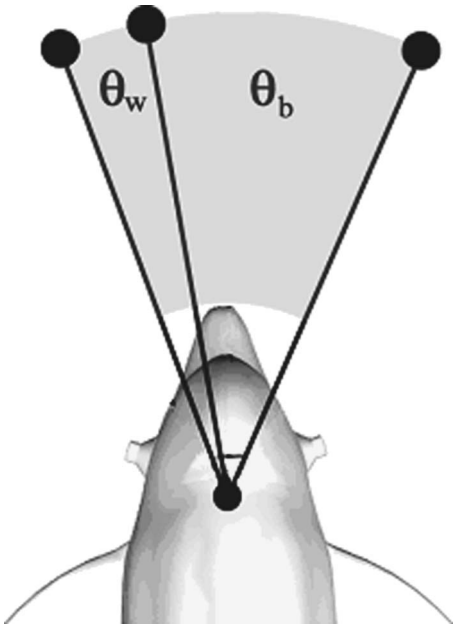


FIG. 3. Angular difference (Δ_θ) between the rods relative to the dolphin is calculated by subtracting θ_w from θ_b.

When Hiapo was positioned inside the stationing hoop as far as his pectoral fins would allow, each rod was approximately 200 radial cm from the tip of his rostrum, 220 radial cm from his mandibular window and 245.7 radial cm from the center of the stationing hoop. This allowed Hiapo to pivot his head in the horizontal plane along a 45.7 cm radial arc (measured from the center of the stationing hoop to the tip of his rostrum) while the stimuli maintained a distance of 200 cm from his rostrum tip. The wooden arc was attached to a pulley system on the aluminum rack that was attached to both the tank wall and a wooden stand on the exterior of the tank (Fig. 1). The rack and the pulley system provided control over stimuli exposure time by allowing an assistant to lower the arrays when prompted. When lowered, the only part of the array submerged was the bottom 55 cm of the six rods. The rack was leveled using a standard bubble level. A visual screen (Fig. 1) prevented the trainer from seeing the stimuli and inadvertently cueing the dolphin towards the correct response.

Sonar emissions were recorded with a hydrophone (LAB-core systems, PZ-1A) placed approximately 15 cm in front of the stationing hoop at a depth of approximately 15 cm. When Hiapo was properly stationed in the hoop, the hydrophone was directly above his head and slightly behind his blow hole. The sounds were digitized at 44.1 kHz sampling rate and recorded with a Sony digital video recorder. Because the sonar emissions were already distorted by the low sampling rate, near field distortion was not considered an issue. Hiapo would emit sonar signals prior to array exposure and, after his response, only the clicks between the onset of array exposure to when he backed out of the hoop (monitored from recorded video) were considered for analy-

sis. Backing out of the hoop was operationally defined as when his eye cups crossed the hoop threshold on his way out of the hoop. Click trains for each trial were then manually counted by two independent observers from a hardcopy printout. Both observers were blind to which trial condition they were reviewing.

Response time (RT) for each trial was also recorded by an observer with a digital stopwatch. RT was defined as the time between the onset of stimulus exposure to when the dolphin pressed a paddle. The trainer (who, because of the visual screen had no knowledge of the stimulus array) vocally reported a paddle press.

C. General procedure

At the beginning of each trial, the eye cups were placed on the dolphin. Hiapo remained at the trainer's station with his jaw held out of the water as an assistant then placed the stimuli on the rack and the tips of the PVC rods were gently submerged a few cm (<5 cm) below the water surface to prevent a potential passive acoustic splash cue. This reduced the possibility of Hiapo passively listening to the stimuli tips while they were placed in the water. In addition, a water-soaked redwood baffle, attached to the aluminum rack was positioned about 5 cm beneath the surface in front of the tips of the rods (Fig. 1). Water-soaked redwood is known for its ability to absorb sound underwater (see Johnson, 1967). The baffle served to reduce any sound the stimuli may have made upon entering the water and prevented the dolphin from echolocating on the tips of the stimuli before they were fully lowered. Once the stimuli were in place, an experimenter in an elevated tower overlooking the tank instructed the trainer to signal the dolphin to enter the hoop. The trainer manually assisted the dolphin to insure that he was positioned in the hoop up to his pectoral fins. The experimenter then verbally instructed the assistant to lower the arrays into the water. The dolphin was required to echolocate and identify the location (left or right) of the *S*+ array by backing out of the hoop and touching the corresponding left or right response paddle. The trainer verbally called out the dolphin's response, either "left" or "right," and an experimenter located on an elevated deck overlooking the pool, identified the choice as correct or incorrect, referring to the preplanned schedule. The trainer also functioned as a blind observer because the visual barrier prevented the trainer from seeing the stimuli and, thus, knowing the correct choice. Correct responses were rewarded with fish, and social reinforcement, followed by an inter-trial interval of approximately 35–40 s. The dolphin was called back to station for incorrect responses, did not receive a fish reward and proceeded directly into an inter-trial interval.

Two experiments were conducted.[2] Their presentation was counterbalanced (ABBA format) to control for any learning effects. Half of experiment I, experiment II were completed in consecutive order. Then, the remaining halves were conducted in reverse order (i.e., Exp. II, and then Exp. III). Prior to conducting these experiments, a Pilot study (Branstetter, 2005) was run to investigate how easily Hiapo would generalize from his initial arrays of four rods vs two rods (Branstetter *et al.*, 2003) to two rods vs one rod, and to determine which angular separations would capture the full psychometric function.

III. EXPERIMENT I

The goal of experiment I was to determine if echoically localizing targets is less difficult when the number of echoes received per incident signal is reduced. In humans, sound localizing performance progressively increases as the number of distracting sounds decreases (Wightman and Kistler, 1997; Langendijk *et al.*, 2001). In the current experiment, the number of potential echoes is reduced to three by decreasing the number of rods presented on each trial to three. If echo complexity, defined as the number of simultaneous echoes received, is a major factor influencing echoic angular discrimination, a threshold decrease is expected in the current experiment relative to the results of Branstetter *et al.* (2003).

A. Stimuli

The *S*+ stimulus was a two-rod array. The *S*− stimulus was a single rod (see Fig. 2). The angle between the rods within the *S*+ array (θ_w) was held at a constant 2° and the angle between the *S*+ and *S*− array (θ_b) varied to produce angular differences (Δ_θ) of 4.00, 3.00, 2.50, 2.25, 2.00, 1.75, 1.50, 1.25, 1.00, and 0.50°. Ten angular differences were tested compared to only eight in Branstetter *et al.* (2003) to provide greater resolution estimating the psychometric function.

B. Procedure

Each session began with six warm-up trials with an angular difference of 4°. The position of the *S*+ array (left or right) was randomized with an equal number of left and right presentations within each session. An approximate 25 s inter-trial interval followed each trial to allow assistants to prepare the next set of stimuli. If the dolphin responded incorrectly on two or more warm-up trials, the testing session was aborted until the next day. If the dolphin responded correctly on five or more warm-up trials ($p<0.05$, cumulative binomial), the test session began after a 1.5–2 min break. The modified method of constant stimuli was used for stimulus presentation. Angular differences were presented in descending order. Each of the ten angular differences was tested twice per session in two descending sweeps. The left or right position of the *S*+ stimulus was randomized for the first ten trials. The position of the *S*+ array in the second ten trials was the opposite of the first ten trials. For example, if the location of the *S*+ array for an angular separation of 2° occurred on the left within the first ten trials, the position of the *S*+ array for the same angular separation in the second ten trials would be to the right. This ensured that any potential right or left bias would affect angular differences equally. The last four trials were cooldown trials with an angular difference of 4°. Thus, a total of six warm-up trials, 20 test trials, and four cooldown trials were conducted each session. If the dolphin was incorrect on three or more of the warm-up and cooldown trials combined, the data from the session were eliminated from analysis.

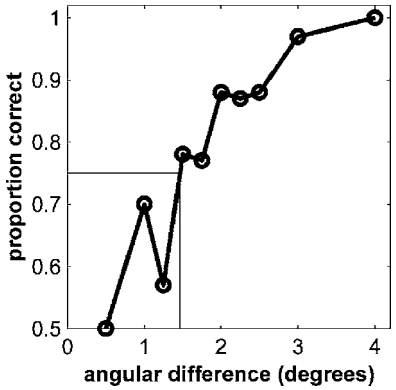


FIG. 4. Experiment I psychometric function with a 75% correct angular difference threshold of 1.5°.

C. Results and discussion

A total of 35 sessions were conducted over the duration of experiment I. Two of the sessions were aborted because Hiapo failed to respond correctly to five or more of the warm-up trials. One session was aborted to repair the stationing hoop that Hiapo dismantled in mid-session. Two additional sessions were aborted because Hiapo would not respond to his trainer. Therefore, data from a total of 30 sessions were used for analysis. A psychometric function describing the dolphin's performance is plotted in Fig. 4. Each data point represents the percentage of correct responses for 60 trials. As the angular difference between the stimuli decreased, so did the dolphin's ability to discriminate the angular difference. Using a criterion of 75% correct responses, a threshold of 1.5° was obtained.

There was no significant linear trend effect of angular difference on the number of clicks (see Fig. 5) the dolphin emitted ($F=0.97$, $p>0.05$, $y=-2.75x+195.45$). On average, the dolphin emitted 190.01 clicks per trial (SD=39.28). The maximum number of emitted clicks on a single trial was 287 while the smallest number was 97. There was a significant linear trend effect of angular difference on the dolphin's reaction time ($F=14.42$, $p<0.05$, $y=-0.14x+4.79$). On average, the dolphin's response latency was 4.50 s (SD=0.81 s).

Maximum and minimum response latencies were 9.25 and 2.91 s, respectively. Although the dolphin appears to require significantly more time to respond as the angular separation decreases, the trend appears to be rather weak with a minor negative slope.

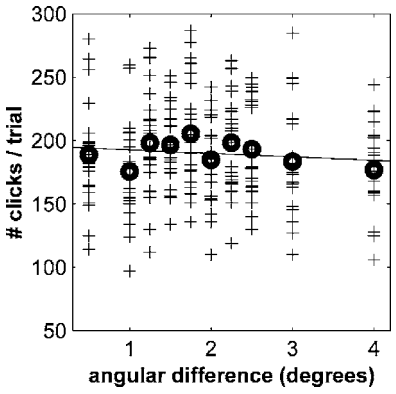


FIG. 5. Number of clicks/trial as a function of angular difference for experiment I. Pluses represent individual trials while bold circles are means for each angular difference.

The current threshold compares very well with the threshold of 1.6 measured by Branstetter *et al.* (2003). These data suggest that the dolphin was able to simultaneously localize six sound sources (i.e., Branstetter *et al.*, 2003) approximately as well as it can simultaneously localize three sound sources (i.e., current study). Thus, within the limited stimuli used between Branstetter *et al.* and the current study, echo complexity does not appear to be a major factor governing echoic angular discrimination.

IV. EXPERIMENT II

The goal of experiment II was to determine if the positions of the rods are less difficult to localize when each rod is more fully resolvable. The thresholds from experiment I and Branstetter *et al.* (2003) were 1.6 and 1.5 respectively. Because θ_w was equal to 2° in both of these experiments, the rods may have been too close together to be fully resolvable. By increasing θ_w to 4° each rod should be more fully resolvable, thus allowing the dolphin to make better judgments about the precise location of each rod relative to one another. If this assumption is true, the dolphin's angular discrimination threshold should be lower than those obtained by experiment I and Branstetter *et al.* (2003)

A. Stimuli

For experiment II, the angle between the rods within the *S*+ array (θ_w) was 4° and the angle between the *S*+ and *S*− array (θ_b) varied to produce angular differences (Δ_θ) of 4.00, 3.00, 2.50, 2.00, 1.50, 1.00, 0.50 and 0.25°.

B. Procedure

Experiment II employed the identical procedure to experiment I, however, there were eight angular differences tested. A subjective decision was made to decrease the number of presented angular differences from ten (in experiment I) to eight to shorten the experimental sessions. The longer sessions appeared to result in a decrease of motivation by the dolphin. Thus, a total of six warm-up trials, 16 test trials, and four cooldown trials were conducted each session.

C. Results and discussion

A total of 38 sessions were conducted. A three month hiatus resulted between the dates of January 8th and March 26th (2001) due to the death of a companion dolphin. The hiatus took place between sessions 18 and 19. During this time, a total of nine practice sessions (half the number of trials/session) were conducted to maintain the behaviors required for the experiment. An *a priori* decision was made to exclude all practice data from analysis. Experiment II resumed after the remaining companion dolphins were trained to facilitate (not interfere with) data collection. Six sessions were aborted because Hiapo failed to respond correctly to five or more of the warm-up trials. An additional two sessions were aborted because Hiapo chose to socialize with the

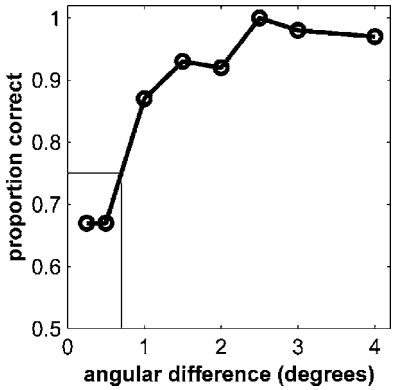


FIG. 6. Experiment II psychometric function with a 75% correct angular difference threshold of 0.7°.

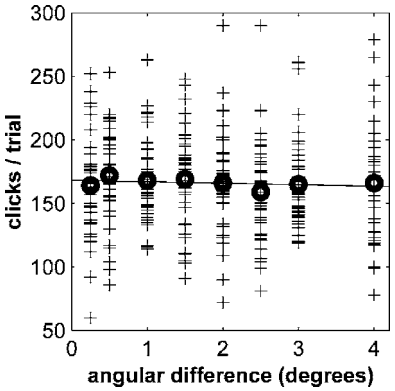


FIG. 7. Number of clicks/trial as a function of angular difference for experiment II. Pluses represent individual trials while bold circles are means for each angular difference.

other dolphins rather than respond to his trainer. Therefore, data from a total of 30 sessions were used for analysis. A psychometric function describing the dolphin's performance is plotted in Fig. 6. Each data point represents the percentage of correct responses for 60 trials. Using a criterion of 75% correct responses, a threshold of 0.7 degrees was obtained.

There was no significant linear trend effect of angular difference on the number of clicks the dolphin emitted ($F=0.40$, $p>0.05$, $y=-1.16x+168.19$), nor was there a significant linear trend effect of angular difference on the dolphin's reaction time ($F=3.98$, $p>0.05$, $y=-0.05x+3.91$). On average, the dolphin emitted 166.06 clicks per trial (SD=38.57). The maximum number of emitted clicks on a single trial (see Fig. 7) was 260 while the smallest number was 90. On average, the dolphin's response latency was 3.83 s (SD =0.67 s). Maximum and minimum response latencies were 8.70 and 2.17 s, respectively.

A point-by-point comparison between proportions correct from experiments I and II was done for each angular separation. The angular separations that experiment II did not have in common with experiment I were interpolated by averaging adjacent points and are denoted with an asterisk in Table I. Differences between points on the psychometric function were large enough to produce significant one-tailed z scores (indicated by bold p values) for all angular separation except for $\Delta\theta=2.00$, 3.00 and 4.00. The proportions associated with the last two angular separations are similar due to a ceiling effect. The angular separations surrounding the 75% correct level are all significantly different, thus we conclude the threshold differences are real. A statistical comparison to Branstetter *et al.* (2003) was not conducted because the results in this study are not counterbalanced with the former study. However, a large leap of faith is not required to notice the similarity between the psychometric function of Exp I. and Branstetter *et al.* (2003), (see Fig. 8).

The number of clicks from experiment I and experiment II were also compared. To prevent any potential biases, only the angular separations the two experiments had in common were chosen for comparison ($\Delta\theta=0.5$, 1, 1.5, 2, 2.5, 3, and 4). Because there was no significant linear trend in either experiment, a grand mean was calculated for each experiment by pooling the data from the angular separation above. The resulting mean clicks per trial for experiments I and II were 186.52 and 166.35, respectively. There was no significant difference between these means, $t(291)=1.47$, $p>0.05$ (two tailed).

Response time was also compared between experiments I and II with only the angular separation in common used for analysis. Despite the fact that there was a significant linear trend for reaction time in Exp. I, a grand mean was calculated for each experiment by pooling the data from the angular separation. There was a significant difference, $t(739)=12.54$, $p<0.01$ (two-tailed) between the means of 4.5 and 3.8 s for experiments I and II, respectively. Because simply pooling data results in the most parsimonious comparison with the least amount of variance accounted for (compared to a linear model with more than a single parameter), the significant result is considered conservative.

The correlation between response time and number of clicks was examined for a random subset of trials within each experiment. A moderate low correlation resulted in Exp. I ($R^2=0.24$, $n=99$) and a weak correlation resulted from Exp. II ($R^2=0.07$, $n=92$).

TABLE I. A point-by-point comparison between proportions correct from experiments I and II was compared for each angular separation. The angular separations that experiment II did not have in common with experiment I were interpolated by averaging adjacent points and are denoted by an asterisk. Differences between points on the psychometric function were large enough to produce significant one-tailed z scores (indicated by bold p values) for all angular separation except for $\Delta\theta=2.00$, 3.00 and 4.00.

Comparison Between Psychometric Functions from Exp. I and Exp. II										
$\Delta\theta$	0.50	1.00	1.25*	1.50	1.75*	2.00	2.25*	2.50	3.00	4.00
z	−1.85	−2.22	−4.13	−2.36	−2.40	−0.61	−1.78	−2.73	−0.58	⋯
p	**0.032**	**0.014**	**<0.001**	**0.009**	**0.008**	**0.248**	**0.038**	**0.003**	**0.281**	⋯

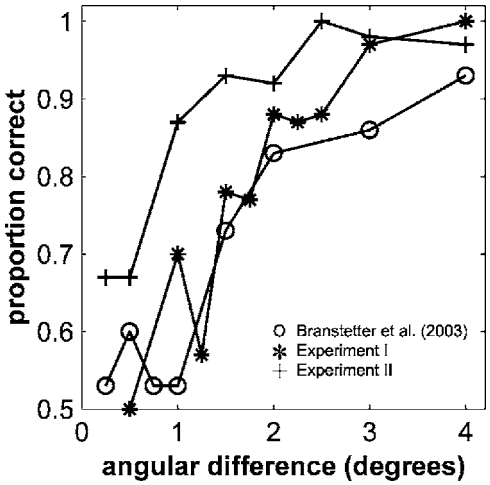


FIG. 8. Psychometric functions from Branstetter *et al.* (2003) and experiments I and II.

V. GENERAL DISCUSSION

The two main findings of this study are: (1) decreasing the number of targets from six to three does not aid in localization, and (2), increasing the space between the rods enhances localization. Although the finding from experiment I (that decreasing the number of targets does not aid in localization), departs from human results (Wightman and Kistler, 1997; Langendijk *et al.*, 2001), this may be a selective adaptation for dolphins. During foraging, dolphins are required to echoically detect, track, and capture multiple fish, often schooling in close spatial proximity. Having the ability to localize multiple targets simultaneously would no doubt be advantageous, if not a basic requirement for this type of auditory predation.

The results were consistent with a passive hearing study that determined the smallest possible angle (minimum audible angle or MAA) between two sound sources that would allow the two sources to be perceived as discrete. Renaud and Popper (1975) reported the horizontal MAA for 30, 60, and 90 kHz pure tones as 2.5, 3.0, and 3.0°, respectively. Horizontal and vertical MAAs for click stimuli (35 μs duration, 64.35 kHz peak frequency) were 0.9 and 0.7°, respectively.

The results from the current study, along with Branstetter *et al.* (2003), suggest that the dolphin's echoic spatial acuity is sufficient to support cross-modal matching of objects between vision and echolocation as shown by Pack and Herman (1995) and Herman *et al.* (1998). Because the dolphin in these studies interrogated relatively large objects (approximately 43 cm^2) at relatively short distances (<1 m), the angle the objects subtended was much greater than the angular differences threshold reported here (For a more extensive discussion on this topic, see Branstetter *et al.*, 2003).

The question remains, how do dolphins segregate individual echoes among several closely spaced echoes? One possible explanation for these findings may be related to the phenomenon of "summing-localization blur" caused by simultaneous sound sources. For human listeners, if two *coherent* sound sources are presented simultaneously at an equal distance from the midsagital plane, a single "phantom" sound is perceived half the distance between the sound sources. Coherence can be defined as two signals that are identical and the degree of coherence can be calculated by the normalized cross correlation function. As the degree of coherence decreases, summing-localization blur decreases (Jeffress *et al.*, 1962). Although the PVC rods in the experiment were identical, during ensonification, adjacent rods are unlikely to produce coherent echoes. The incident signal of the bottlenose dolphin is not only directional with respect to amplitude, but also highly directional and asymmetric in the frequency domain. Au (1980) demonstrated that peak frequency of the incident signal was 122 kHz directly in front of the dolphin but was 38 kHz only 10° to the right and 115 kHz, 10° to the left of the dolphin. Thus, targets separated in space will produce echoes with different peak frequencies and, thus, varying degrees of coherence. Although Au (1980) did not measure peak frequencies for small angles off the median, we fit a fourth order polynomial to the data suggesting that at 2 and 4°, the peak frequencies of the echoes would be 109 and 94 kHz, respectively. Thus, echoes off the median plane would have spectral differences of 13 and 28 kHz for 2 and 4°, respectively. Although a polynomial is overly simplistic, spectral differences are nevertheless likely to be larger for rods separated by 4° compared to 2°. These spectral differences decrease signal coherence of adjacent rod echoes, and may aid the dolphin in degrading summing-localization blur for closely spaced objects. Even when the dolphin is pivoting its head, the target directly in front of the dolphin is likely to have a different frequency and amplitude signature than peripheral targets.

In addition, lower frequencies will not reflect well off of a small object compared to higher frequencies due to their relatively large wavelengths. For example, two incident signals with a 38 and a 122 kHz peak frequencies will have 4.0 and 1.2 cm wavelengths, respectively. Because the PVC rods have a 1.9 cm outer diameter, the 38 kHz signal will produce a much more attenuated echo than the 122 kHz signal. Thus only a small portion of the array will probably produce salient echoes per each individual outgoing click. During echoic interrogation, Hiapo would pivot his head in the horizontal plane, suggesting the dolphin was ensonifying only a portion of the stimuli at a time. Adding multiple targets (e.g., four vs two rods from Branstetter *et al.*, 2003) may not affect his localization ability through interference because peripheral targets will produce increasingly attenuated echoes.

Another mechanism that may aid in producing noncoherent echoes will be the dolphin head related transfer function (HRTF). The dolphin's head behaves as a position dependant spectral filter (Branstetter and Mercado, 2006; Aroyan, 2001; Supin and Popov, 1993; Ketten, 2000) as well as a shaded receiver (Mohl *et al.* 1999). Echoes from one position in auditory space (relative to the head) will be spectrally different in another position. Unlike the pinna which is responsible for HRTF in mammals, the most likely candidates for dolphin HRTFs are internal anatomical structures such as the lipid jaw channels (Ketten, 2000; Aroyan, 2001).

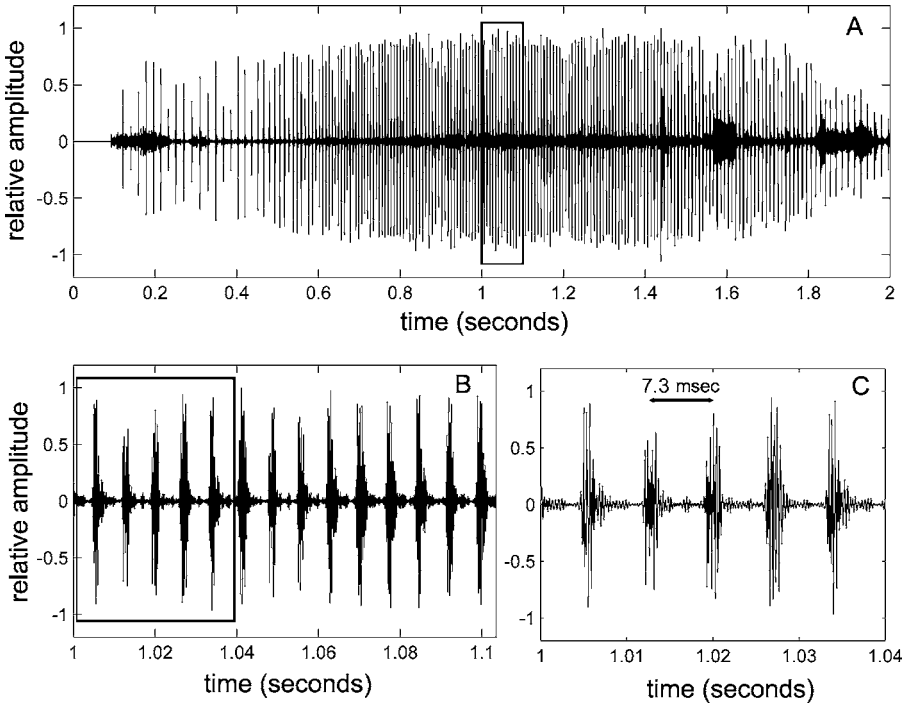


FIG. 9. An example of a standard click train from a trial in experiment I. (A) represents an entire click train from a single trial. (B) is a time expanded section from the rectangle in (A). (C) is a time expanded section from the rectangle in (B). Sampling rate was 44.1 kHz.

A. Click trains

Figure 9 is a plot of a typical click train from the current study. The number of clicks a dolphin emits during a sonar task is highly variable but typically related to how much effort is used to interrogate an object (Roitblat *et al.*, 1990). For example, a dolphin's click quantity has been positively correlated to level of masking noise (Au and Penner, 1981) and target range (Thompson and Turl, 1990) during detection tasks. In addition, dolphins emit more clicks for objects that are harder to identify in delayed matching-to-sample tasks (Roitblat *et al.*, 1990). The dolphin in the current study displayed similar (but weak) trends in adaptive acoustic behavior (Figs. 5 and 7); however, none of the trends were above statistical chance.

Inter-click intervals (ICIs) from wild dolphins have been reported to fit a bimodal distribution (Lammers *et al.*, 2003). Echolocation signals typically have ICIs greater than 15 ms while burst-pulse signals (observed in social contexts) have ICIs that begin, persist and end below 10 ms (Lammers *et al.*, 2003). Laboratory studies suggest dolphins require a finite amount of time (lag time) to process each incoming click before generating another click. Lag times between 19 and 45 m typically are added to the sound speed, two-way travel time for a number of experiments (see Au, 1993; p. 115–117). Although foraging dolphins sometime produce ICIs below the apparent minimum lag time (immediately before a prey capture), how these clicks are processed (or if they were processed at all) was a topic of speculation (Lammers *et al.*, 2003). Figure 10 displays ICIs from ten randomly sampled click trains from experiments I and II. Because of the large variability in the number of clicks between trials, only the first 20 clicks, the middle 20 clicks, and the last 20 clicks were used for analysis. All click trains had these three segments despite the overall number of clicks. The click trains in all of the experiments reported here typically start and end with longer and more variable inter-click intervals (Figs. 9 and 10) and lower amplitude clicks (Fig. 9). The stereotyped decrease and increase of the ICIs during the first 20 clicks and the last 20 clicks, respectively, may reflect the dolphin "focusing" the range of its sonar signal by dynamically adjusting the ICI to account for the delay of the object echo. The most stereotyped ICIs have an average interval of about 7.3 ms and occur during the vast majority of all click trains (Figs. 9 and 10). The two-way sound speed travel time for a target at a 200 cm distance is 1.3 ms. This suggests a lag time of 6.0 ms, considerably lower than previously reported (Au *et al.*, 1974; Au, 1980). Another possibility is that dolphins do not process echoes individually at short ranges, but perceive rapid echo returns as an auditory stream. ICIs would then be adjusted or "tuned" to decrease interference with the incident signal.

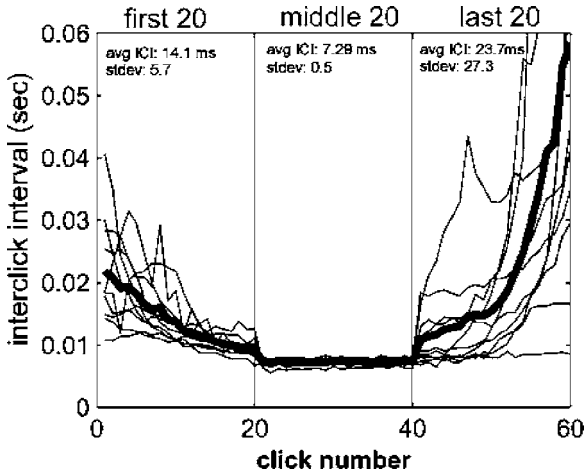


FIG. 10. Inter-click intervals sampled from ten randomly selected trials. Due to variable click train lengths, only the first, middle, and last 20 clicks in every train are compared. The bold line represents the average of the ten trials. The click trains in all trains start and end with longer and more variable ICIs. ICIs toward the center of the clicks trains are very stable.

One shortcoming of the current study is the absence of broadband acquisition of the dolphin's sonar signals. Dolphins are known to alter their sonar signals in the amplitude and frequency domain in response to environmental (Au *et al.*, 1985) and sonar task conditions (Au and Penner, 1981; Au, 1980). Because high frequency signals reflect better off of smaller object, the dolphin may selectively increase the higher frequency components of its sonar signal when attempting to resolve fine spatial details. This possibility remains untested.

B. Response time

Pooling the results from the above studies (significant trend in experiment I and significant difference between experiments I and II) suggests there is a subtle relationship between task difficulty and the amount of time the dolphin uses to make a response. Spending more time on a more difficult task makes intuitive sense. Because there was no significant difference in the number of clicks within on between the experiments, the difference in response time may reflect decision making time rather than observation time.

C. Auditory mechanisms

Of the auditory mechanisms used for sound source localization, the easiest to evaluate are ITDs. Using a simple two-receiver model (see Branstetter *et al.*, 2003), ITDs were calculated from the obtained angular discrimination thresholds. ITDs for thresholds of 1.6 (Branstetter *et al.*, 2003), 1.5 (Experiment I) and 0.7 (Experiment II) resulted in ITDs of 2.6, 2.4, and 1.1 μs, respectively. The smallest dolphin ITD threshold measured by Moore *et al.* (1995) was 7 μs for click stimuli with a peak frequency of 30 kHz. For click stimuli with peak frequencies between 60 and 90 kHz, ITD thresholds were between 17 and 18 μs. The ITD thresholds from the angular discrimination studies are several times smaller than those reported by Moore *et al.* (1995). In addition, Moore *et al.* (1995) calculated an ITD, for the dolphin MAA of 1° (Renaud and Popper, 1975), would be about 1.3 μs. Again, the MAA ITD was several times smaller than those reported by Moore *et al.* (1995). There are at least two possibilities for the resulting discrepancies. First, the model used for calculating ITD in the current study was overly simplistic. The model did not take into account reflective or refractive properties of the dolphin's head such as those produced by the skull or the lipid channels in the lower jaw. The internal structures of the dolphin head could possibly increase the ITD threshold. However, it is unlikely that the ITD threshold could be increased by several factors to be consistent with ITDs measured by Moore *et al.* (1995). Second, the dolphin may not have used ITDs for fine angular discrimination. IIDs, binaural spectral differences or monaural spectral cues may have been used instead. IIDs are more salient at higher frequencies for terrestrial mammals and dolphins (Supin and Popov, 1993). The high degree of sound shadowing produced by the dolphin head ($>$20 dB; Supin and Popov, 1993) and the high level of IID sensitivity ($<$1 dB) measured by Moore *et al.* (1995) suggests IIDs may play a significant and possibly dominant role in sound source localization for fine horizontal angular discrimination. The anatomical structures responsible for producing IIDs and binaural spectral differences remain speculative.

Although IIDs and ITDs can provide the horizontal position of a sound source, these cues typically provide little if any information about source elevation. For terrestrial mammals, vertical and monaural localization are due to the spectral filtering properties of the pinna. The pinna behaves as a position-dependent spectral filter that produces subtle differences in the received sound depending on the location of the sound source. The direction dependent transfer function created by the pinna (and to a lesser degree, the head and torso) is known as the head related transfer function (HRTF). Interestingly, auditory predators often display elaborately convoluted pinnae (e.g., members of the order *Chiroptera*) and asymmetry in external auditory apparatus (e.g., skull asymmetry in the barn owl, *Tyto alba*). Human studies suggests HRTF generated cues are greatest for higher frequencies (Kisler and Wightman, 1992; Middlebrooks and Green, 1992) and broadband sounds. Because dolphin hearing is both broadband with good sensitivity at higher frequencies, HRTFs may be a likely candidate for vertical and perhaps horizontal sound localization (for a more detailed discussion see Branstetter and Mercado, 2006).

VI. CONCLUSION

The ability to localize fine spatial details through sonar no doubt has several advantages for wild dolphins living in visually restricted environments. For example, the Amazon River dolphin (*Inia geoffrensis*) feeds in shallow murky waters with very limited visibility. Still others such as the Hawaiian Spinner dolphin (*Stenella longirostris*) are nocturnal feeders (Perrin and Gilpatrick, 1994). Most dolphins typically prey on small fish and squid and are restricted to capturing prey in a serial fashion. Thus, a dolphin may be required to echoically isolate and capture prey tens to hundreds of times within a 24 h period. This taxing requirement demands a sonar system well adapted for precise target localization through echolocation. Echoic localization may also be important for protection from potential predators such as sharks or other marine mammals. Echoically locating and identifying predators at a distance may provide a dolphin with enough time to evade predation.

The results reported here demonstrate one of many methods of measuring spatial acuity through echolocation. Because we wanted these data to generalize well to the cross-modal data, we constructed the stimuli out of similar materials (PVS) with similar target complexity (multiple targets). A simpler test would require the dolphin to discriminate between two objects (example: large sphere vs small sphere). The angle between the objects could be adjusted in a similar manner as the studies presented here, perhaps yielding a finer estimate of the dolphin's angular resolution.

Although horizontal echoic angular resolution has been measures for a handful of conditions, only a single passive listening study has measured sound source localization in the vertical plane (Renaud and Popper, 1975). Despite the lack of binaural stimulus differences, the animal's MAA was

slightly smaller in the vertical plane. If this result can be corroborated, an even stronger case that dolphins employ HRTFs for fine sound source localization and echoic imaging can be made.

ACKNOWLEDGMENTS

This work would not have been possible without the hard work from the many interns, volunteers, graduate students and staff at the Kewalo Basin Marine Mammal Laboratory. I would like to extend special thanks to the "angular resolution interns" who were the backbone of this project from the first day we began to habituate Hiapo to his "eyecups" to the last day of data collection. Once again, thanks to Matthias Hoffmann-Kuhnt for the 3D renderings of the experimental setup. Thanks to Paul Nachtigall and Whitlow Au for suggestions on earlier versions of this paper. Additional thanks to two anonymous reviewers for their helpful comments. This work was in part funded by NSF Grant No. 0090744.

APPENDIX

The arrays were centered on the wooden arc so the far end of the *S*+ array and the *S*− rray were an equal distance to the center of the arc and, hence, the dolphin's median plane. The center of the stimuli can be calculated by

$$\frac{\theta_w + \theta_b}{2}.$$

As θ_b was varied, the distance from the center of the rack to the furthermost ends of the stimuli also varied. The position of the center of θ_b relative to the center of the arc can be calculated by

$$D = \left(\theta_w + \frac{\theta_b}{2}\right) - \left(\frac{\theta_w + \theta_b}{2}\right),$$

where D is the distance from the center of the arc to the center of θ_b. The solution for D is the constant $\theta_w/2$.

[1]Henceforth, the word "dolphin" will be used to refer to the bottlenose dolphin (*Tursiops truncatus*).

[2]Although only two experiments are reported here, a pilot study and a control study were also conducted, and counterbalanced (Branstetter, 2005). The control study demonstrated the dolphin was not using a potential time-delay confound from echoes reflecting off the tank wall.

Aroyan, J. L. (**2001**). "Three-dimensional modeling of hearing in *Delphinus delphis*," J. Acoust. Soc. Am. **110**(6), 3305–3318.

Au, W. W. L. (**1993**). *The Sonar of Dolphins* (Springer-Verlag, New York).

Au, W. W. L., Carder, D. A., Pennerm, R. H., and Scronce, B. L. (**1985**). "Demonstration of adaptation in beluga wale echolocation signals," J. Acoust. Soc. Am. **77**, 726–730.

Au, W. W. L., and Penner, R. H. (**1981**). "Target detection in noise by echolocating Atlantic bottlenose dolphin," J. Acoust. Soc. Am. **70**, 251–282.

Au, W. W. L., Floyd, R. W., Penner, R. H., and Murchison, A. E. (**1974**). "Measurement of echolocation signals of the Atlantic bottlenose dolphin, *Tursiops truncates Montagu*, in open waters," J. Acoust. Soc. Am. **54**, 1280–1290.

Au, W. W. L. (**1980**). "Echolocation signals of the Atlantic bottlenose dolphin (*Tursiops truncatus*)." in *Animal Sonar Systems*, edited by R. G. Busnel and J. F. Fish (Plenum, New York), pp. 251–282.

Branstetter, B. K., and Mercado III, E. (**2006**). "Sound localization by cetaceans," Int. J. Comp. Psychol. **19**, 25–61.

Branstetter, B. K. (**2005**). "Sound localization and auditory perception by an echolocating bottlenose dolphin (*Tursiops truncatus*)," unpublished Doctoral Dissertation, University of Hawaii, Manoa.

Branstetter, B. K., Mevissen, S. J., Herman, L. M., Pack, A. A., and Roberts, S. P. (**2003**). "Horizontal angular discrimination by an echolocating bottlenose dolphin *Tursiops truncatus*," Bioacoustics 14, 15–34.

Herman, L. M., Pack, A. A., and Hoffmann-Kuhnt, M. (**1998**). "Seeing through sound: Dolphins (*Tursiops truncatus*) perceive the spatial structure of objects through echolocation," J. Comp. Psychol. **112**, 292–305.

Jeffress, L. A., Blodgett, H. C., and Deatherage, B. H. (**1962**). "Effects of interaural correlation on the precision of centering noise," J. Acoust. Soc. Am. **34**, 1122–1123.

Johnson, S. C. (**1967**). "Sound detection thresholds in marine mammals," *Marine BioAcoustics*, edited by W. Tavolga (Pergamon, New York), pp. 240–260.

Ketten, D. R. (**2000**). "Cetacean ears," *Hearing by Whales and Dolphins*, edited by W. W. L. Au, A. N. Popper, and R. R. Fay (Springer, New York), pp. 43–108.

Kisler, D. J., and Wightman, F. L. (**1992**). "A model of head-related transfer functions based on principle components analysis and minimum-phase reconstruction," J. Acoust. Soc. Am. **91**, 1637–1647.

Lammers, M. O., Au, W. W. L., and Herzing, D. L. (**2003**). "The broadband social acoustic signaling of spinner and spotted dolphins," J. Acoust. Soc. Am. **114**, 1629–1639.

Langendijk, E. H. A., Wightman, F. L., and Kistler, D. J. (**2001**). "Sound localization in the presence of one or two distracters," J. Acoust. Soc. Am. **109**, 2123–2134.

Middlebrooks, J. C., and Green, D. M. (**1992**). "Observations on a principle components analysis of head-related transfer functions," J. Acoust. Soc. Am. **92**, 597–599.

Mohl, B., Au, W. W. L., Pawloski, J. L., and Nachtigall, P. E. (**1999**). Dolphin hearing: Relative sensitivity as a function of point of application of a contact sound source in the jaw and head region. J. Acoust. Soc. Am. **105**, 3421-3424.

Moore, P. W. B., Pawloski, D. A., and Dankiewicz, L. (**1995**). "Interaural time and intensity difference thresholds in the bottlenose dolphin (*Tursiops truncatus*)," *Sensory Systems of Aquatic Mammals*, edited by R. A. Kastelein, J. A. Thomas, and P. E. Nachtigall (De Spil, Woerden, The Netherlands).

Moore, P. W. P., Hall, R. W., Friedl, W. A., and Nachtigall, P. E. (**1984**). "The critical interval in dolphin echolocation: What is it?" J. Acoust. Soc. Am. **76**, 314–317.

Pack, A. A., and Herman, L. M. (**1995**). "Sensory integration in the bottlenosed dolphin: Immediate recognition of complex shapes across the senses of echolocation and vision," J. Acoust. Soc. Am. **98**, 722–733.

Pack, A. A., Herman, L. M., and Hoffmann-Kuhnt, M. (2004). "*Dolphin echolocation shape perception: From sound to object," Echolocation in Bats and Dolphins*, edited by J. A. Thomas, C. F. Moss, and M. Vater (University of Chicago Press, Chicago), pp. 288–298.

Pack, A. A., Herman, L. M., Hoffmann-Kuhnt, M., and Branstetter, B. K. (**2002**). "The object behind the echo: Dolphins (*Tursiops truncatus*) perceive object shape globally through echolocation," Behav. Processes **58**, 1–26.

Perrin, W. F., and Gilpatrick, J. W. Jr. (**1994**). "Spinner dolphin—Stenella longirostris (Gray, 1828)," in *Handbook of Marine Mammals, The First Book of Dolphins*, edited by S. H. Ridgway, S. R. Harrison, Vol. 5 (Academic, London), pp. 99–128.

Renaud, D. L., and Popper, A. N. (**1975**). "Sound localization by the bottlenose porpoise, *Tursiops truncatus*," J. Exp. Biol. **63**, 569–585.

Roitblat, H. L., Penner, R. H., and Nachtigall, P. E., (**1990**). Jorurnal of Experimental Psychology: Annimal Behavior Processes, **16**, 85-95.

Supin, Y. A., and Popov, V. V. (**1993**). "Direction-dependant spectral sensitivity and interaural spectral difference in a dolphin: Evoked potential study," J. Acoust. Soc. Am. **93**, 3490–3495.

Thompson, J. A., and Turl,C. W., (**1990**). "Echolocation Characteristics and Range Detection by a False Killer Whale (Pseudorca Crassidens)," edited by J. A. Thomas and R. Kasterlein in *Cetacean Sensory Systems: Field and Laboratory Evidences* (Plenum Press, New York), pp. 321-334.

Wightman, F. L., and Kistler, D. J. (**1997**). "Sound localization in the presence of multiple distracters," J. Acoust. Soc. Am. **101**, 3105.

Two-dimensional ultrasound detection with unfocused frequency-randomized signals

Gregory T. Clement[a)]
Department of Radiology, Harvard Medical School, Brigham and Women's Hospital, Boston, Massachusetts 02115

(Received 26 June 2006; revised 20 October 2006; accepted 1 November 2006)

A method is described for detecting scattering in two-dimensions using an unfocused ultrasound field created from a continuously driven source array. The frequency of each element on the array is unique, resulting in a field that is highly variant as a function of both time and position. The scattered signal is then received by a single receiving line. The method, as currently written, is valid under the first order Born approximation. To demonstrate the approach, a series of simulations within the frequency range of 0.10–1.25 MHz are performed and compared with a simulated B-Scan in the same frequency range. The method is found to be superior in resolving closely spaced objects, discerning 1.4 mm separation in the radial and 0.5-mm separation in the axial direction. The method was also better able to determine object size, resolving scatters less than 10% of wavelength associated with the center frequency. © *2007 Acoustical Society of America.*
[DOI: 10.1121/1.2400847]

PACS number(s): 43.80.Qf, 43.80.Vj, 43.20.El, 43.20.Ye, 43.35.Yb [FD] Pages: 636–647

I. INTRODUCTION

Ultrasound imaging methods have generally employed temporal bandwidth and frequency to determine image resolution; the former setting the axial resolution and the latter determining the radial beamwidth. Imaging has, therefore, turned to increasingly higher frequencies as a way to achieve higher imaging resolution via a more localized focus in the transmission and/or receiving array.[1–3] These high-frequency approaches, however, are generally performed at the expense of higher beam attenuation as a function of frequency, limiting their application to more superficial locations.

Alternatively, the ability to detect small phase distortions in an ultrasound field opens the possibility of imaging objects smaller than the focal size. Various ultrasound methods for creating images that are small relative to the imaging acoustic wavelength have been investigated. For example, near field imaging[4] records information within a distance of several wavelengths from the source in order to collect information from evanescent wave components of the signal, where the magnitude decays exponentially with distance from an object. Time reversal[5] uses information from scattering centers in a highly inhomogeneous field to focus beyond that which is obtainable in a homogeneous medium. Spectral methods[6,7] seek to reconstruct higher spatial frequencies beyond the cut-off frequency using *a priori* information about the image.

Presently, an alternative method of ultrasound image construction that creates a signal composed of randomized and spatially separated time-harmonic frequencies is examined. It will be shown that both transmission and interpretation of the resultant signal does not require a focused beam, allowing for a center frequency that is significantly below that which is required by traditional backscattered ultrasound methods. Using a source array, each element emits a signal differing in frequency from any other element. The individual frequencies are selected so that the net signal is comprised of a relatively large bandwidth, but in this manner, the field is unfocused and seemingly incoherent. A single point-like receiver is used to record the time history of the signal scattered from objects in front of the transducer. Reconstruction of the entire planar region of interest (ROI) is performed from analysis of the time trace acquired from this single receiver. To improve the image, the process is repeated a selected number of times, each time using a new randomized frequency pattern. Signal analysis consists of a Fourier-based approach that determines scattering locations using the spatially unique phasing patterns created by randomization of the array.

The physical basis of the reconstruction is rooted in its ability to create a known pressure field, with a frequency spectrum whose phase is highly varying as a function of position. In this manner, each point within the ROI possesses a unique, or nearly unique, waveform. The receiver records a superposition of these waveforms, as scattered by any objects present in the field. This signal is used to construct an image of the scattering medium by searching for the presence of each unique waveform within the signal.

This paper aims to describe the method as well as perform a preliminary feasibility assessment of the feasibility of the approach. A detailed description of the ultrasound field and the reconstruction of a scattering region is provided. It is shown that the method provides a means for inverting the linear scattering equation, allowing solutions to the corresponding scattering functions over a limited ROI. A numeric study is then performed to assess the method's ability to construct a two-dimensional scattering field from data gathered in a single temporal dimension. The method is evaluated by comparison of two-dimensional image reconstructions with simulated B-scanned images, assuming a transducer

[a)]Electronic mail: gclement@hms.harvard.edu

having geometry and bandwidth identical to that used by the randomized frequency method. The resolution limits of the two approaches are determined by measuring two closely situated scattering objects. Scattering of multiple objects and larger inclusions is also demonstrated. It will be shown that the random-frequency approach has the potential to detect small objects beyond the abilities of current widely used methods.

II. THEORY

To illustrate the random-frequency approach, the proposed emitter is first approximated as an array of simple sources, each radiating in a continuous fashion at a unique angular frequency ω. The pressure at any point in a homogeneous space due to a single source at r_0 on the radiator is then given by

$$p_\omega(\mathbf{r},t) = -ic_0k_0\rho_0S_\omega g_\omega(\mathbf{r_{S_\omega}}|\mathbf{r_0}), \tag{1}$$

with sound speed c_0, wave number k_0, density ρ_0, source strength S_ω, and the function

$$g_\omega(\mathbf{r_{S_\omega}}|\mathbf{r_0}) = \frac{e^{-ik|\mathbf{r}-\mathbf{r_0}|}}{4\pi|\mathbf{r}-\mathbf{r_0}|}, \tag{2}$$

that will serve as a Green's function for the relevant wave equation.

If the radiating wave encounters a region of spatially varying density ρ and sound speed c, the time-harmonic acoustic pressure may then be described by the acoustic wave equation

$$\rho\,\nabla\cdot\left(\frac{1}{\rho}\nabla p_\omega\right) + \frac{\omega^2}{c^2}p_\omega = 0. \tag{3}$$

Following the structure of Morse and Ingard,[8] Eq. (3) is first multiplied by ρ_0/ρ and then $-(\nabla^2+\omega^2/c_0^2)p_\omega$ is added to both sides of the equation

$$\nabla^2 p_\omega + k_0^2 p_\omega = \nabla\cdot\left(\left\{1-\frac{\rho_0}{\rho}\right\}\nabla p_\omega\right) + \left\{k_0^2 - \frac{\rho_0}{\rho}k^2\right\}p_\omega, \tag{4}$$

giving the form of a harmonically driven distributed source, which in the absence of the scattering region reduces to a Helmoltz equation. Equation (4) may now be written in the form of a Lippmann Schwinger integral equation[9]

$$p_\omega(\mathbf{r}_R) = -ic_0k_0\rho_0S_\omega g_\omega(\mathbf{r_{S_\omega}}|\mathbf{r}_R) + \int\int\int_{\mathrm{ROI}} (\nabla\cdot(q_\rho(\mathbf{r})\,\nabla p_\omega) + q_\kappa(\mathbf{r})k_0^2 p_\omega) g(\mathbf{r}|\mathbf{r}_R)dV, \tag{5}$$

which represents the incident wave plus the scattered wave. The function $q_\rho(\mathbf{r}) = 1-\rho_0/\rho$ provides a measure of the spatial variation in density while $q_\kappa(\mathbf{r}) = 1-(\rho_0c_0^2/\rho c^2)$ is a function of variation in compressibility. It is further assumed that the scattered field is weak, such that the first-order Born approximation holds,[10] as has been widely used in ultrasound scattering in standard and modified[11] forms. The scattered pressure recorded at a point receiver located at $\mathbf{r_R}$ will be linearly dependent on the initial source function and Eq. (5) becomes

$$p(\mathbf{r_R}) \approx ic_0k_0\rho_0S_\omega\left[g_\omega(\mathbf{r}|\mathbf{r_0}) + \int\int\int_{\mathrm{ROI}} [g_\omega(\mathbf{r}_R|\mathbf{r})\,\nabla \cdot (q_\rho(\mathbf{r})\,\nabla\, g_\omega(\mathbf{r}|\mathbf{r_0})) + q_\kappa(\mathbf{r})k_0^2 g_\omega(\mathbf{r_R}|\mathbf{r})g_\omega(\mathbf{r}|\mathbf{r_0})]dV\right]. \tag{6}$$

The integral in Eq. (6) may be expanded using the vector identity

$$\phi(\nabla\cdot\mathbf{A}) = \nabla\cdot(\phi\mathbf{A}) - \mathbf{A}\cdot\nabla\phi, \tag{7}$$

so that by the divergence theorem

$$\int\int\int \nabla\cdot(\phi\mathbf{A})dV = \oiint \phi\mathbf{A}\cdot d\mathbf{S}, \tag{8}$$

where S is the surface surrounding the ROI, the first term in the identity given by Eq. (7) integrates to zero. The equation presented by Eq. (6) becomes

$$p(\mathbf{r_R}) \approx ic_0k_0\rho_0S_\omega\left[g_\omega(\mathbf{r}|\mathbf{r_0}) + \int\int\int_{\mathrm{ROI}} [q_\rho(\mathbf{r})\,\nabla\, g_\omega(\mathbf{r}_R|\mathbf{r})\,\nabla \times g_\omega(\mathbf{r}|\mathbf{r_0}) + q_\kappa(\mathbf{r})k_0^2 g_\omega(\mathbf{r_R}|\mathbf{r})g_\omega(\mathbf{r}|\mathbf{r_0})]dV\right]. \tag{9}$$

The central problem lies in the classic problem of inverting Eq. (9), to provide solutions of the scattering functions $q_\rho(\mathbf{r})$ and $q_\kappa(\mathbf{r})$. To tailor the field in a way that will allow this inversion to be performed, n simple sources are now considered, each radiating continuously at its own independent frequency $\omega(r_0)$. The scattered acoustic pressure at $\mathbf{r_R}$ will then be time-dependent as described by

$$p(\mathbf{r_R},t) = \sum_{\mathbf{r_0}}^{n} e^{i\omega(\mathbf{r_0})t}\int\int\int_{\mathrm{ROI}} [q_\rho(\mathbf{r})P_\rho(\mathbf{r_R},\mathbf{r},\mathbf{r_0}) + q_\kappa(\mathbf{r})P_\kappa(\mathbf{r_R},\mathbf{r},\mathbf{r_0})]dV. \tag{10}$$

where the kernels $P_\rho(\mathbf{r_R},\mathbf{r},\mathbf{r_0})$ and $P_\kappa(\mathbf{r_R},\mathbf{r},\mathbf{r_0})$, having the dimension of pressure per unit volume, are obtained by combining terms in Eq. (9). Successful reconstruction requires that Eq. (10) be invertible, or at least pseudo-invertible given $p(\mathbf{r_R},t)$ and the kernels.

The problem of inversion is presently approached by writing the Fourier integral transform of the received signal

$$\tilde{p}(\mathbf{r_R},\omega) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} p(\mathbf{r_R},t)e^{i\omega t}dt, \tag{11}$$

as a discrete summation of the scattered signal from finite volumes, ΔV, in space

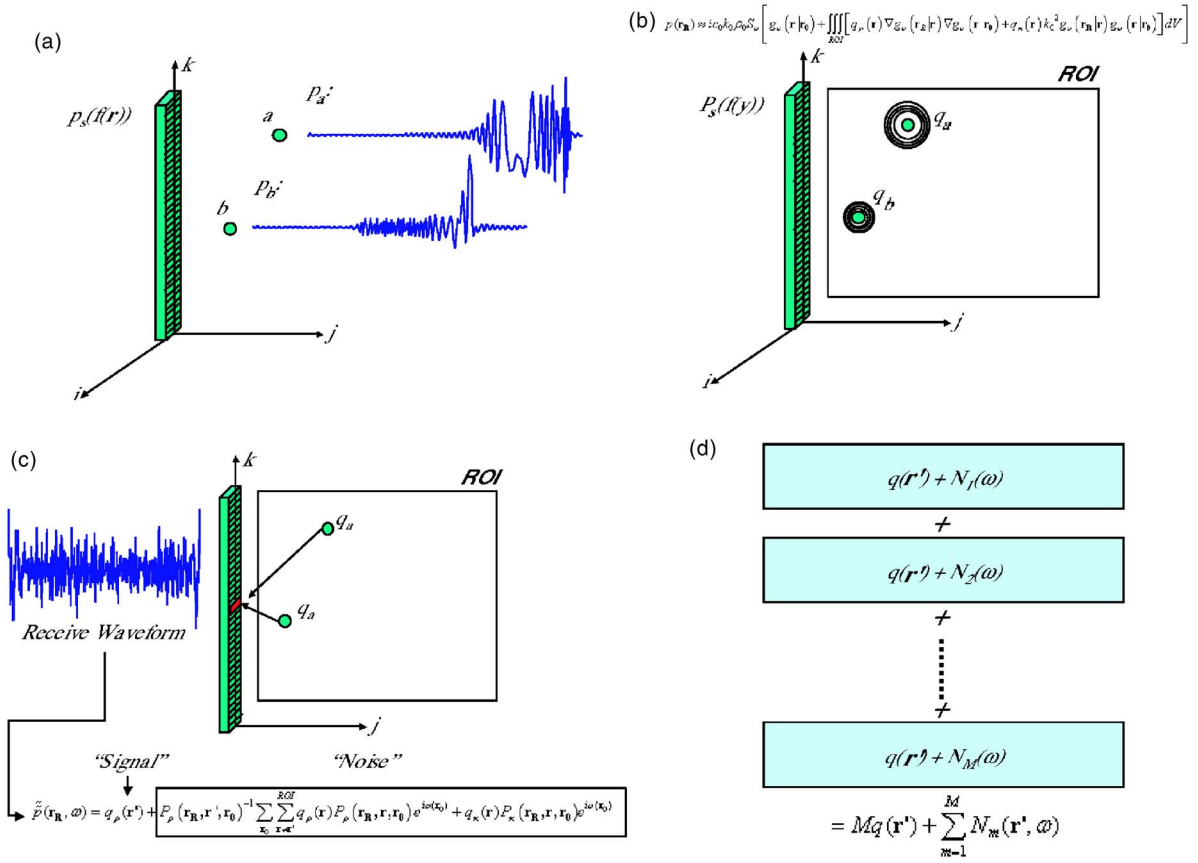


FIG. 1. (a) A pressure field is generated by driving each of N transducer elements with a unique frequency within the bandwidth of the transducer, creating a complicated broadband field pattern. (b) This field is assumed to encounter variation in density and sound speed within a region of interest (ROI) causing weak scattering, $q(r)$ (first-order Born approximation). (c) Each point in the ROI is tested using the inverse of the kernel for the location. The rest of the signal will present itself as noise N. (d) If this process is repeated with a new randomization, the "noise" averages toward zero, while the signal gets stronger.

$$\tilde{p}(\mathbf{r_R},\omega)=\sum_{\mathbf{r_0}}^{n}\sum_{\mathbf{r}}^{\mathrm{ROI}} q_\rho(\mathbf{r})P_\rho(\mathbf{r_R},\mathbf{r},\mathbf{r_0})\delta(\omega-\omega(\mathbf{r_0})) + q_\kappa(\mathbf{r})P_\kappa(\mathbf{r_R},\mathbf{r},\mathbf{r_0})\delta(\omega-\omega(\mathbf{r_0}))\Delta V, \tag{12}$$

where the Fourier transform properties of the Dirac delta function δ have been used. The scattering strength at a specific location in the ROI, $\mathbf{r}=\mathbf{r}'$, is analyzed by multiplying both sides of Eq. (12) by

$$\left[\sum_{\mathbf{r_0}}^{n} P_\rho(\mathbf{r_R},\mathbf{r}',\mathbf{r_0})\delta(\omega-\omega(\mathbf{r_0}))\right]^{-1}$$

or $$\left[\sum_{\mathbf{r_0}}^{n} P_\kappa(\mathbf{r_R},\mathbf{r}',\mathbf{r}_0)\delta(\omega-\omega(\mathbf{r_0}))\right]^{-1}.$$

Without loss of generalization, the case of solving for $q_\rho(\mathbf{r}')$ will be considered and applied to both sides of Eq. (12)

$$\tilde{p}(\mathbf{r_R},\mathbf{r}',\omega)=q_\rho(\mathbf{r}')+\frac{\sum_{\mathbf{r_0}}^{n}\sum_{\mathbf{r}\neq\mathbf{r}'}^{\mathrm{ROI}} q_\rho(\mathbf{r})P_\rho(\mathbf{r_R},\mathbf{r},\mathbf{r_0})\delta(\omega-\omega(\mathbf{r_0})) + q_\kappa(\mathbf{r})P_\kappa(\mathbf{r_R},\mathbf{r},\mathbf{r_0})\delta(\omega-\omega(\mathbf{r_0}))}{\sum_{\mathbf{r_0}}^{n} P_\rho(\mathbf{r_R},\mathbf{r}',\mathbf{r_0})\delta(\omega-\omega(\mathbf{r_0}))}. \tag{13}$$

In this form, the equation is separated in to the desired value $q_\rho(\mathbf{r}')$, referred herein as the *signal*, and the remaining terms, which will be regarded as *noise*, N. The inverse Fourier transform of Eq. (13) with respect to frequency

$$\tilde{\tilde{p}}(\mathbf{r_R},t)=q_\rho(\mathbf{r}')\delta(t)+\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}N(\mathbf{r_R},\mathbf{r}')e^{-i\omega t}d\omega, \tag{14}$$

will produce a localized peak centered about time $t=0$ that is

proportional in amplitude to $q_\rho(\mathbf{r}')$. The ability to detect this signal will require that the "noise" terms are sufficiently small relative to the signal. If the amplitudes and phases of each noise term are randomized over the frequency spectrum, their contribution in the time domain will generally *not* be localized, but rather will resemble a stochastically varying signal as a function of time.

With this observation, it is now assumed that the frequency of each simple source on the emitting transducer surface is randomly selected over the range $(\omega_{\min}, \omega_{\max})$, where the range is evenly divided based on the number of transducer elements. This construction process is summarized conceptually in Fig. 1. To enhance the signal strength, an additional independent random frequency distribution may be generated at the source and the two resulting signals combined. In this manner, if a series of M signals are recorded and summed, the signal at $q_\rho(\mathbf{r}')$ will increase linearly with M while the noise N will be further randomized over the frequency domain

$$\tilde{\tilde{p}}_M(\mathbf{r}', \omega) = Mq(\mathbf{r}') + \sum_{m=1}^{M} N_m(\mathbf{r}', \omega), \tag{15}$$

and consequently further distributed over the time domain. If the transform of N is sufficiently small, the signal-to-noise ratio will be large enough to recover the signal, which is taken to be the value in the time domain at $t=0$. This process is repeated for all $\mathbf{r}'$ over the region of interest to form an image.

III. METHODS

A. Simulated array

In this preliminary investigation of the randomized frequency method, fields are selected that could feasibly be produced by a one-dimensional array. An operating frequency range between 0.1 and 1.25 MHz is chosen, motivated in part by the potential medical implications of these frequencies, which are below the range of existing medical diagnostic transducers. The simulated array has a length $y=40$ mm and width of $x=10$ mm that is segmented into 202 linear sources with no kerf, propagating into an otherwise homogeneous medium containing a region of scattering sources. The transducer surface is assumed to be situated in the x-y plane with the axis of symmetry along the positive z-axis, which extends from the geometric center of the surface.

To simulate the acoustic pressure field, the transducer surface is divided into simple sources with diameters equal to $\frac{1}{4}$ wavelength corresponding to the highest transmitted frequency in water ($c=1500$ m/s). A linear distribution of 202 frequencies between 0.1 and 1.25 MHz is determined with a frequency resolution of 5.7 kHz, allowing each element on the transducer to be assigned exactly one frequency as obtained using a uniformly distributed pseudo-random number generator. Again by random selection, a single element is also selected to serve as the receiver. A scattering field is placed within the ROI, and the scattered signal reaching the receiver is simulated as a discrete approximation to Eq. (10). This calculated signal is stored for later processing. In practice, the simulation is performed by calculating the scattering element-by-element, and then linearly superimposing the individual time-domain waveforms at the receiver to form the resultant signal.

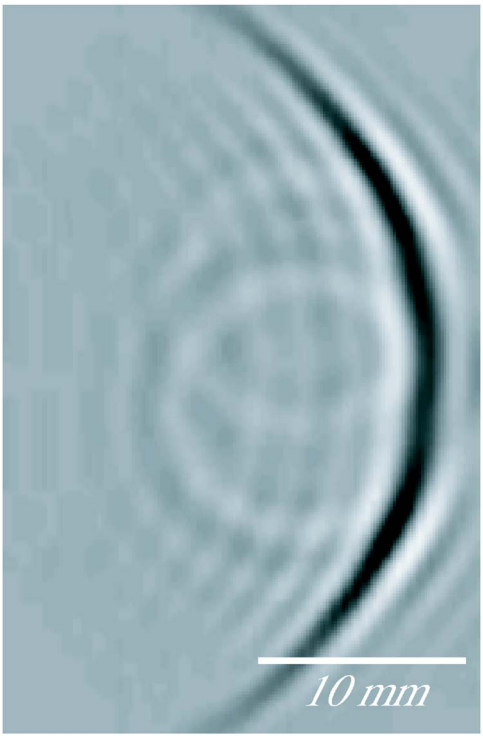


FIG. 2. Slice from a three-dimensional k-space simulation of a focused, pulsed wave traveling through a medium with scatterers smaller than the imaging wavelengths. Scattering from two objects is present in the image.

The scattering kernels, P_ρ and P_κ, given in (10) are also determined for every point over the ROI for a particular random field pattern. The Fourier transform of the received signal is calculated using a discrete approximation to the integral given by (11) and the reconstruction indicated by (13) is then performed. The entire process is repeated M times, and the results summed as given by (15). This sum is inverse transformed with respect to time, providing the image intensity at a single point.

B. B-scan simulation

B-Scan images are simulated using focused pulses with bandwidths that span the same frequencies used in the randomized imaging, but with time-localized impulse responses within a 0.1–1.25 MHz bandwidth. The emitting array is also geometrically identical to that used in the random-frequency case. Images are assembled by acquiring a series of echoes oriented by phase-controlled beam steering. This phasing is restricted to the y-direction only, with phases calculated according to $\theta_n = 2\pi\sqrt{(y-y_n')^2+Z^2}/c$, where Z defines the distance of the focal line on the z axis, and y_n' is the location of the center of the nth element on the array. In each image, twenty-one focal positions a distance $z=30$ mm in front of the transducer are created along the line from $y=+10$ mm to $y=-10$ mm. Each of these waveforms is propagated into the ROI and the scattered signal recorded by the center element

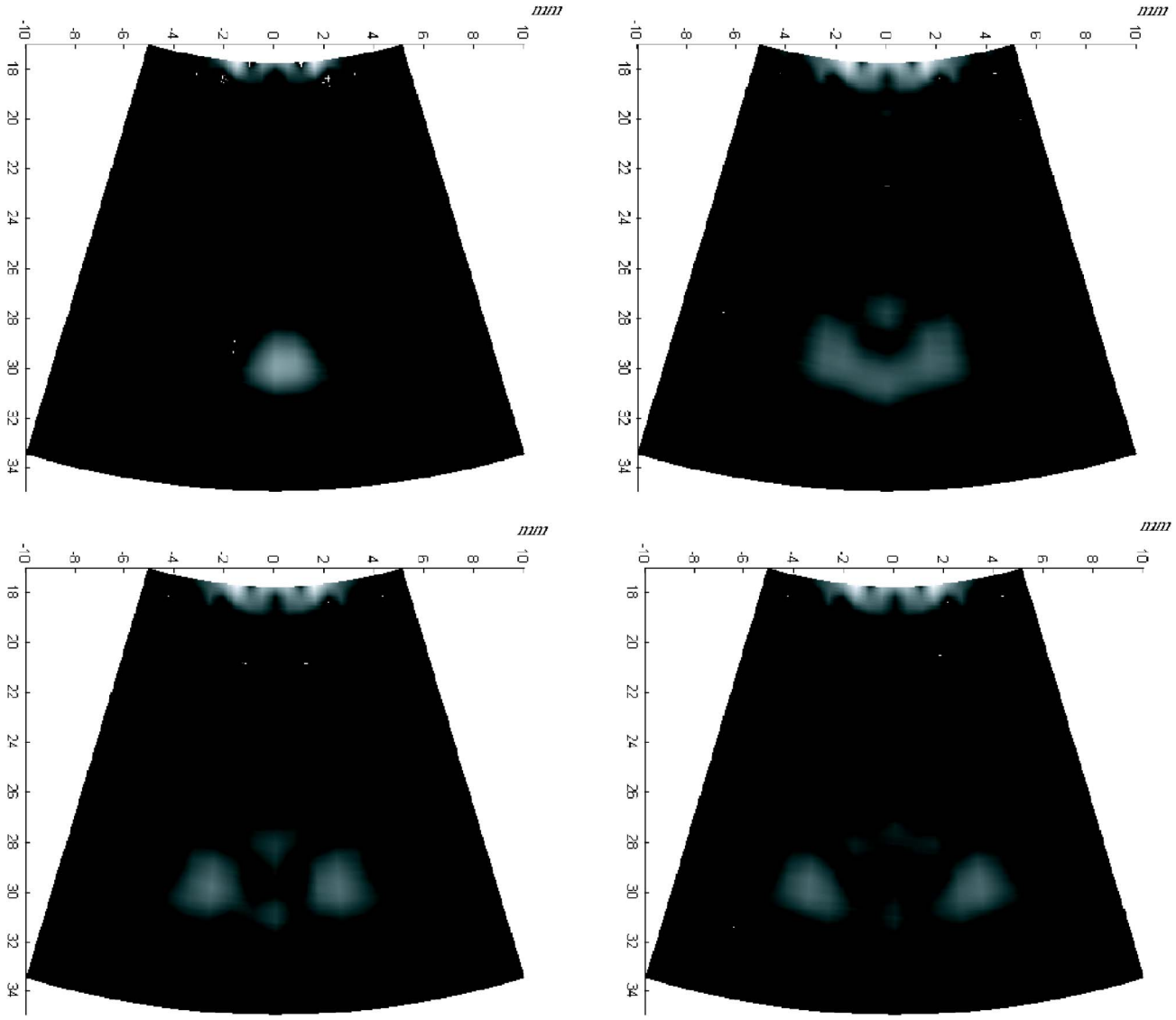


FIG. 3. Simulated B-scan images (f_{center}=0.67 MHz) resulting from scattering from two wires (diam=0.2 mm) separated by (a) 2 mm, (b) 4 mm, (c) 6 mm, and (d) 8 mm.

of the array. The received lines are then extended along their respective directions of propagation and combined to produce an image.

Simulation of the propagation and scattering is performed using a wave-vector space time domain method.[12–14] Related techniques have been described for propagation in inhomogeneous media,[15] including tissues.[16] These previously described methods involve elimination of a pressure gradient term, which is based upon simplifications of the wave equation laid out by Pourjavid and Tretiak.[17] The present algorithm follows a procedure similar to past work, but provides a full solution to the wave equation, given the pressure field at some initial time t_0 as well as the density and sound speed over all space. The starting point is time-dependent equivalent to Eq. (4)

$$\nabla^2 p - \frac{1}{c_0^2}\ddot{p} = \nabla \cdot [q_\rho \nabla p] - q_\kappa \ddot{p}. \tag{16}$$

The equation is simplified by assuming only changes in compressibility (q_ρ=0) and by defining $f(\mathbf{r},t)=q_\kappa(\mathbf{r})p(\mathbf{r},t)$. Equation (16) may be written in terms of the three-dimensional spatial Fourier transforms with respect to p and f, allowing it to be written as an inhomogeneous Helmholtz equation

$$\ddot{P} + K^2 c_0^2 P = \ddot{F}, \tag{17}$$

where capital letters represent respective Fourier transforms, and the wavenumber in Cartesian coordinates is given by $K^2=k_x^2+k_y^2+k_z^2$. The integral solution to (17) may be expressed using the one-dimensional (1D) Green's function

$$g(t|t_0) = \frac{e^{ic_0 K|t-t_0|}}{c_0 K}, \tag{18}$$

giving

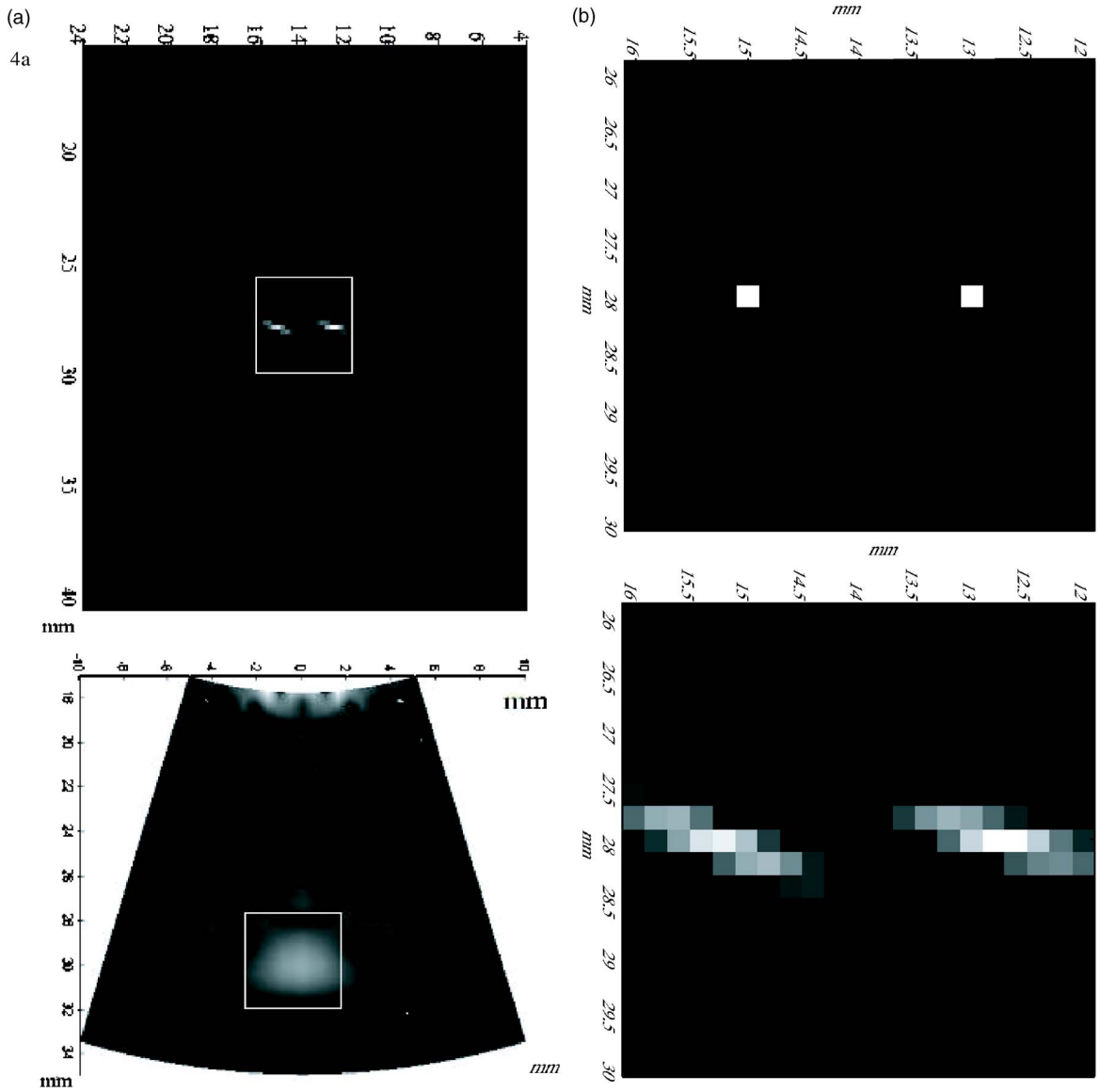


FIG. 4. (a) Comparison of the random frequency method (top) with a traditional B-scan (bottom) using the same frequency bandwidth. (b) A more magnified view compared with the actual position of the scatterers.

$$P(\mathbf{k},t) = P_i(\mathbf{k},t) + \int_0^\infty g(t|t')\ddot{F}(t')dt'. \tag{19}$$

The integral in (19) has the form of a convolution integral with respect to time. Using the property of the convolution integral

$$P_i \otimes_t \ddot{F} = \ddot{P}_i \otimes F, \tag{20}$$

the solution to the pressure in k-space may finally be expressed as

$$P(\mathbf{k},t) = P_i(\mathbf{k},t) - c_0 K \int_0^\infty g(t|t')F(t')dt'. \tag{21}$$

In general, F is not known but rather may be *grown* using an iterative method to step forward in time by some step Δt. Presently this is performed using Simpson's method in composite form.[18] Meanwhile the incident field P_i may be projected forward in time using the exponential transfer function, explained in detail in Ref. 19.

The summarized algorithm for calculating the B-scan image is as follows:

(1) Provide the initial fields $p(r,t_0)$ and $f(r,t_0)$;
(2) transform with respect to r to give $P(r,t_0)$ and $F(r,t_0)$;
(3) determine $P_i(t+dt)$ and $P_s(t+dt)$;
(4) inverse transform the field to give pfg, p_x, p_y, and p_z
(5) let $p=p_i+p_s$
(6) $f=p\gamma$
(7) go to step 2.

The time-step, dt, in step 3 may be varied, according to the complexity of the media and the location of the ultrasound beam. In the limiting case of a completely homogeneous media, dt may be arbitrarily long. An example of a field

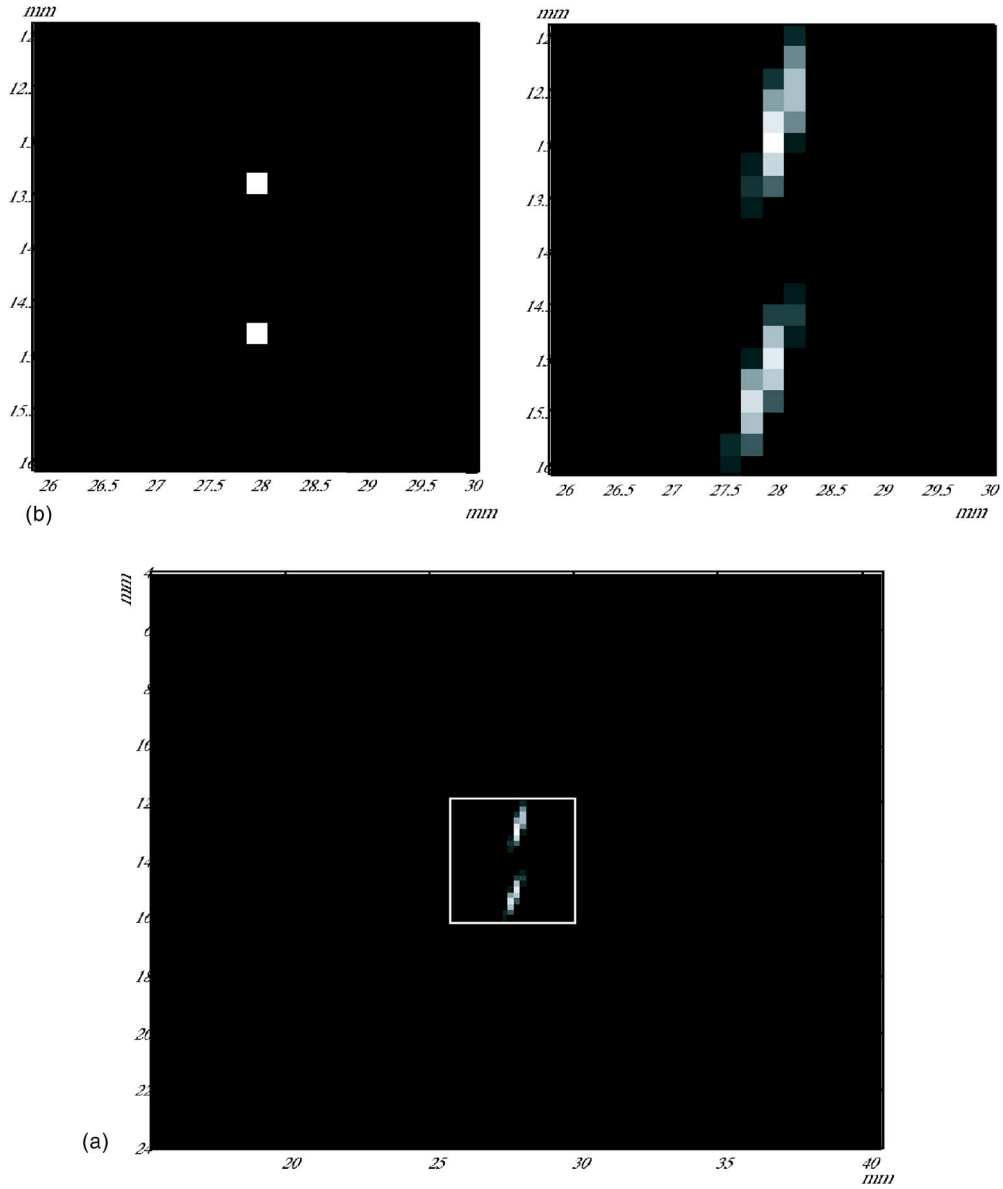


FIG. 5. (a) An image of two scatterers separated in the direction parallel to the transducer face at the resolution limit of 1.35 mm over the entire ROI and (b) a magnified view compared with the actual position of the scatterers.

propagating past two scatterers is provided in Fig. 2, representing a two-dimensional (2D) slice of the acoustic pressure taken out of the three-dimensional field at a given time.

IV. RESULTS

To simulate a B-scan, the acoustic impulse response directly in front of the array was calculated for a given phase-determined beam orientation. This impulse was then projected forward in time via the *k*-space projection algorithm using a 200 ns resolution. To produce one image it was necessary to repeat the simulation for each of the 21 scan directions. A planar region of interest having dimensions (40,40) mm, and with a spatial resolution of 0.2 mm was selected within the three-dimensional field volume in the plane $x=0$. Two 0.2 mm diameter wires were added to the ROI, representing the smallest linear scattering object that could be simulated under the current (0.2 mm) resolution. Both objects were given a sound speed of 3500 m/s. The distance between the wires was varied to represent separations of 1, 2, 4, 6, and 8 mm. The case of a single wire was also calculated.

Images resulting from the B-scans are presented in Fig. 3. In all figures, the transducer is located at the top of the figure, with the axis of forward propagation oriented downward. At 1 mm (not shown) and 2 mm separation [Fig. 3(a)], little difference was observed between these images and the images produced with the single wire. The images, which were constructed from the envelope of the amplitude of the backscattered signal, appeared as a single object with a diameter of approximately 2 mm, based on the full-width-at-half-maximum (FWHM) on the *y* and *z* axes. At 4 mm [Fig. 3(b), the two wires appeared as a single object elongated in the direction parallel with the transducer face. By 6 mm [Fig. 3(c)], two individual objects are discernable in the image, but with an on-axis artifact. At 8 mm [Fig. 3(d)], the objects are clearly separated.

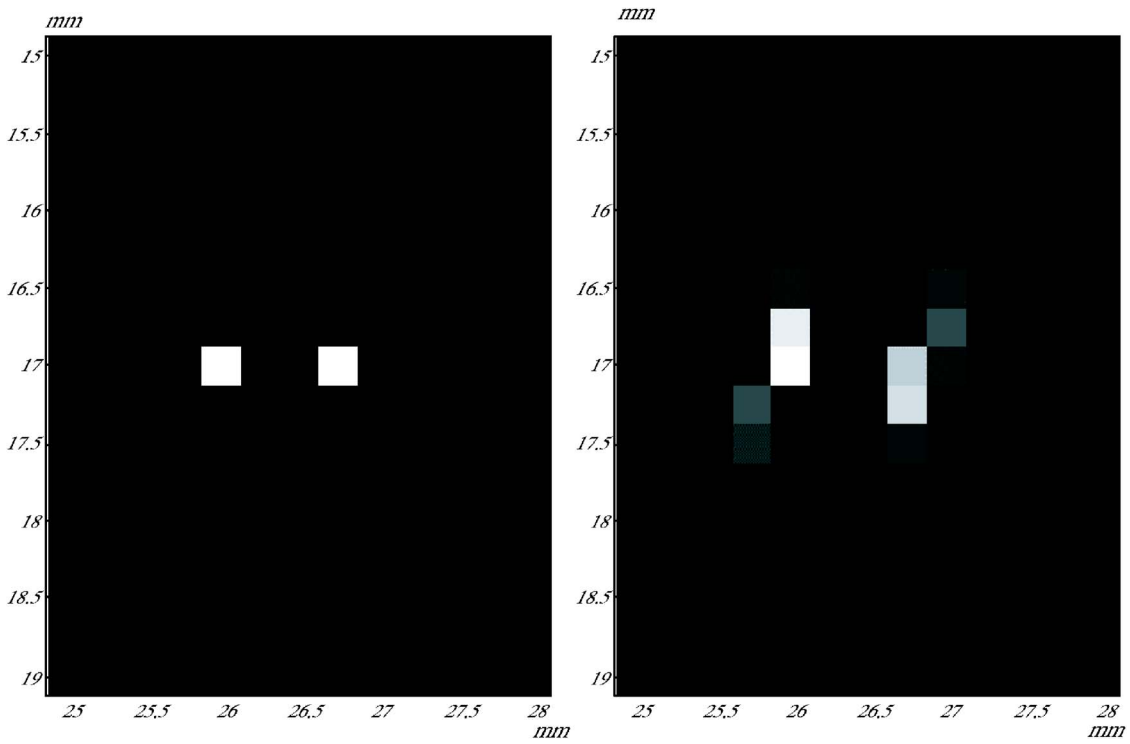


FIG. 6. (a) An image of two scatterers separated in the axial direction at the resolution limit of 0.5 mm over the entire ROI and (b) a magnified view compared with the actual position of the scatterers indicates the distortion observed at small separations.

Resolution limits for the spacing between wires was next examined with the randomized method. The transducer was identical in geometry and number of elements to that used in the B-mode simulations and all frequencies were selected within the B-mode bandwidth. Equation (15) was constructed with $M=15$ randomized signals. The field was calculated about the wires, using a spatial resolution of 0.2 mm. Based on initial trials, it was observed that a 2 mm separation of the objects, which appeared as a single object in the B-scan, produced two clearly discernable objects, which were also more localized than in the former case. This reconstruction is shown in Fig. 4. At 1 mm separation, the objects appeared as a single elongated object, similar to the effect that appeared below 4 mm in the B-scan. The smallest discernable separation for the present configuration was determined to be 1.35 mm (Fig. 5), below which the two objects appeared as one. A view over the entire ROI [Fig. 5(a)] indicates that artifacts are not present away from the scattering objects, while a magnified view [Fig. 5(b)] compares the reconstructed locations with the actual object placement.

Additional simulations were performed, varying the location of the two objects and their relative positions. Slightly better ability to separate objects was observed along the direction of propagation, where separations of 0.5 mm were observable, but with a distortion in the localization, making the objects appear approximately 0.75 mm apart (Fig. 6). This distortion was not apparent beyond 1 mm separation.

The ability to differentiate between two objects with different scattering strength was next examined. Two point objects (O_1 and O_2) with different speeds of sound were placed within the ROI and reconstructed. In all simulations, the objects were situated 4 mm apart and 17 mm from the transducer surface at locations (17, 8) and (17,11) mm, respectively, as indicated in Fig. 7(a). A series of reconstructions were performed for a range of O_1 sound speeds equal to 1517, 1857, 2143, 2428, 2427, and 3000 m/s, while the speed of sound of O_2 was set at a constant value of 1510 m/s. Figure 7(b) plots the scattering strength of O_1 relative to the constant O_2 ($q_2=0.51$). This plot is compared with the reconstructed image at the point (17,8) mm relative to the value at (17,11) mm. The plot indicates a similar trend between the scattering strength and the reconstruction, with increased distortion with higher sound speeds. The corresponding error ranged from 4.3% at 1857 m/sto 13% at 3000 m/s.

A series of reconstructions were then simulated in order to demonstrate the possibility of using the method to detect multiple objects and larger occlusions. Three such examples are provided, representing varying spatial configurations and distances from the source transducer face. The first example [Fig. 8(a)] consists of three scatterers placed diagonally within a relatively small ROI (8 mm$\times$8 mm) plus an additional scattering source situated approximately 4 mm from the others. Each object was given equal scattering strength ($q=0.1$). The reconstruction detected all objects with minimal spatial distortion [Fig. 8(b)], but with clear variation in the scattering intensity. The second example is presented in Fig. 8(c) as an inverted "V" situated approximately 18 mm from the ultrasound source. Here the objects were detected [Fig. 8(d)], but with blurring of the object and enhanced reflection at the extremities. The third example occurred at a mean distance of 27.3 mm from the source [Fig. 8(e)] consisting of a linear change in sound speed extending for 3 mm

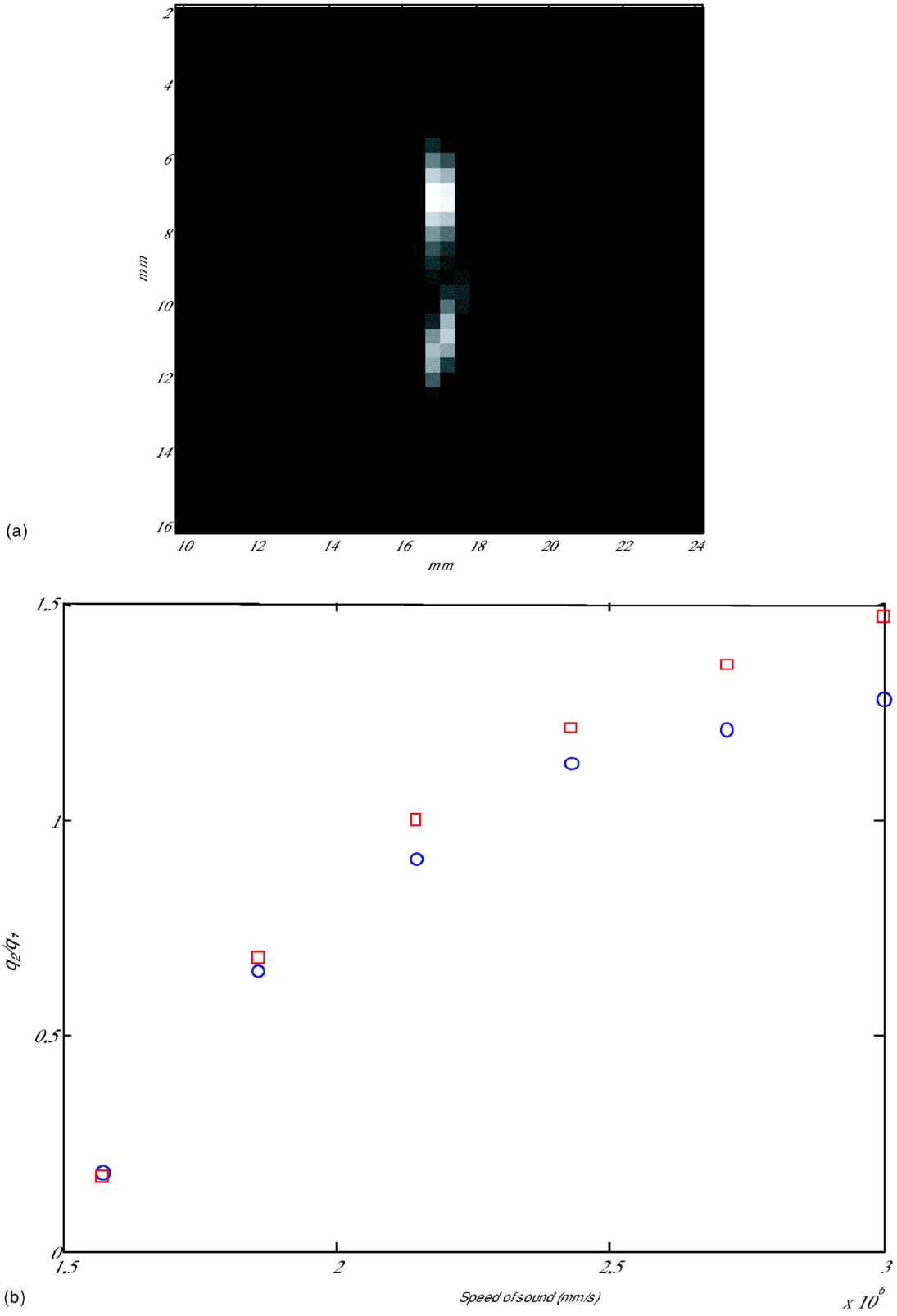


FIG. 7. (a) Two reconstructed objects in the ROI with scattering strengths q_1=0.51 and q_2=0.62. The speed of sound of q_2 is varied and the ratio of the measured scattering ratio of the two coefficients q_2/q_1 is plotted as a function of frequency (squares). The reconstructed values are compared with the actual (circles).

at a diagonal relative to the source. As with the previous example the line was blurred with a half-maximum intensity drop-off of under 1 mm, dimensionally consistent with the separation resolution between two objects.

V. DISCUSSION

The process of locating an object with a randomized and unfocused field seems, at first, counterintuitive. However, it

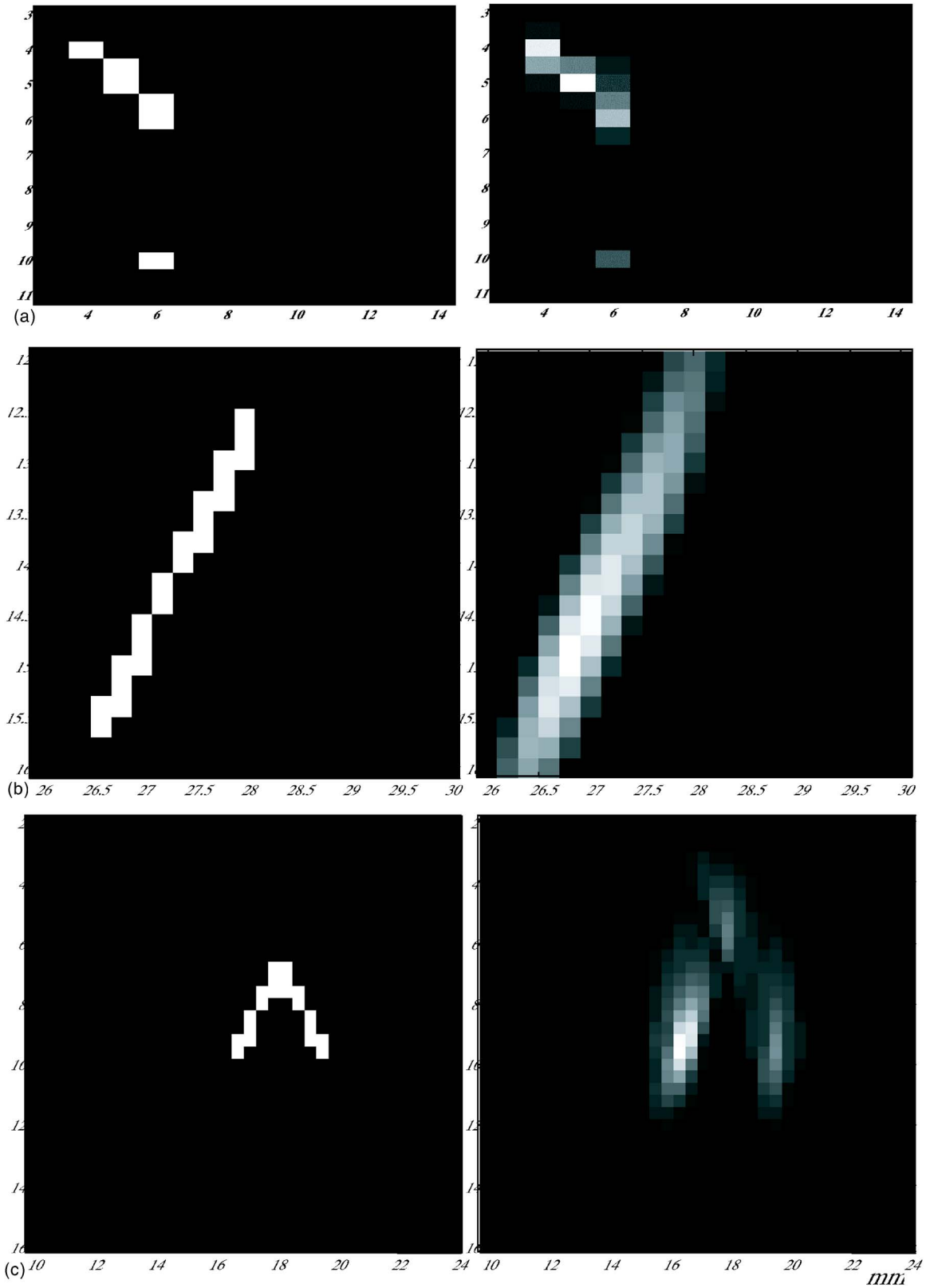


FIG. 8. A series of simulations with multiple scatterers (right column) and their reconstruction (left column) demonstrating application at distances of (a) 3–14 mm, (b) 10–24 mm, and (c) 26–30 mm in front of the transducer.

is in fact the large variation in the imaging field that makes it possible to localize the position of scattering sources. It is particularly interesting that the two-dimensional reconstruction described here can be achieved using only a single receive channel. In the preliminary numeric investigation, it was observed that objects were better defined and more spatially localized, as compared to a synthetic and idealized B-mode signal. The smallest objects considered were less than $0.1\times$ the center wavelength of the signal in the surrounding fluid.

Although in its present basic form, little more than speculation can be applied toward assessing utility to the method; the ability to detect objects would suggest that further investigation is warranted. It is stressed that although the theory was based on a point source/point receiver approximation, the simulations were full three-dimensional simulations, and considered the distortion introduced by finite size elements and radiation in all dimensions. Yet, the conditions displayed in this preliminary work are clearly idealistic, leaving much work still to be performed. For medical and biological application, experimental data is necessary in order to confirm the approach in actual tissues and with noise levels representative of *in vivo* situations.

Small objects, which can be hard to detect, and even harder to localize using present methods, may be both detectable and localizable with the random-frequency approach. Further, the approach acquires only a single channel of information, leaving the potential to acquire a large amount of data for creating time sequence imaging. One suggested medical use for such "M-mode" images is in conjunction with contrast agents to trace a single bubble along a vessel. While such images would probably not be possible in real time due to the processing time needed for image reconstruction, near-real-time motion image is conceivable. For example, the reconstruction phase of the algorithm typically took approximately 15 s, when benchmarked on a WINDOWSXP 64-bit operating platform controlling two dual-core 3 GHz Pentium Xenon processors.

Since the reconstruction approach analyzes changes in density and compressibility separately, there is a prospect of creating separate density and speed-of-sound images. In this configuration, two objects with identical impedances could be placed in the ROI: One varying from the surrounding medium in density and the second varying in speed-of-sound. Provided the objects are adequately separated, traditional backscatter imaging, which is sensitive to impedance mismatch, would display both objects with equal intensity on the one image. In contrast, the randomized frequency reconstruction offers the potential ability to distinguish between the two.

The ability to perform imaging at reduced frequencies may find particular utility in transcranial applications. Clinical methods are presently limited to imaging through the temporal bone acoustic window near the ear, which allows only a limited view of the brain. However, recent research has shown that ultrasound can be delivered through thicker bone and with reduced distortion using shear modes of propagation within the bone itself.[20] Further, reflected signals can return to the transducer in a similar manner and used to form ultrasound images.[21] Despite reduced beam refraction, absorption in shear propagation is a stronger function of frequency than its longitudinal counterpart,[22] and conditions at or below 1 MHz are significantly more favorable than existing transcranial probes that operate at or above 2 MHz. Development of lower frequency reconstruction methods will be necessary to create images with this approach.

It is not expected that hardware will limit the practicality of the method. The approach requires a broadband multichannel driving system that, in fact, reflects design trends which are increasingly utilizing low cost broadband digital arbitrary waveform circuitry in combination with analog amplification and filtering components to form multichannel amplifier systems. One such system, capable of over a thousand channels at 50 channels per circuit board has been described by Sokka *et al.*,[23] with a low cost $20/channel estimate for a broadband (0–10 MHz) system.

VI. CONCLUSIONS

This preliminary study demonstrates the process of two-dimensional object detection by a single receiving channel when using an unfocused field of randomized frequency sources. The algorithm was effective in detecting and separating 0.2 mm diam objects placed 1.5 mm from each other when using frequencies in the 0.25–1.25 MHz bandwidth. Based on the results, it is concluded that an array could be used for noninvasive monitoring in a variety of acoustic situations. Future experimental work will involve verification of the approach in the laboratory, and examine its ability to detect objects and create images in human tissues. Meanwhile much work is required in optimizing the method for a particular bandwidth, frequency and region of interest.

ACKNOWLEDGMENTS

This work was supported in part by Grant Nos. R21EB004353 and U41RR019703 of the National Institutes of Health.

[1]D. A. Knapik, B. Starkoski, C. J. Pavlin, and F. S. Foster, "A 100–200 MHZ ultrasound biomicroscope," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **47**, 1540–1549 (2000).
[2]M. P. Spriet, C. A. Girard, S. F. Foster, K. Harasiewicz, D. W. Holdsworth, and S. Laverty, "Validation of a 40 MHz B-scan ultrasound biomicroscope for the evaluation of osteoarthritis lesions in an animal model," Osteoarthritis Cartilage **13**(2), 171–179 (2005).
[3]K. A. Snook, C.-H. HU, T. R. Shrout, and K. K. Shung, "High-frequency ultrasound annular-array imaging. Part I: Array design and fabrication," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **53**(2), 300–308 (2006).
[4]M. Ueda and T. Sato, "Superresolution by holography," J. Opt. Soc. Am. **61**(3), 418–419 (1971).
[5]M. A. Fink, "Time-reversed acoustics," Phys. Today **20**, 34–40 (1997).
[6]G. T. Clement, J. Huttunen, and K. Hynynen, "Superresolution ultrasound imaging using back-projected reconstruction," J. Acoust. Soc. Am. **118**, 3953–3960 (2005).
[7]G. T. Clement and K. Hynynen, *Superresolution Ultrasound for Imaging and Microscopy, 2004 IEEE Ultrasonics Symposium Proceedings* (2005).
[8]P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (Princeton University Press, Princeton, New Jersey, 1968).
[9]B. A. Lippmann and J. Schwinger, "Variational principles for scattering processes I," Phys. Rev. **79**, 469–480 (1950).
[10]W. Tobocman, D. Driscoll, N. Shokrollahi, and J. A. Izatt, "Free of speckle ultrasound images of small tissue structures," Ultrasonics **40**(9), 983–996 (2002).
[11]R. K. Saha and S. K. Sharma, "Validity of a modified Born approximation for a pulsed plane wave in acoustic scattering problems," Phys. Med. Biol. **50**(12), 2823–2836 (2005).
[12]N. N. Bojarski, "The k-space formulation of the scattering problem in the time domain," J. Acoust. Soc. Am. **72**(2), 570–584 (1982).
[13]N. N. Bojarski, "The k-space formulation of the scattering problem in the time domain: An improved single propagator formulation," J. Acoust. Soc. Am. **77**(3), 826–831 (1985).
[14]G. T. Clement, R. Liu, S. V. Letcher, and P. R. Stepanishen, "Temporal backward planar projection of acoustic transients," J. Acoust. Soc. Am. **103**(4), 1723–1726 (1998).
[15]C. J. Vecchio, M. E. Schafer, and P. A. Lewin, "Prediction of ultrasonic field propagation through layered media using the extended angular spec-

trum method," Ultrasound Med. Biol. **20**(7), 611–622 (1994).
[16]T. D. Mast, L. P. Souriau, D.-L. D. Liu, M. Tabei, A. I. Nachman, and R. C. Waag, "A k-space method for large-scale models of wave propagation in tissue," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **48**(2), 341–354 (2001).
[17]S. Pourjavid and O. J. Tretiak, "Numerical solution of the direct scattering problem through the transformed acoustical wave equation," J. Acoust. Soc. Am. **91**(2), 639–645 (1992).
[18]E. W. Weisstein, *Simpson's Rule. From MathWorld—A Wolfram Web Resource* (CRC Press LLC, Boca Raton, FL, 2006).
[19]G. T. Clement and K. Hynynen, "Field characterization of therapeutic ultrasound phased arrays through forward and backward planar projection," J. Acoust. Soc. Am. **108**(1), 441–446 (2000).
[20]G. T. Clement, P. J. White, and K. Hynynen, "Enhanced ultrasound transmission through the human skull using shear mode conversion," J. Acoust. Soc. Am. **115**(3), 1356–1364 (2004).
[21]G. T. Clement, "Spectral image reconstruction for transcranial ultrasound measurement," Phys. Med. Biol. **50**, 5557–5571 (2005).
[22]P. J. White, G. T. Clement, and K. Hynynen, "Longitudinal and shear mode ultrasound propagation in human skull bone," Ultrasound Med. Biol. **32**(7), 1085–1096 (2006).
[23]S. D. Sokka, J. Juste, and K. Hynynen, "Design and evaluation of a broadband multi-channel ultrasound driving system for large scale therapeutic phased arrays," *2003 IEEE Ultrasonics Symposium Proceedings*, 1638–1640 (2004).

Microbubble spectroscopy of ultrasound contrast agents

Sander M. van der Meer and Benjamin Dollet
Physics of Fluids Group, Department of Science and Technology, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands

Marco M. Voormolen, Chien T. Chin,[b] Ayache Bouakaz,[c] and Nico de Jong[d]
Department of Experimental Echocardiography, Erasmus MC, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands

Michel Versluis[a] and Detlef Lohse
Physics of Fluids Group, Department of Science and Technology, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands

(Received 13 July 2006; revised 16 October 2006; accepted 16 October 2006)

A new optical characterization of the behavior of single ultrasound contrast bubbles is presented. The method consists of insonifying individual bubbles several times successively sweeping the applied frequency, and to record movies of the bubble response up to 25 million frames/s with an ultrahigh speed camera operated in a segmented mode. The method, termed microbubble spectroscopy, enables to reconstruct a resonance curve in a single run. The data is analyzed through a linearized model for coated bubbles. The results confirm the significant influence of the shell on the bubble dynamics: shell elasticity increases the resonance frequency by about 50%, and shell viscosity is responsible for about 70% of the total damping. The obtained value for shell elasticity is in quantative agreement with previously reported values. The shell viscosity increases significantly with the radius, revealing a new nonlinear behavior of the phospholipid coating.
 [DOI: 10.1121/1.2390673]

PACS number(s): 43.80.Qf, 43.80.Vj, 43.35.Ei [CCC] Pages: 648–656

I. INTRODUCTION

Medical ultrasound imaging is based on scatter and reflection of sound from inhomogeneities in the tissue.[1] The scatter from blood is much weaker than the scatter from tissue. To increase the scattering properties from the blood pool, an ultrasound contrast agent (UCA) is introduced in the blood. An UCA is a liquid, containing small encapsulated microbubbles, which very efficiently scatter ultrasound.[2–4] In this way, it is possible to visualize and quantify the perfusion of tissue, like for instance the heart muscle, liver, or kidney. Contrast agents are nowadays used in various medical investigations. e.g., in obtaining diagnostic information from the volume and shape of the heart ventricles, or to quantify the perfusion of various organs, like liver or kidney.

The fundamental understanding of the dynamics of contrast bubbles is a field of ongoing research. For example, the quantification of the response of contrast bubbles to ultrasound is an important research aspect. Until now, bubbles are characterized mainly by studying acoustically a representative sample of the UCA, containing many microbubbles.[5–7] From this data the overall resonance behavior of the sample can be deduced. Ideal contrast agents would be monodisperse in size, but in practice they have a size distribution which can be measured with, e.g., a Coulter counter, resulting in a mean size and size range.[8] For SonoVue™ and also for BR-14 (Bracco SA, Geneva), e.g., the mean radius is 1.5 μm, with 95% of the bubbles smaller than 10 μm. The polydispersity of the microbubbles makes it difficult to extract information on the physical properties of single bubbles, since the acoustical response of a bubble strongly depends on its size.[8] Furthermore, the acoustic pressure signal emitted by the bubbles is distorted by frequency dependent scattering and attenuation. On the other hand, measuring the acoustic response of a single individual contrast bubble is a difficult task.[9] First, it is difficult to isolate a single bubble in the focal region of a transducer: this would require at least 1 mm of distance between the bubbles. Second, extracting the absolute pressure emitted by the bubble from the measured response requires an accurate calibration of the transducer transfer function.

To overcome the difficulties associated with acoustical characterization, optical methods have been proposed.[10–14] Such methods are based on the direct measurement of the bubble radius, which, unlike the acoustical response, is not subject to distortion and in principle does not require difficult calibration. Furthermore, the interaction between bubbles (secondary Bjerknes forces) decays as the inverse square of their distance, fast enough to consider different bubbles to oscillate independently as soon as they are separated by a few (roughly ten) bubble radii. Isolating the response of a single bubble optically is thus less constraining than for

[a]Author to whom correspondence should be addressed. Electronic mail: m.versluis@utwente.nl
[b]Present address: Philips Research, Briarcliff Manor, NY.
[c]Present address: Inserm U619, B1A, CHU Bretonneau, 2 bd Tonnell, 37044 Tours Cedex, France.
[d]Also affiliated with: Physics of Fluids Group, Department of Science and Technology, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands.

acoustical measurements. However, optical methods exhibit other drawbacks: very high frame rates are required to resolve microbubble oscillations at several MHz, and the resolution is limited, since the microbubble size is just a little higher than the optical resolution.

Here, we present a new optical method, that we term bubble spectroscopy, to characterize individual contrast bubbles. To resolve the oscillations of such bubbles, we use the ultrahigh speed camera Brandaris,[15] used in a segmented mode described in detail in Sec. II: we scan the insonation frequency to reconstruct a resonance curve, from which we extract the resonance frequency and the total damping coefficient. The experimental methods are detailed in Sec. III. In Sec. IV, we present the main results: we quantify the change of the resonance frequency with the radius, and discuss the influence of shell elasticity. We also quantify damping and show the influence of shell viscosity. We further discuss the accuracy of the proposed method in Sec. V.

II. BUBBLE SPECTROSCOPY

A. The microbubble as a linear oscillator

For small enough acoustic forcing, it is well known[16,17] that a bubble behaves as a linear oscillator; its relative radial excursion x, defined as $R=R_0(1+x)$, obeys the equation

$$\ddot{x} + \omega_0 \delta \dot{x} + \omega_0^2 x = F(t), \tag{1}$$

with $f_0=\omega_0/2\pi$ the eigenfrequency of the system and δ its (linear) dimensionless damping coefficient (equivalently, one can define the quality factor $Q=1/\delta$), and F is the forcing term.

The amplitude of the radial variation of the bubble depends of the driving frequency $f=\omega/2\pi$. Writing $F(t)=F_0 \sin \omega t$ and $x(t)=x_0 \sin(\omega t+\varphi)$, one gets from Eq. (1):

$$x_0(\omega) = \frac{F_0}{\sqrt{(\omega_0^2-\omega^2)^2 + (\delta\omega\omega_0)^2}}. \tag{2}$$

This equation defines the resonance curve, displaying a maximum at the resonance frequency

$$f_{\mathrm{res}} = f_0 \sqrt{1-\frac{\delta^2}{2}}, \tag{3}$$

which is lower than the eigenfrequency in the presence of damping. Strictly speaking, Eq. (3) holds only if the damping coefficient δ is independent of ω. The main objective of our bubble spectroscopy method is to fully characterize the linear response of single bubbles, by constructing its resonance curve and extracting from that the eigenfrequency and the damping coefficient. We show this on a simulation example in the following subsection, and on experiments in Sec. III.

B. Simulation example: Power spectrum and resonance curve

As an example, we derive the resonance frequency and the damping coefficient from a numerical simulation. We compute the time evolution of the radius of a coated bubble subjected to an acoustic pressure, using the following model, adapted from Marmottant *et al.*[18]

$$\rho_\ell \left(R\ddot{R} + \frac{3}{2}\dot{R}^2 \right) = \left[P_0 + \frac{2\sigma_w}{R_0} \right] \left(\frac{R}{R_0} \right)^{-3\gamma} \left(1 - \frac{3\gamma}{c}\dot{R} \right) - P_0 - \frac{2\sigma_w}{R} - 4\chi \left(\frac{1}{R_0} - \frac{1}{R} \right) - \frac{4\mu\dot{R}}{R} - \frac{4\kappa_s \dot{R}}{R^2} - P(t), \tag{4}$$

where R, $\dot{R}$, and $\ddot{R}$ represent the radius, velocity, and acceleration of the bubble wall (R_0 being the equilibrium radius), $\rho_\ell=10^3$ kg/m^3 the volumetric mass of water, $P_0=10^5$ Pa is the ambient pressure and $P(t)$ the driving acoustic pressure, and γ the polytropic exponent; since the oscillations are fast, $\mathrm{Pe}=R_0^2\omega/D_{\mathrm{th}} \gg 1$[18] for bubbles of several microns in size in the MHz regime (here, $D_{\mathrm{th}}=2\times10^{-6}$ m^2/s is the thermal diffusivity of C_4F_{10}), we assume that this exponent equals the ratio of specific heats, $\gamma=1.07$ for C_4F_{10}. Furthermore, $c=1.5\cdot10^3$ m/s is the speed of sound in the fluid, $\sigma_w=0.072$ N/m the surface tension, μ the dynamic viscosity, and $P(t)$ is the acoustic pressure. Equation (4) is based on the Rayleigh-Plesset equation, commonly used to model the behavior of bubbles (see Refs. 19–21 for general reviews on this subject) with two additional parameters to model the shell: an elasticity parameter χ (in N/m), and a shell viscosity κ_s (in kg/s).[27] Various models including the shell properties in the Rayleigh-Plesset equation have already been proposed to model contrast agent bubbles.[13,18,22–26] Most of these models[13,23–25] consider a shell of finite thickness, modeled as a three-dimensional continuous medium, which may not be satisfactory for a monolayer shell. This is why Eq. (4) is inspired from models considering the shell as a two-dimensional viscoelastic medium.[18,26] More precisely, Eq. (4) is closely related to the model of Marmottant *et al.*,[18] who suggested modeling shell elasticity through a radius-dependent surface tension over a certain range of radii (elastic range), below which the bubble buckles and above which the shell breaks. Basically, Eq. (4) would correspond to an infinite elastic range, which is relevant in this study, since we use low enough acoustic pressures to avoid both buckling (associated with nonspherical oscillations) and rupture (which would lead to fast bubble dissolution). Equation (4) is also related to the model of De Jong[22] by a more physical description of radiation and viscous damping. For the sake of simplicity, we account this model for thermal damping as an effective viscosity, taking $\mu=2\times10^{-3}$ Pa s, thus twice that of water.

Linearization of Eq. (4) yields the following eigenfrequency in the elastic regime:

$$f_0 = \frac{1}{2\pi} \sqrt{\frac{1}{\rho R_0^2} \left[3\gamma P_0 + \frac{2(3\gamma-1)\sigma_w}{R_0} + \frac{4\chi}{R_0} \right]}. \tag{5}$$

This eigenfrequency has two contributions: the Minnaert frequency,[28]

$$f_M = \frac{1}{2\pi} \sqrt{\frac{1}{\rho R_0^2} \left[3\gamma P_0 + \frac{2(3\gamma-1)\sigma_w}{R_0} \right]}, \tag{6}$$

i.e., the eigenfrequency of an uncoated bubble, and a shell contribution which increases the eigenfrequency.

The linearization of Eq. (4) also gives the expression of the total damping coefficient: $\delta_{\text{tot}}=\delta_{\text{rad}}+\delta_{\text{vis}}+\delta_{\text{shell}}$, with a contribution coming from the sound reradiated by the bubble, which writes at $\omega=\omega_0$[16]

$$\delta_{\text{rad}}=\frac{\omega_0 R_0}{c}, \tag{7}$$

a viscous contribution

$$\delta_{\text{vis}}=\frac{4\mu}{R_0^2\rho\omega_0}, \tag{8}$$

and a shell viscosity contribution

$$\delta_{\text{shell}}=\frac{4\kappa_s}{R_0^3\rho\omega_0}. \tag{9}$$

To be rigorous, a direct derivation from Eq. (4) gives an expression of δ_{rad} slightly different than the classical expression (7), but the order of magnitude remains the same, and we will see in Sec. IV B that radiation is a secondary contribution to the total damping. We will thus keep the expression (7) for the radiation damping.

We compute Eq. (4) with a bubble of ambient radius $R_0=2.8$ μm. We numerically solve the bubble response $R(t)$ to an ultrasound burst of eight cycles, whose two first and two last cycles are modulated by a Gaussian envelope, as in experiments. The acoustic amplitude is $P_a=1$ kPa to minimize nonlinear effects, and the frequency range was chosen around the estimated resonance frequency, from 1.5 to 2.5 MHz. Next, we apply a fast Fourier transform algorithm on each $R(t)$ curve to compute its Fourier transform, and we take the square of this quantity: we get thus the power spectrum P_R of the radius-time curve. Typical $R(t)$ curves and their corresponding power spectra are shown in Fig. 1. At resonance, more energy goes into the oscillation than off-resonance. We quantify this effect by taking the area in the power spectrum in a band of $\Delta f=175$ kHz around the maximum frequency $f_{\max}$, which we term the response:

$$\text{Re}=\int_{f_{\max}-\frac{1}{2}\Delta f}^{f_{\max}+\frac{1}{2}\Delta f} P_R df.$$

The choice for a value of 175 kHz was found to be a suitable bandwidth for our data analysis. The total area in the power spectrum of the signal is an equivalent measure for the resonance, because nearly all the area in the power spectrum of the signal is located in the fundamental peak. We decided to take the area in a band around the peak, and we checked that the results are indeed in accordance.

We now redetermine the eigenfrequency and the total damping out of the resonance curve, in order to establish and verify the method we want to apply to the experimental data of Sec. III. To do so, we fit the data points to the response of a harmonic oscillator [Eq. (2)], which we rewrite as

$$\text{Re}(f)=\frac{\text{Re}_0}{(1-f^2/f_0^2)^2+(\delta f/f_0)^2}. \tag{10}$$

From the best fit of the data points (shown in Fig. 2 for our simulation example), we extract the eigenfrequency f_0^{fit}

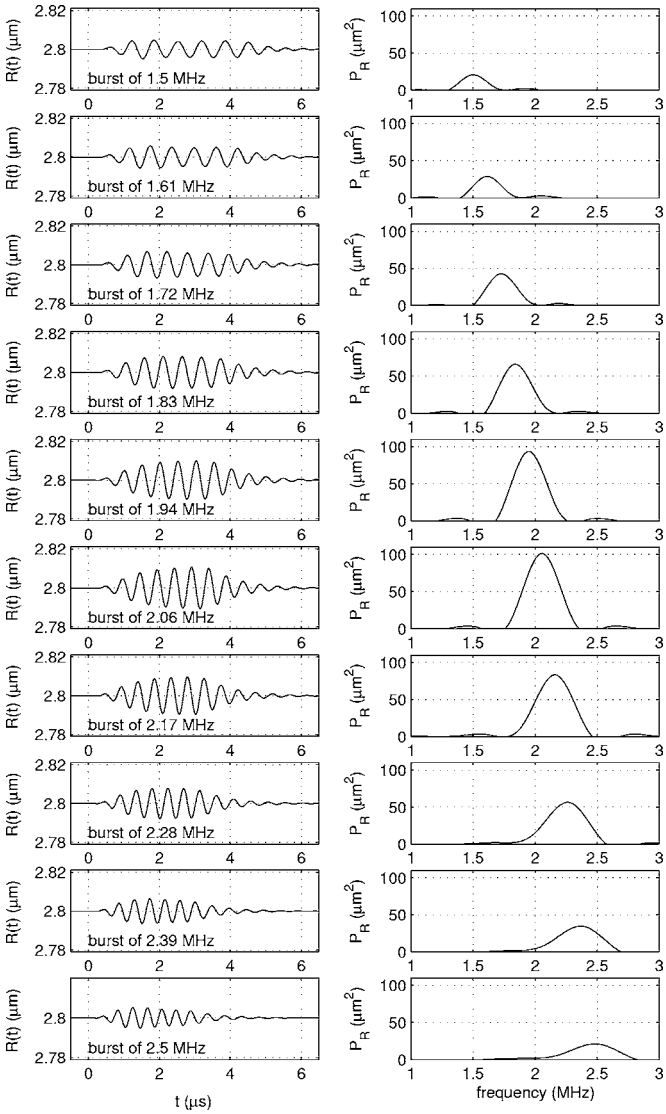


FIG. 1. Simulated response of a bubble of initial radius $R_0=2.8$ μm to an ultrasound wave of 1 kPa of different frequencies. Left: radius-time response; right: power spectrum. The resonance frequency is at 2.07 MHz.

$=2.07$ MHz and the damping coefficient $\delta_{\text{tot}}^{\text{fit}}=0.24$. We will motivate why we analyze the power spectrum, and not directly the radial oscillations, in Sec. III B.

C. Analysis of the resonance curve

From the resonance curve obtained in the simulation example, we can determine the position where the amplitude of the oscillation is maximum, i.e., the resonance frequency: $f_{\text{res}}^{\text{curve}}=2.02$ MHz. The best fit curve to the simulation data displays a resonance frequency compatible within 1% error: $f_{\text{res}}^{\text{fit}}=2.04$ MHz. The resonance frequency is lower than the eigenfrequency $f_0^{\text{fit}}=2.07$ MHz because of damping, but the shift remains very small. The resonance frequencies $f_{\text{res}}^{\text{curve}}$ and $f_{\text{res}}^{\text{fit}}$ fully agree with the theoretical estimate (Table I). Second, the peak width is directly related to the total damping of the system δ_{tot}. A sharp peak indicates low damping, whereas a broad peak indicates high damping. More precisely, the width Δf of the response peak at half the maximum amplitude obeys: $\Delta f/f_0=\delta_{\text{tot}}$. Measuring the width of

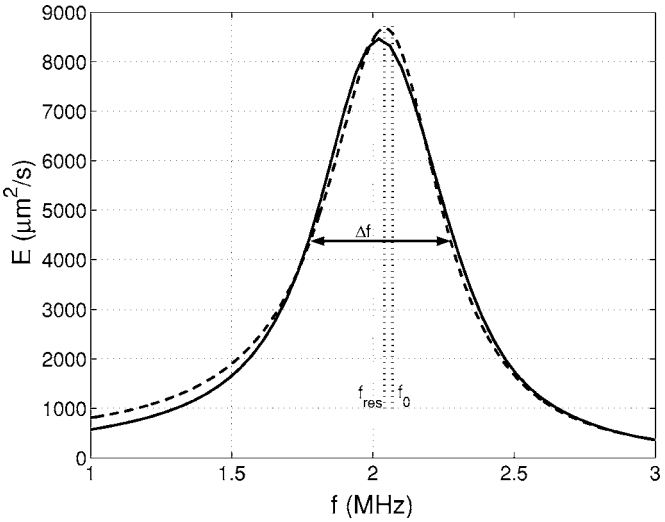


FIG. 2. The simulated response (solid line) of a free 2.8 μm gas bubble on ultrasound bursts of 1 kPa, ranging in frequency from 1 to 3 MHz. The dashed curve is the fit to a harmonic oscillator [Eq. (10)] giving $f_0^{\text{fit}}=2.07$ MHz and $\delta_{\text{tot}}^{\text{fit}}=0.24$.

the peak in Fig. 2 gives the following value for the damping coefficient: $\delta_{\text{tot}}^{\text{curve}}=0.25$, in excellent agreement with the theoretical value: $\delta_{\text{tot}}^{\text{th}}=0.24$ and the value obtained from fit: $\delta_{\text{tot}}^{\text{fit}}=0.24$. This excellent agreement between the theory, the simulation curve, and the fitting curve (Table I) shows that our fitting procedure on the oscillation response is an accurate method to extract both the eigenfrequency and the damping coefficient.

For a linear oscillator, the phase lag φ between the forcing term and the oscillator response is, according to Eq. (1),

$$\tan\varphi=\frac{\delta_{\text{tot}}}{\frac{f_0}{f}-\frac{f}{f_0}}. \tag{11}$$

If the oscillator is driven at frequencies well below resonance, it is in phase with the driving force ($\varphi=0$). As the frequency is increased towards resonance, the displacement tends to lag behind the driving, so that at resonance the displacement has a phase shift of $\pi/2$. When the oscillator is driven at frequencies much greater than resonance, the displacement is in antiphase with the driving force ($\varphi=\pi$). In principle, also measuring the phase difference between ultrasound driving and bubble response could be used to determine the resonance frequency of contrast bubbles from experimental data. However, this requires precise timing. Therefore, presently we focus on the oscillation response.

TABLE I. Comparison of the eigenfrequency f_0, the damping coefficient δ_{tot}, and the resonance frequency f_{res} [related through Eq. (3)], from the theory [Eqs. (5)–(9)], the simulation curve, and its fitting curve (Fig. 2).

	f_0 (MHz)	δ_{tot}	f_{res} (MHz)
Theory	2.06	0.24	2.03
Simulation curve	⋯	0.25	2.02
Fitting curve	2.07	0.24	2.04

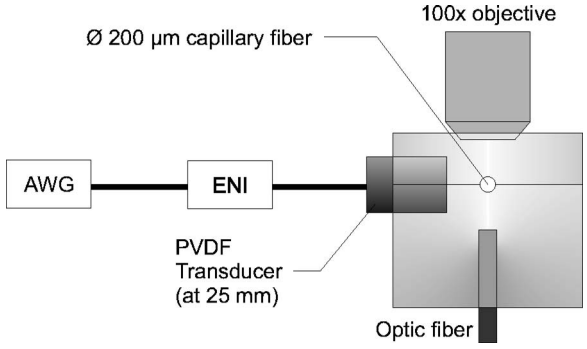


FIG. 3. The experimental setup: an arbitrary waveform generator produces ultrasound signals that are amplified by an amplifier and led to a transducer. Contrast bubbles are injected through a 200 μm capillary fiber perpendicular to the plane of the figure. The contrast bubbles are imaged from the top through a 100× objective and illumination is provided from the bottom.

III. EXPERIMENTS

A. The setup

The experimental setup is schematically drawn in Fig. 3. A dilute solution of individual BR-14 contrast bubbles (Bracco, Geneva) is prepared and injected through a capillary fiber of 200 μm diameter immersed in water. The bubbles are illuminated from below with an optical fiber and an image is produced by an Olympus microscope with a 100× water-immersed objective and a 2× magnifier. We carefully check that only single bubbles are present in the field of view of the microscope. The image is relayed onto a charge coupled device (CCD) camera for orientation and size estimation, and simultaneously onto the Brandaris high-speed camera.[15] The camera can record six movies of 128 frames at up to 25 million frames/s. Furthermore, the camera was designed to operate in a segmented mode, in fact specifically to accommodate microbubble spectroscopy. In practice the conventional single acquisition of 128 frames was replaced by recording two segments of 64 frames each, or four segments of 32 frames each. The camera houses memory space for six conventional acquisitions of 128 frames, before the images are transferred to the PC. Using two segments, this procedure results in the recording of 12 sets of 64 frames. Using four segments has the advantage of an increased frequency resolution (24 instead of 12), but reduces the sampling of the movies from 64 to 32 frames. The camera is operated at a framing rate of 15 million frames/s. The segmented mode allows us to construct a resonance curve of the bubble in a single acquisition, in less than 1 s.

The experiments described here were always done within 6 h after the preparation of the contrast bubbles. The bubbles were introduced through the capillary fiber in the focus of the microscope. Once we identified a single bubble, we estimated its radius from the images of the CCD camera, and estimated its approximate resonance frequency through the Minnaert equation [Eq. (6)], taking into account that the shell elasticity shifts the resonance frequency up. The bubble was then subjected to a scan of 12 (or 24) different frequencies, in a range of roughly 1 MHz below and above the expected resonance frequency. The contrast bubbles were insonified from the side by a broadband single element

transducer (Precision Acoustics, PA081) with a center frequency of 1.7 MHz and a calibrated range of frequencies from 0.7 to 6 MHz. An arbitrary waveform generator (AWG), a Tabor 8026, connected to a PC, was used to produce the required waveforms, which were then amplified by an ENI 350L amplifier. The length of the ultrasound waveforms was eight cycles, of which the first two and the last two were tapered taking a Gaussian envelope. The bubbles were investigated with sequential bursts of the ultrasound waveforms, with an acoustic pressure kept as low as possible to minimize nonlinear responses. In this paper, we present only results at driving pressures lower than 40 kPa. This prevented us from studying bubbles smaller than 1.5 μm in radius, since these bubbles do not oscillate significantly at these small pressures.[40] The pressures generated by the broadband single element transducer were calibrated with a 0.2 mm needle hydrophone (Precision Acoustics, SN1033). To maintain a constant pressure in the focus, we compensated for the frequency-dependent response of the transducer by adjusting the amplitude of the waveforms of different frequencies. The programming of the waveforms was done in MATLAB. The waveforms were transferred to the AWG via a general purpose interface bus (GPIB).

B. Radius-time curves determination

From the images, we extract the radius-time information. The radius-time curves of individual bubbles were measured using a so-called dynamic programming algorithm.[29] The center of the bubble of interest is annotated in the first frame of the recording. This center point was then used to radially resample the bubble and its direct surrounding. The resulting image was used as an input to the algorithm to find the optimal path along the contour of the bubble. After transforming the contour back into the recorded frame the average bubble radius and corresponding center are determined. This center point is then used to repeat the above procedure for the next frame. After running through all frames the radius-time curve of the bubble is obtained. Through a calibration grid the conversion between pixels and micrometers is performed.

To quantify the amplitude of oscillations of a bubble, the simplest method would be to find the maximum and minimum radial excursion during the insonation. However, in experiments, this proved not to be the most accurate measurement for several reasons. First of all, due to the finite sampling frequency of the signal, there is a difference between the recorded extrema and their actual values. Furthermore, the extrema need to be determined from only a few cycles of the bubble oscillation. Second, in experiments an off-resonance oscillating bubble often shows an amplitude overshoot in the first few cycles. Although we tried to minimize this by using a Gaussian envelope, it is somehow arbitrary to choose which of the extrema shows the real amplitude for that particular frequency and pressure. For all these reasons, we rather work on the Fourier transform of the radius-time curve. It also presents a maximum amplitude at resonance, and since it is an integrated quantity over the full $R(t)$ signal, it is much less sensitive to the sampling rate and

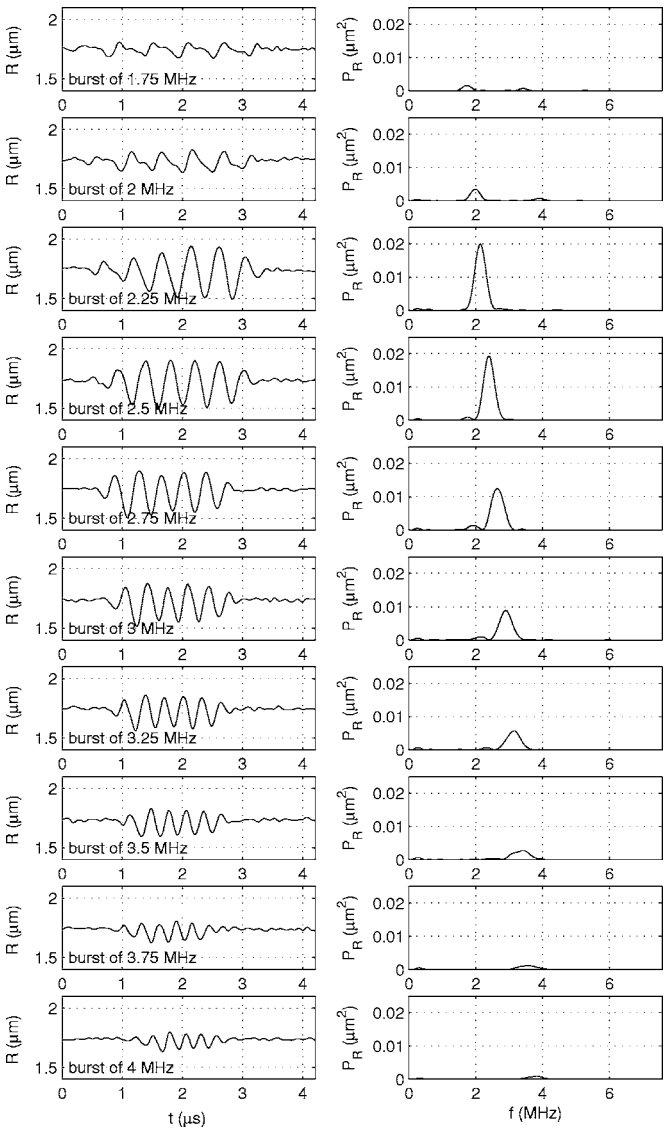


FIG. 4. Experimental radius-time curves (left column), and corresponding power spectra (right column), for a bubble with an ambient radius of 1.7 μm during a scan of insonifying frequencies.

to the short transients and overshoots of the $R(t)$ curve. In the following section we describe the subsequent analysis.

C. Data processing

From the images, the radius-time curves for each individual bubble were measured for each frequency component. An example of such a curve is shown in Fig. 4.

We treat the radius-time curves as in Sec. II B: for each scanned frequency, we calculate the area in the power spectrum in a band around the maximum frequency, as the quantifier of the amplitude of bubble oscillation (Fig. 4). Plotting the results yields an experimental resonance curve which we fit to the linear oscillator expression [Eq. (10)]. Three examples of such resonance curves are displayed in Fig. 5. We then record for each experiment the two following fitting parameters of Eq. (10): the eigenfrequency f_0 and the damping coefficient δ_{tot}. After careful selection (correct pressure, no significant shrinking of the bubble by loss of gas during insonation), we present 22 experimental data points.

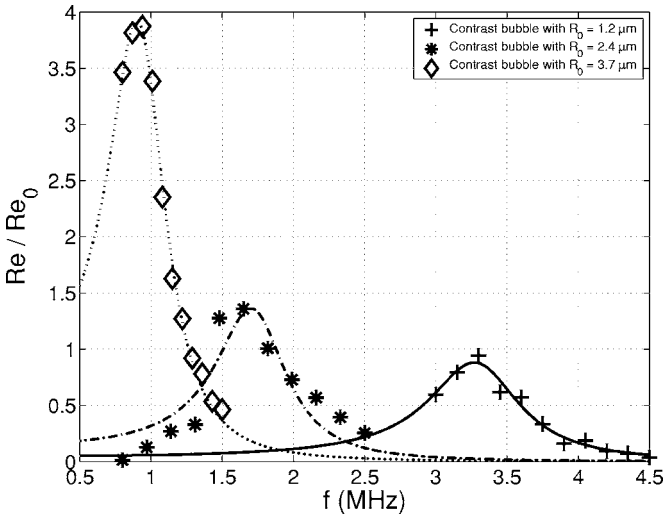


FIG. 5. Experimental resonance curves for three different bubbles, and the corresponding fits by Eq. (10).

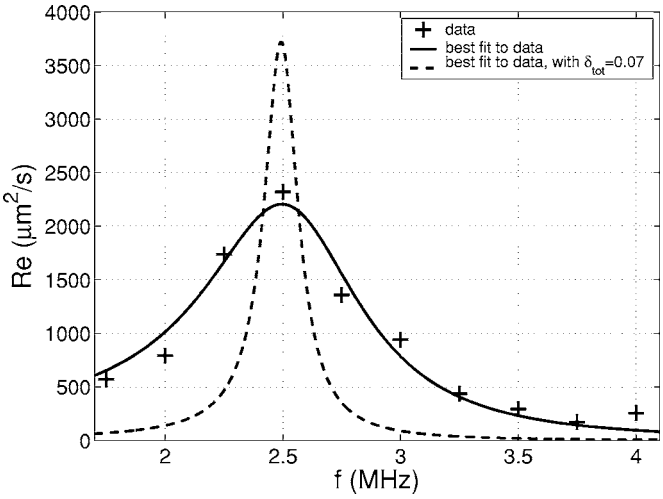


FIG. 7. Example of an experimental resonance curve for a bubble with a resting radius $R_0=1.9\ \mu$m. The solid curve is the fit to Eq. (10), giving $f_0=2.56$ MHz and $\delta_{tot}=0.26$. The dashed line shows the fit to Eq. (10) with an imposed damping coefficient $\delta_{tot}=0.07$ (see text) without shell damping.

IV. RESULTS

A. Eigenfrequency

In Fig. 6, we report the eigenfrequency $f_0=\omega_0/2\pi$, obtained as explained in Sec. III C, for 22 experiments. As expected, the measured eigenfrequencies decrease with the bubble radius. They are also significantly higher than the Minnaert frequency [Eq. (6)], which confirms the influence of the shell. To quantify this influence, we fit the data points to Eq. (5). The best fit yields the following value of the shell elasticity parameter: $\chi=0.54\pm0.10$ N/m, the error bar coming from the dispersion of the experimental data.

The value of χ is fully compatible with the previously reported values for Sonovue™. Using the model of De Jong,[22] which is similar to the model of Marmottant *et al.*[18] used here in the linear regime, Gorce *et al.*[8] gave a value of $\chi=0.55$ N/m based on four analyzed samples. Marmottant *et al.* derived from their model a value of $\chi=1$ N/m on a single example for the shell elasticity.

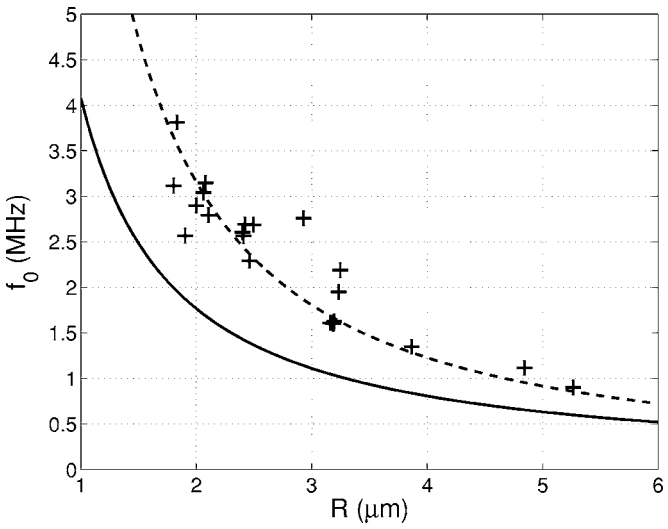


FIG. 6. Experimentally determined eigenfrequency, plotted vs bubble radius. The solid curve shows the Minnaert frequency [Eq. (6)]. The dashed curve shows the resonance frequency including shell elasticity. The best fit with Eq. (5) yields the following value for shell elasticity: $\chi=0.54\pm0.10$ N/m.

B. Damping

We now consider the damping coefficient, which is derived from the experimental data points by fitting to Eq. (10) (see Fig. 5). The damping coefficient is determined as a function of the bubble radius. The variation in the data points is not clearly correlated to the radius, so we rather present a typical example for a bubble of resting radius $R_0=1.9\ \mu$m, see Fig. 7.

Damping arises from various sources: reradiation of sound by the bubbles, thermal diffusion, bulk, and shell viscosity. In the example of Fig. 7, we compute from Eqs. (7) and (8) $\delta_{vis}=0.032$ and $\delta_{rad}=0.022$, given the radius $R_0=1.9\ \mu$m and the eigenfrequency $f_0=2.56$ MHz. Thermal damping is more difficult to express; from Ref. 30 and the value of the thermal diffusivity for C_4F_{10}, we get $\delta_{th}=0.02$. These three sources give a contribution of 0.07 for the damping coefficient. We plot the resonance curve corresponding to this value in Fig. 7: clearly, this curve is too sharp to fit the data correctly, showing the significance of a fourth source of damping, arising from shell viscosity. More precisely, the fit to Eq. (10) with all free parameters gives a significantly higher value of the total damping coefficient, namely $\delta_{tot}=0.26$. This implies a value of $\delta_{shell}=0.19$ for the extra damping through the shell: in this example, shell viscous damping is thus responsible for 73% of the total damping; an average, this proportion is 68%: the shell is therefore the major source of damping. This result confirms existing studies,[31] which showed the significant influence of the shell viscosity on the resonance properties of contrast bubbles. For the total set of 22 data points the value of the shell damping ranged from 0.05 to 0.4. The shell damping shows no clear dependence with the radius, see Fig. 8(A).

From δ_{shell} we can easily calculate the shell viscosity κ_s, see Eq. (9). In the example of Fig. 7, this gives $\kappa_s=2.3\times10^{-8}$ kg/s. Analyzing all data points, we find a significant

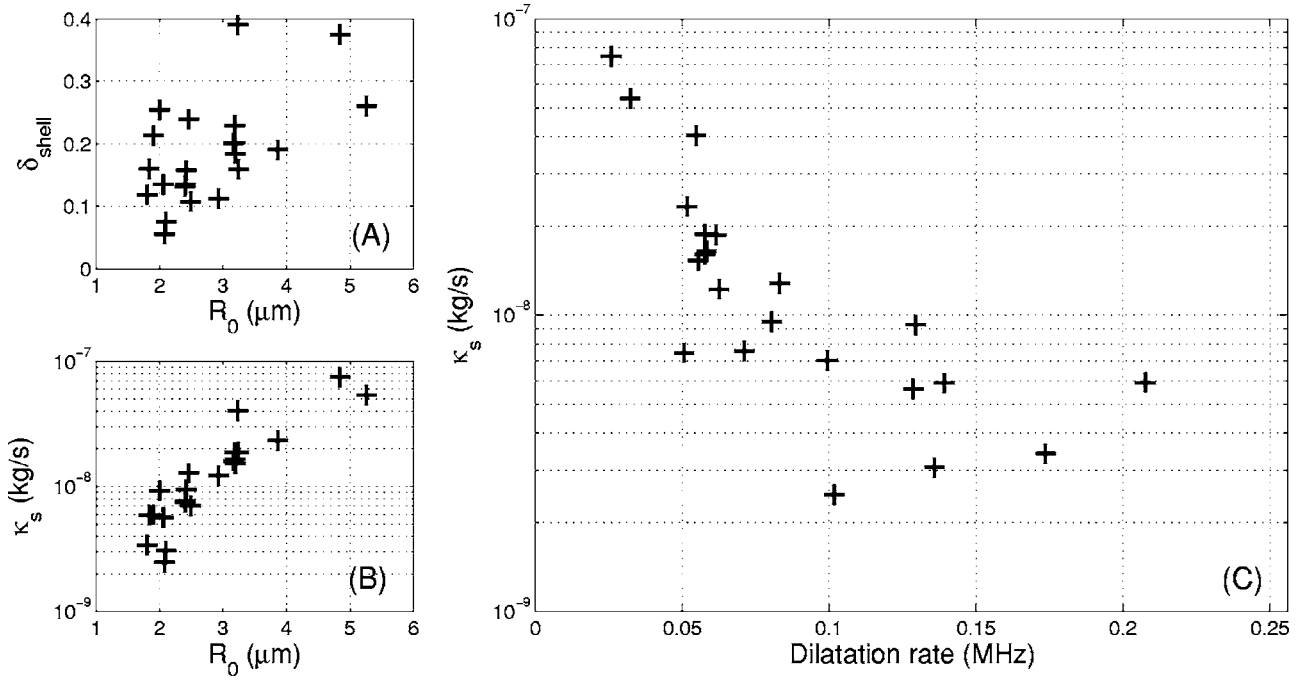


FIG. 8. (A) Experimentally determined shell damping δ_{shell}, plotted vs the ambient bubble radius R_0. (B) Experimentally determined shell viscosity κ_s, plotted vs the ambient bubble radius R_0. (C) Plot of the shell viscosity as a function of the estimate of the dilatation rate $2\pi f_0 \Delta R/R_0$.

increase of the shell viscosity with the radius of the bubble [see Fig. 8(B)], as already reported.[13] However, plotting κ_s as a function of the dilatation rate of the shell offers a more physical picture in terms of surface rheology. We estimate the dilatation rate as: $\dot{R}/R \approx \omega \Delta R/R_0 \approx 2\pi f_0 \Delta R/R_0$, where ΔR is the maximum amplitude of the radial oscillations. We plot κ_s versus the estimated dilatation rate in Fig. 8(C). The plot shows a clear decrease of the shell viscosity with the dilatation rate, which may be the signature of a rheological thinning behavior of the phospholipid monolayer shell. Such a behavior has already been observed for monolayers of myristic acid[32] and of poly(vinylacetate).[33] Furthermore, the order of magnitude of the shell viscosity is 10^{-8} kg/s, which is compatible with previously reported values of $0.72 \cdot 10^{-8}$ kg/s[8]and $1.5 \cdot 10^{-8}$ kg/s.[18]

V. DISCUSSION AND CONCLUSIONS

A. Accuracy of the measurements

We discuss here various sources of bias of the measurements: shrinking of bubbles, compression-only behavior, and uncertainty on the measured radius.

After each burst of ultrasound, some gas may escape from the bubble, reducing the resting radius of the contrast bubble. This unwanted effect is minimized by keeping the ultrasound pressure as low as possible, yet high enough to be able to distinguish the oscillation of the contrast bubble at different frequencies. Great care was taken to verify that the initial resting radius was equal to the final resting radius. In our experiments, we only considered oscillations with a relative decrease in radius less than 10%, hence we assume that the properties of the bubble do not change significantly during the insonation cycle. There is also no difference seen in an ascending or a descending frequency sweep.

Another difficulty is a nonlinear phenomenon referred to as "compression-only" behavior,[34] due to the shell mechanical properties.[18] In the data analysis we included only data for which the ratio between the maximal extension and compression from the equilibrium radius is higher than 0.9: compression only is then negligible.

There is also an uncertainty in the resting radius: since it equals only a few microns, bubbles behave like Mie scatterers[35] hence they create a complicated combination of diffraction and scattering in the focal plane of the camera. This leads to some uncertainty in the image analysis, since the transition in contrast between bubble and background is gradual. The edge detection, described in Sec. III B, at the dark-bright interface at the edge of the bubble will in many cases not give the correct initial bubble radius. Only in the case of in-focus, on-axis, incoherent illumination with a sufficiently high imaging resolution will the lowest cost at the dark-bright interface give the correct bubble radius estimation. In all other cases, the cost needs to be slightly different, depending on the imaging system characteristics like system coherency, objective NA, and abberations (like focus error). In our case we estimate the error we make in the radius estimation at around 10%.

B. Nonlinear pressure and wall effects

The frame rate of 15 million frames/s that was used in the experiment enabled us to resolve the oscillations of the bubbles, to get precise power spectra (Fig. 4), and resonance curves (Fig. 5). We extract a resonance frequency by fitting the observed resonance curves with a linear oscillator response, neglecting the nonlinear influence of the acoustic amplitude, which may be questionable: numerical calculations for uncoated bubbles show that resonance curves be-

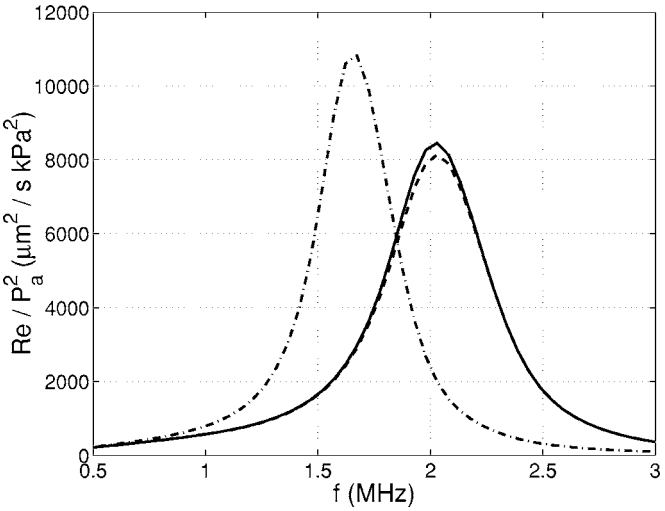


FIG. 9. Resonance curves computed from Eq. (4) for an acoustic amplitude P_a=1 kPa (solid line, same curve as Fig. 2) and 40 kPa (dashed line). Here, the response has been rescaled by P_a^2 to allow for comparison. For P_a =40 kPa we find f_0=2.07 MHz and δ_{tot}=0.24. The third line is a simulation with wall present [dash-dotted line, Eq. (4) computed with an additional term modeling the wall as an image bubble (Ref. 39)]. With wall, we find f_0=1.69 MHz and δ_{tot}=0.21.

come asymmetrical (skewed), and the maximum shifts to a lower frequency. For a driving pressure of 40 kPa the relative decrease of the resonance frequency has been reported to be as large as 10%.[36,37] To address this question for coated bubbles, we perform the same simulation as in Sec. II B with P_a=40 kPa, which is the maximum value used in experiments, and compare the resonance curve with the one of Fig. 2 (see Fig. 9). We find no significant difference, neither in eigenfrequency nor in damping coefficient: the shell elasticity seems to counterbalance the nonlinear effect of pressure.

The presence of the top capillary wall, against which the bubbles rest because of buoyancy, is also expected to affect the resonance frequency.[38,41] As a first approximation, the wall can be modeled as an image bubble, which yields the following prediction of the eigenfrequency of a bubble in contact with a wall: $f_{\text{wall}}=f_0\sqrt{2/3}\simeq 0.82 f_0$.[39] More precisely, we compare the bubble response to an acoustic pressure of amplitude 1 kPa with and without wall in the same figure, Fig. 9. It shows indeed that the eigenfrequency is lowered in the presence of the wall by this factor 0.82. Therefore, the wall tends to lower the resonance frequency; this means that the shell elasticity, 0.54 N/m, that we measured by fitting the experimental data (Fig. 6), is underestimated.

C. Conclusions

We have presented a new, optical method to determine the resonance frequency of individual ultrasound contrast agents bubbles. This method relies on the use of an ultrahigh speed camera, fast enough to resolve bubble oscillations at several Mfps. We operate the camera in a segmented mode, and scan the frequency over the bubble resonance. The bubble response is then recorded at each frequency, and analyzed to construct its resonance curve, from which a resonance frequency as well as a damping coefficient are extracted. The results confirm the influence of the viscoelastic properties of the shell on bubble behavior: The shell elasticity increases the resonance frequency compared to the uncoated bubble case and the shell viscosity proves to be a significant source of damping. Moreover, we showed that the shell viscosity increases with the bubble radius and suggested an explanation in terms of surface rheology.

The measured value of the shell elasticity, 0.54 N/m, and the order of magnitude of the shell viscosity, 10^{-8} kg/s, are in good agreement with previous, independent measurements, giving confidence in this method as an efficient probe of bubble shell properties.

ACKNOWLEDGMENTS

We would like to thank Marlies Overvelde and Jeroen Sijl for fruitful discussions.

This work is part of the research program of the Stichting FOM, which is financially supported by NWO. We acknowledge the Interuniversity Cardiology Institute of the Netherlands (ICIN) and the Senter Novem agency of the Dutch Ministry of Economic Affairs for their financial support. We acknowledge Bracco Research, Geneva, for supplying contrast agents (BR-14 and SonoVue™).

[1] T. Szabo, *Diagnostic Ultrasound Imaging: Inside Out* (Academic, New York, 2004).
[2] N. de Jong, "Acoustic properties of ultrasound contrast agents," Ph.D thesis, Erasmus University Rotterdam (1993).
[3] L. Hoff, *Acoustic Characterization of Contrast Agents for Medical Ultrasound Imaging* (Kluwer, Dordrecht, 2001).
[4] A. L. Klibanov, "Ultrasound contrast agents: Development of the field and current status," Top. Curr. Chem. **222**, 73–106 (2002).
[5] P. J. A. Frinking and N. de Jong, "Acoustic modeling of shell-encapsulated gas bubbles," Ultrasound Med. Biol. **24**(4), 523–533 (1998).
[6] P. J. A. Frinking, N. de Jong, and E. I. Céspedes, "Scattering properties of encapsulated gas bubbles at high ultrasound pressures," J. Acoust. Soc. Am. **105**(3), 1989–1996 (1999).
[7] W. T. Shi and F. Forsberg, "Ultrasonic characterization of the nonlinear properties of contrast microbubbles," Ultrasound Med. Biol. **26**, 93–104 (2000).
[8] J. M. Gorce, M. Arditi, and M. Schneider, "Influence of bubble size distribution on the echogenicity of ultrasound contrast agents. A study of SonoVue™," Invest. Radiol. **35**(11), 661–671 (2000).
[9] S. H. Bloch, R. E. Short, K. W. Ferrara, and E. R. Wisner, "The effect of size on the acoustic response of polymer-shelled contrast agents," Ultrasound Med. Biol. **31**, 439–444 (2005).
[10] J. E. Chomas, P. A. Dayton, D. May, J. Allen, A. L. Klibanov, and K. W. Ferrara, "Optical observation of contrast agent destruction," Appl. Phys. Lett. **77**, 1056–1058 (2000).
[11] N. de Jong, P. J. A. Frinking, A. Bouakaz, M. Goorden, T. Schourmans, J. P. Xu, and F. Mastik, "Optical imaging of contrast agent microbubbles in an ultrasound field with a 100 MHz camera," Ultrasound Med. Biol. **26**, 487–492 (2000).
[12] Y. Sun, D. E. Kruse, P. A. Dayton, and K. W. Ferrara, "High-frequency dynamics of ultrasound contrast agents," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **52**, 1981–1991 (2005).
[13] K. E. Morgan, J. S. Allen, P. A. Dayton, J. E. Chomas, A. L. Klibanov, and K. W. Ferrara, "Experimental and theoretical evaluation of microbubble behavior: Effect of transmitted phase and bubble size," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **47**, 1494–1508 (2000).
[14] P. A. Dayton, J. S. Allen, and K. W. Ferrara, "The magnitude of radiation force on ultrasound contrast agents," J. Acoust. Soc. Am. **112**, 2183–2192 (2002).
[15] C. T. Chin, C. Lancée, J. Borsboom, F. Mastik, M. E. Frijlink, N. de Jong, M. Versluis, and D. Lohse, "Brandaris 128: A digital 25 million frames per second camera with 128 highly sensitive frames," Rev. Sci. Instrum. **74**(12), 5026–5034 (2003).
[16] T. G. Leighton, *The Acoustic Bubble* (Academic, London, 1994).
[17] S. Hilgenfeldt, D. Lohse, and M. Zomack, "Response of bubbles to diag-

nostic ultrasound:A unifying theoretical approach," Eur. Phys. J. B **4**, 247–255 (1998).
[18]P. Marmottant, S. M. van der Meer, M. Emmer, M. Versluis, N. de Jong, S. Hilgenfeldt, and D. Lohse, "A model for large amplitude oscillations of coated bubbles accounting for buckling and rupture," J. Acoust. Soc. Am. **118**(6), 3499–3505 (2005).
[19]M. S. Plesset and A. Prosperetti, "Bubble dynamics and cavitation," Annu. Rev. Fluid Mech. **9**, 145–185 (1977).
[20]C. E. Brennen, *Cavitation and Bubble Dynamics* (Oxford University Press, Oxford, 1995).
[21]M. Brenner, S. Hilgenfeldt, and D. Lohse, "Single bubble sonoluminescence," Rev. Mod. Phys. **74**(2), 425–484 (2002).
[22]N. de Jong, R. Cornet, and C. T. Lancée, "Higher harmonics of vibrating gas-filled microspheres Part one: Simulations," Ultrasonics **32**(6), 447–453 (1994).
[23]C. C. Church, "The effects of an elastic solid surface layer on the radial pulsarions of gas bubble," J. Acoust. Soc. Am. **97**, 1510–1521 (1995).
[24]L. Hoff, P. C. Sontum, and J. M. Havem, "Oscillations of polymeric microbubbles: Effect of the encapsulating shell," J. Acoust. Soc. Am. **107**, 2272–2280 (2000).
[25]D. B. Khismatullin and A. Nadim, "Radial oscillations of encapsulated microbubbles in viscoelastic liquids," Phys. Fluids **14**, 3534 (2002).
[26]K. Sarkar, W. T. Shi, D. Chatterjee, and F. Forsberg, "Characterization of ultrasound contrast microbubbles using in vitro experiments and viscous and viscoelastic interface models for encapsulation," J. Acoust. Soc. Am. **118**(1), 539–550 (2005).
[27]We thank Charles C. Church for pointing out a typo in Ref. 18. The dilatational viscosity κ_s is expressed in units of kg/s.
[28]M. Minnaert, "On musical air-bubbles and the sounds of running water," Philos. Mag. **16**, 235–248 (1933).
[29]M. Sonka, V. Hlavac, and R. Boyle, *Image Processing, Analysis, and Machine Vision*, 2nd ed. (PWS, Pacific Grove, CA, 1999).
[30]A. Prosperetti, "Thermal effects and damping mechanisms in the forced radial oscillations of gas bubbles in liquids," J. Acoust. Soc. Am. **61**, 17–27 (1977).
[31]D. B. Khismatullin, "Resonance frequency of microbubbles: Effect of viscosity," J. Acoust. Soc. Am. **116**, 1463–1473 (2004).
[32]C. Lemaire and D. Langevin, "Longitudinal surface waves at liquid interfaces. Measurement of monolayer viscoelasticity," Colloids Surf. **65**, 101–112 (1992).
[33]F. Monroy, F. Ortega, and R. G. Rubio, "Dilatational rheology of insoluble polymer monolayers: Poly(vinylacetate)," Phys. Rev. E **58**(6), 7629–7641 (1998).
[34]N. de Jong, C. T. Chin, A. Bouakaz, F. Mastik, D. Lohse, and M. Versluis, " 'Compression-only' behavior of phospholipid-coated contrast bubbles," Ultrasound Med. Biol. (in press).
[35]C. F. Bohren and D. R. Huffman, *Absorption and Scattering of Light by Small Particles* (Wiley, New York l983).
[36]W. Lauterborn, "Nonlinear oscillations of gas bubbles," J. Acoust. Soc. Am. **59**(2), 283–293 (1976).
[37]A. Prosperetti, "Bubble phenomena in sound fields: Part two," Ultrasonics **22**, 115–124 (1984).
[38]E. M. B. Payne, S. Illesinghe, A. Ooi, and R. Manasseh, "Symmetric mode resonance of bubbles attached to a rigid boundary," J. Acoust. Soc. Am. **118**, 2841–2849 (2005).
[39]M. Strasberg, "The pulsation frequency of nonspherical gas bubbles in liquids," J. Acoust. Soc. Am. **25**, 536–537 (1953).
[40]J. S. Allen, D. E. Kruse, P. A. Dayton, and K. W. Ferrara, "Effect of coupled oscillations on microbubble behavior," J. Acoust. Soc. Am. **114**(3), 1678–1690 (2003).
[41]M. Emmer, A. van Wamel, D. E. Goertz, and N. de Jong, "The onset of microbubble vibration," Ultrasound Med. Biol. (in press).