CODEN: JASMAN

The Journal of the Acoustical Society of America

ISSN: 0001-4966

Vol. 125, No. 3 March 2009

(Continued)

CONTENTS—*Continued from preceding page*

UNDERWATER SOUND [30]

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

TRANSDUCTION [38]

STRUCTURAL ACOUSTICS AND VIBRATION [40]

NOISE: ITS EFFECTS AND CONTROL [50]

ARCHITECTURAL ACOUSTICS [55]

(Continued)

CONTENTS—*Continued from preceding page*

ACOUSTIC SIGNAL PROCESSING [60]

PHYSIOLOGICAL ACOUSTICS [64]

PSYCHOLOGICAL ACOUSTICS [66]

SPEECH PRODUCTION [70]

SPEECH PERCEPTION [71]

(Continued)

CONTENTS—*Continued from preceding page*

Steep and shallow phase gradient distortion product otoacoustic emissions arising basal to the primary tones

Glen K. Martin
Research Service (151), VA Loma Linda Healthcare System, 11201 Benton Street, Loma Linda, California 92357 and Department of Otolaryngology—Head & Neck Surgery, Loma Linda University Medical Center, Loma Linda, California 92350
glen.martin2@va.gov

Barden B. Stagner
Research Service (151), VA Loma Linda Healthcare System, 11201 Benton Street, Loma Linda, California 92357
barden.stagner@va.gov

Paul F. Fahey
Department of Physics/Electrical Engineering, University of Scranton, Scranton, Pennsylvania 18510
faheyp1@scranton.edu

Brenda L. Lonsbury-Martin
Research Service (151), VA Loma Linda Healthcare System, 11201 Benton Street, Loma Linda, California 92357 and Department of Otolaryngology—Head & Neck Surgery, Loma Linda University Medical Center, Loma Linda, California 92350
blonsbury-martin@llu.edu

Abstract: Distortion product otoacoustic emission (DPOAE) level/phase maps were collected in humans with and without an interference tone (IT) near the DPOAE frequency place (f_{dp}) at primary-tone levels of 75 dB SPL. A DPOAE component with the expected steep phase gradient could be extracted at f_{dp}, however, considerable vertical-phase banding, presumably indicative of reflection emissions, remained. An IT placed 0.33 oct above f_2 removed most of this banding, revealing DPOAE components originating basal to the IT frequency place. These findings suggest that the commonly accepted two-source model of DPOAE generation may need to be qualified when higher primary-tone levels are utilized.

PACS numbers: 43.64.Bt, 43.64.Jb, 43.64.Kc [WS]
Date Received: September 15, 2008 **Date Accepted:** December 22, 2008

1. Introduction

Kim (1980) first proposed that distortion-product otoacoustic emissions (DPOAEs) are produced by a combination of two emission-generator components, one originating at the f_2 place and the other arising from the DPOAE frequency place (f_{dp}). The now commonly accepted two-source model (e.g., Zwieg and Shera, 1995; Talmadge *et al.*, 1998) for apical DPOAE generation (i.e., $f_{dp}<f_1$; e.g., $2f_1$-f_2 DPOAE), supported by numerous studies in humans (e.g., Talmadge *et al.*, 1999; Knight and Kemp, 2000, 2001; Kalluri and Shera, 2001; Konrad-Martin *et al.*, 2001; Dhar *et al.*, 2005), proposes that within the cochlea, there are two separate contributors to the DPOAE levels and phases measured in the ear canal when elicited by low- to moderate-level primaries. The first contributor, referred to as the "generator" or "distortion" source, is thought to arise from the nonlinear interaction between the f_1 and f_2 primary tones. The primary-tone "overlap" region for apical DPOAEs is assumed to be near the tonotopic location on the basilar membrane (BM) of the higher-frequency f_2 primary tone, which creates energy at the DPOAE frequency that then travels both apically and basally within the cochlea.

The apically traveling energy reaches the f_{dp} place on the BM and is then reflected back basally, thereby providing the second source (reflection) of the otoacoustic emission (OAE) measured in the ear canal. This energy reflection is attributed to the presence of randomly distributed inhomogeneities on the BM in the f_{dp} region. The interaction of these two sources results in the commonly observed fine structure in human DP-grams.

Emissions that Kemp and Knight (2000, 2001) called "place-fixed" and "wave-fixed" OAEs correspond to the more mechanistic terms "coherent linear reflection" and "nonlinear distortion," respectively, adopted by Shera and Guinan (1999) to refer to these same emission properties. The Shera and Guinan (1999) formulation emphasizes the unique phase characteristics of the DPOAE, depending upon the site of emission generation. For example, for the wave-fixed component, a constant f_2/f_1-ratio sweep maintains the relative phases of the f_1 and f_2 primary tones and the DPOAE. Thus, because of the property of cochlear-scale invariance, DPOAE phase remains relatively unvarying for a constant f_2/f_1-ratio sweep resulting in shallow phase gradients. Because the DPOAE-place component (reflection) comes from the f_{dp}, for a constant f_2/f_1-ratio sweep, the phase for this constituent changes rapidly and is associated with steep phase gradients.

In the DPOAE level/phase (L/P) maps described by Knight and Kemp (2000, 2001), when DPOAE phase is plotted as a function of DPOAE frequency, a constant vertical-phase band for a given DPOAE frequency presumably indicates a place-fixed (reflection) source. In contrast, a constant horizontal-phase band for a fixed f_2/f_1-ratio sweep as a function of DPOAE frequency is considered indicative of a wave-fixed (distortion) source.

It is important to emphasize that the majority of the research designed to explore the two-source DPOAE model in humans has been conducted in exceptional subjects having very robust DPOAEs (e.g., Dhar *et al.*, 2005) using low to moderate primary-tone levels (e.g., Konrad-Martin *et al.*, 2001; Kalluri and Shera, 2001). Recently, our laboratory assembled DPOAE L/P maps like those described by Knight and Kemp (2000, 2001), at similar higher primary-tone levels with $L_1, L_2 = 75, 75$ dB SPL. These higher primary-tone levels are required to obtain robust $2f_1$-f_2 and $2f_2$-f_1 DPOAEs across a wide range of f_2/f_1 ratios, especially in subjects with various types of sensorineural hearing loss (e.g., Stagner *et al.*, 2007). In these L/P maps, both horizontal- and vertical-phase banding were obtained that were very similar to that originally described by Knight and Kemp (2000, 2001).

In an attempt to "unmix" the two DPOAE components presumably indicative of the two emission mechanisms, interference tones (ITs) were placed near the f_{dp}, and DPOAE L/P maps were obtained with and without the IT. Vector differences computed between the two conditions extracted a reflection component at f_{dp} for both the $2f_1$-f_2 and $2f_2$-f_1 DPOAEs; however, significant vertical banding still remained. In subsequent experiments, ITs placed 0.33 oct above f_2 removed most of the observable vertical-phase banding for both DPOAEs. The present report describes these theoretically important findings consistent with the presence of other DPOAE sources at higher primary-tone levels in humans.

2. Methods

2.1 Subjects

DPOAEs were measured in three ears of three normal-hearing human subjects between 18 and 30 years of age. Each subject had normal DPOAEs as compared to our laboratory's database. All subjects provided informed consent and received monetary compensation for participation in the study. DPOAEs were obtained with subjects seated comfortably in a reclining chair within a single-walled sound booth situated in a quiet laboratory setting. The human-research protocol was approved by the institutional review board (IRB) of the VA Loma Linda Healthcare System.

2.2 DPOAE measures

To assure normal baseline DPOAEs, "DP-grams" as a function of f_2 frequency were measured with f_2 ranging from 0.275 to 15.3 kHz in 0.1-oct steps with $L_1, L_2 = 55, 55$; 65,65; 75,75; and

65,55 dB SPL. Primary tones were produced by two digital-to-analog (D/A) channels of a digital-signal processing (DSP) board (Digidesign, Audiomedia), mounted in a microcomputer (Apple, Macintosh Quadra 700). The f_1 and f_2 signals were presented using two ear-speakers (Etymotic Research, ER-2), and the level of the ear-canal sound pressure was measured using a low-noise microphone assembly (Etymotic Research, ER-10B+). The ear-canal signal was synchronously sampled at 44,100 kHz and averaged ($n=4$) by an analog-to-digital (A/D) channel of the DSP board. A 4096-point fast Fourier transform (FFT) of the averaged time sample was performed by customized software. The $2f_1$-f_2 and $2f_2$-f_1 DPOAEs and associated noise-floor (NF) levels were extracted from the FFT. The NF was based upon the average of eight frequency bins on either side of the DPOAE frequency bin, excluding the first bin on either side of the DPOAE frequency.

2.3 DPOAE level/phase (L/P) maps

To generate DPOAE L/P maps, DPOAEs were measured in response to constant f_2/f_1 ratio sweeps varied in 0.025 increments from $f_2/f_1=1.025$ to 1.5, with DPOAE frequency steps of ~43 Hz, from 0.5 to 6 kHz for both $2f_1$-f_2 and $2f_2$-f_1, resulting in f_1 ranging from 0.258 to 12.016 kHz and f_2 from 0.366 to 18.023 kHz. For this study, the L/P maps were collected at primary-tone levels of $L_1,L_2=75,75$ dB SPL, using a 2048-point FFT and four- or eight-time averages. DPOAE level was directly plotted (Microsoft Excel 2003, v.11.5), while phase was corrected for primary-tone phase variation and unwrapped by "looking" in two directions (f_2/f_1 ratio and DPOAE frequency) using custom-developed Excel-based routines before plotting. Final, more detailed analysis and plotting were performed in MATLAB.

DPOAE L/P maps were obtained with and without an IT placed at either 44 Hz below f_{dp} (IT=65 dB SPL) or at 0.33 oct above f_2 (IT=75 dB SPL). ITs were presented on alternate trials throughout the protocol to minimize any effects due to time-dependent changes in DPOAEs. The IT, which was digitally mixed with f_1, was rotated in phase by 90° over the four presentations and then time averaged to eliminate the majority of emission components produced by the IT. Vector differences were computed between control (no IT) and IT conditions to produce residual DPOAE L/P maps consisting of DPOAE components removed by the IT. DPgrams with and without the IT and for the residual were extracted from the DPOAE L/P maps for wide ($f_2/f_1=1.20$) and narrow ($f_2/f_1=1.025$) primary-tone ratios. The phase of the residual was unwrapped in the frequency direction and plotted as a function of DPOAE frequency and phase in cycles. Plotted this way, the slope is in units of time and can be thought of as a delay. The contrast of these delays suggests different physical mechanisms of emission production.

3. Results

The results of obtaining DPOAE L/P maps with and without an IT near f_{dp} are shown for a representative subject in Fig. 1. Similar results were found for the other subjects, but the subject illustrated had the most robust DPOAEs, making it more straightforward to appreciate the findings in the DPOAE L/P maps. On the left of Fig. 1 are six DPOAE L/P maps [Figs. 1(a), 1(c), and 1(e) =level; Figs. 1(b), 1(d), and 1(f) =phase] corresponding to the control [Figs. 1(a) and 1(b)], IT [Figs. 1(c) and 1(d)], and residual [Figs. 1(e) and 1(f)] experimental conditions. In the control condition of Fig. 1(a), DPOAE fine structure is evident in the level plot at f_2/f_1 ratios of ~1.2 (upper dashed black line) as peaks and valleys (arrows). In the corresponding phase plot of Fig. 1(b), the two frequently observed phase behaviors for the $2f_1$-f_2 DPOAE are clearly evident. That is, horizontal-phase banding dominates at wide f_2/f_1 ratios, which presumably represents distortion emissions, while vertical-phase banding associated with reflection emissions is apparent at f_2/f_1 ratios less than about 1.1. For the $2f_2$-f_1 DPOAE [bottom half of Fig. 1(b)], vertical-phase banding was obtained for all f_2/f_1 ratio values. In the IT condition shown in the level plot of Fig. 1(c), DPOAE fine structure is substantially reduced (arrows). However, in the corresponding phase plot of Fig. 1(d) for the $2f_1$-f_2 DPOAE, significant vertical banding remained, especially above 3 kHz. Likewise, much of the vertical-phase structure associated with the $2f_2$-f_1 DPOAE remained in this frequency region. In the residual map of Fig. 1(e),

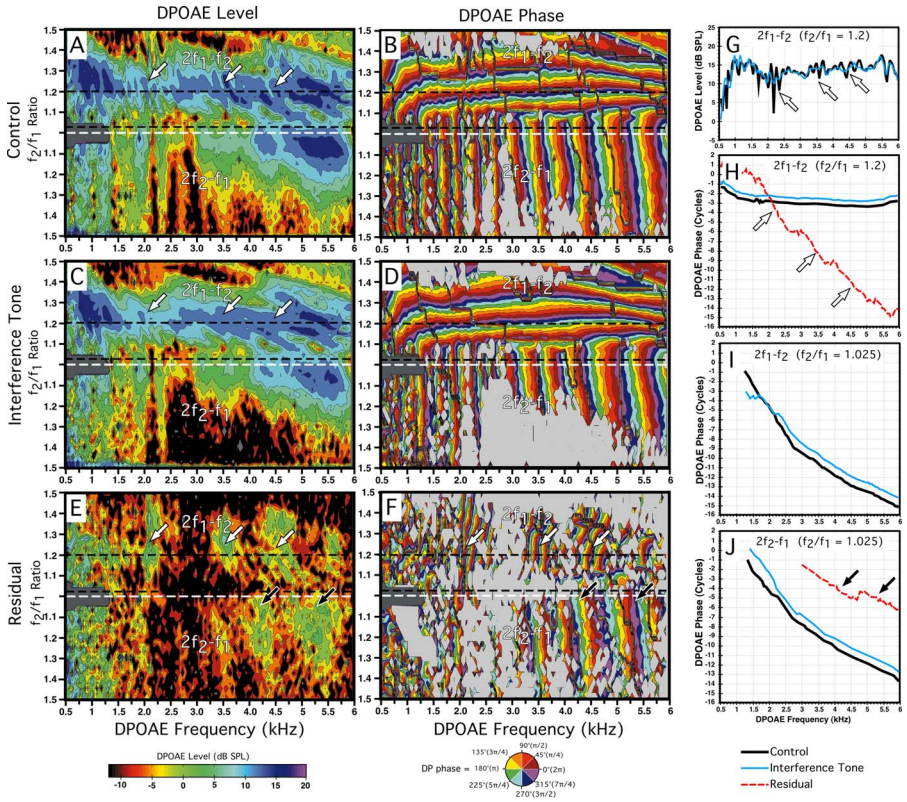


Fig. 1. (Color online) DPOAE L/P maps obtained with and without an IT placed 44 Hz below f_{dp}. The residual map in (e) demonstrates that DPOAE components can be extracted near f_{dp} that have vertical-phase banding properties in (f) (arrows). Red/gray dashed line in (h) demonstrates that these emissions have the expected steep phase gradients consistent with a DPOAE-reflection component. However, in the presence of the IT, large DPOAE components remain (c) associated with similar vertical banding (d) and phase gradients as indicated by the blue/gray line in (i). Black dashed lines on the DPOAE L/P maps indicate f_2/f_1 ratios where data were extracted for the plots in (g) and (h) (see text for complete details of this and Fig. 2).

based upon the vector difference between the control (no IT) and IT maps, patchy residuals are evident where the IT removed DPOAE components for the $2f_1$-f_2 (white arrows) and for the $2f_2$-f_1 (black arrows). Finally, in the corresponding phase map of Fig. 1(f), the phase behavior of these residuals showed vertical banding as commonly described for a reflection component from f_{dp}. It should be emphasized that the phase banding was notably much narrower (i.e., steeper phase gradient) for the residual $2f_1$-f_2 DPOAEs at wide ratios [Fig. 1(f), white arrows] as compared to the narrow ratio $2f_1$-f_2 or $2f_2$-f_1 for all conditions [Figs. 1(b), 1(d), and 1(f)].

The four plots to the right of the DPOAE L/P maps illustrate more familiar analyses. For example, Fig. 1(g) demonstrates DP-grams extracted from these maps at an f_2/f_1 ratio of 1.2 for the $2f_1$-f_2 DPOAE without (solid black line) and with (solid blue/gray line) the IT near f_{dp}. It is clear that in this situation the IT essentially eliminated the fine structure supporting the visual impression gained by comparing the central regions (white arrows) of the control [Fig. 1(a)] and IT [Fig. 1(c)] DPOAE L/P maps.

The remaining three plots show phase curves extracted for the $2f_1$-f_2 DPOAE at wide [Fig. 1(h)] and narrow [Fig. 1(i)] f_2/f_1 ratios, as well as for the $2f_2$-f_1 DPOAE at a narrow ratio [Fig. 1(j)] for control, IT, and residual conditions. For the $2f_1$-f_2 at a standard ratio (f_2/f_1=1.2), shallow phase slopes [Fig. 1(h)] were obtained for both the control (solid black line) and IT (solid blue/gray line) conditions, consistent with the horizontal-phase bands seen in the corre-

sponding phase maps at this ratio [Figs. 1(b) and 1(d)]. For this same f_2/f_1 ratio, the residual showed a steep phase slope (dashed red/gray line) associated with the vertical-phase banding [Fig. 1(f), white arrows] extracted by the IT near f_{dp}. These findings illustrate the ability to extract a reflection component from the f_{dp}. However, most notably, significant vertical-phase banding remained [Fig. 1(d)] that could not be removed by the IT.

At narrow f_2/f_1 ratios of 1.025 [Fig. 1(i)], steep phase-gradient DPOAEs were obtained for both the control (solid black line) and IT (solid blue/gray line) conditions accompanied by vertical-phase structure. However, at this narrow ratio setting, the residual was too near the NF to measure its phase characteristics. This more conventional analysis confirms the observations noted above on the L/P maps that the slopes of the phase gradients for the wide ratio $2f_1$-f_2 DPOAE residual [white arrows in Fig. 1(h)] are roughly twice those of the narrow ratio $2f_1$-f_2 [Fig. 1(i)], or $2f_2$-f_1 [Fig. 1(j)] components for the control, IT, or residual (black arrows) conditions suggesting that the $2f_1$-f_2 DPOAE from f_{dp} has a much longer latency than these narrow-ratio $2f_1$-f_2 and $2f_2$-f_1 components associated with steep phase gradients.

Phase curves for the $2f_2$-f_1 DPOAE are shown in Fig. 1(j) for the narrow f_2/f_1-ratio condition (1.025) equidistant below the white dashed centerline at $f_2/f_1=0$. For this DPOAE, all conditions revealed steep phase gradients associated with the vertical-phase banding observed in the phase maps, which is consistent with the place-fixed behavior of this emission.

Figure 2 displays the results of repeating the same experiment depicted in Fig. 1 in the same subject in a different session, but with the IT placed 0.33 oct above f_2. First, the striking similarity of the DPOAE L/P maps between Figs. 1(a) and 2(a) with respect to DPOAE level should be noted thus demonstrating the excellent test/retest reliability of these maps within the same individual. In the control condition of Fig. 2(a), fine structure is again evident in the DPOAE levels at wide f_2/f_1 ratios of $\sim$1.2 as peaks and valleys. In the corresponding phase plot of Fig. 2(b), the two phase behaviors can again be observed for the $2f_1$-f_2 DPOAE. For the $2f_2$-f_1 DPOAE (bottom half of plot), the phase banding is vertical, which supports the presence of a reflection-based generation mechanism. In the IT condition illustrated in Figs. 2(c) and 2(d), as compared to Fig. 1(c), it can be seen that the fine structure at $f_2/f_1=1.2$ in the level plot [Fig. 2(c)] is not affected by the IT, when it is placed 0.33 oct above f_2. This outcome suggests that the IT had minimal effects on the f_2 source at this distance basal to f_2, otherwise the DPOAE fine structure would be modified by a reduction in this DPOAE component. In contrast to the previous experiment, in the related phase plot of Fig. 2(d), a significant amount of the vertical-phase banding that was present in the control condition [Fig. 2(b)] is removed by the IT for both DPOAEs. In the residual maps of Figs. 2(e) and 2(f), a large residual DPOAE for both the $2f_1$-f_2 and $2f_2$-f_1 emissions is evident in Fig. 2(e), which is indicative of the DPOAE components removed by the IT that presumably originate, or are modified, from BM regions situated basal to the IT place. In the corresponding phase map of Fig. 2(f), the phase behavior of these residuals shows both horizontal- and vertical-phase banding for the $2f_1$-f_2 DPOAE, and vertical-phase banding for the $2f_2$-f_1 DPOAE.

The top plot of Fig. 2(g) at the right of the DPOAE L/P maps shows DP-grams extracted for an f_2/f_1 ratio of 1.2 for the $2f_1$-f_2 emission collected without (solid black line) and with an IT (solid blue/gray line) placed 0.33 oct above f_2. These DP-grams traverse the center of the fine structure regions in Figs. 1(a) and 1(c) (dashed black line at 1.2). It is clear that both DP-grams superimposed on one another convincingly indicate that in this situation the 75 dB SPL IT did not influence DPOAE sources originating from the f_2 place that interact with the source reflected from the f_{dp} to produce this fine structure. This observation reinforces the notion that the modifications produced by the IT resulted from its influence on DPOAE components located basal to the IT place.

The other three plots depict DPOAE phase gradients extracted from the DPOAE L/P maps for the $2f_1$-f_2 and $2f_2$-f_1 DPOAEs for wide (1.2) and narrow (1.025) f_2/f_1 ratios for the control, IT, and residual experimental conditions. Figure 2(h) shows phase curves for the wide 1.2-ratio condition, a circumstance where all the phase curves have shallow slopes that are in agreement with the horizontal-phase banding observed across all three experimental conditions. At this ratio, the IT presumably removed a distortion component, which is in accord with

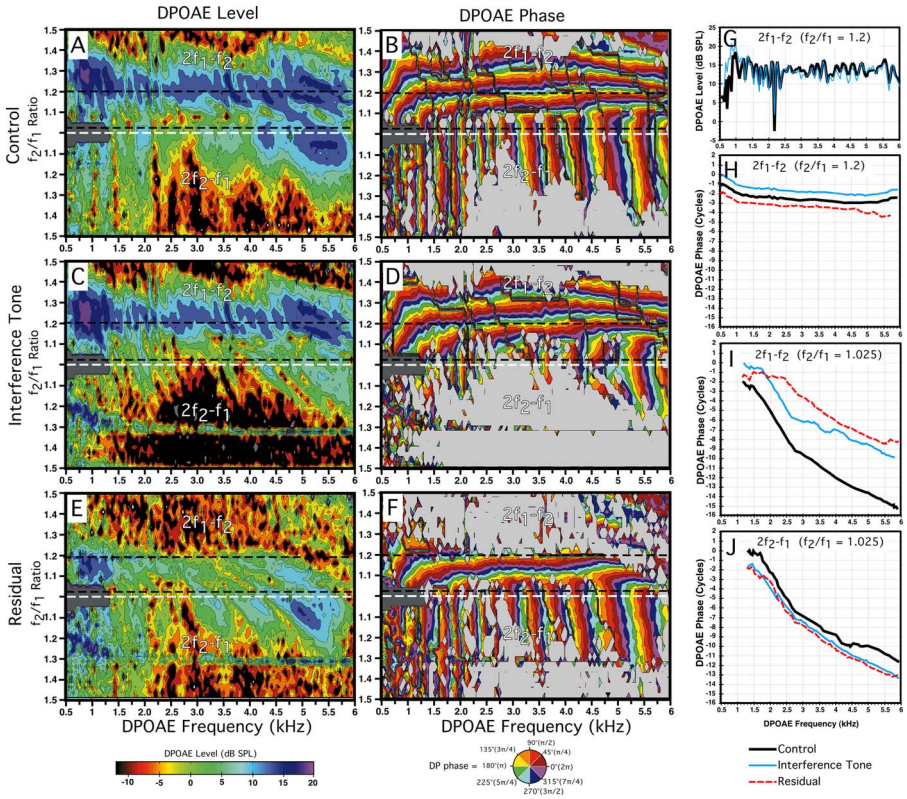


Fig. 2. (Color online) DPOAE maps obtained with and without an IT placed 0.33 oct above f_2. The residual map in (e) demonstrates that by placing the IT above f_2, large DPOAE components can be extracted that have phase properties of both distortion and reflection emissions, depending upon the f_2/f_1 ratio (f). Shallow and steep phase gradients associated with these two components were confirmed by extracting the phase curves at a wide f_2/f_1 ratio in (h), and at a narrow ratio in (i) (dashed red/gray lines). In (d), almost all of the narrow f_2/f_1-ratio vertical-phase banding was removed by the IT for the narrow ratio $2f_1$-f_2 DPOAE, and similarly for the $2f_2$-f_1 DPOAE, suggesting that both of these reflectionlike components originate in regions of the cochlea that are basal to the IT place.

the shallow phase slope for the residual (red/gray dashed line) constituent. However, this component apparently was not generated at f_2, but rather basal to the IT frequency place above f_2.

For the narrow ratio situation at 1.025 shown in Fig. 2(i), steep phase gradients were obtained for all conditions. The phase gradient for the residual (red/gray dashed line) seen in Fig. 2(f) as vertical banding is consistent with the IT removing a DPOAE source with reflection-like phase properties that does not arise from the f_{dp}.

In Fig. 2(j), the phase slopes are steep for all three experimental conditions as commonly observed for the place-fixed behavior of the $2f_2$-f_1 DPOAE. However, in this situation, the IT was considerably above the f_{dp} until the f_2/f_1 ratio values were equal to or >1.3 in contrast to Fig. 1(j) where it is just below the f_{dp} for all ratios. Thus, these place-fixed components also appear to originate from regions that are substantially basal to the f_{dp}.

4. Discussion

A number of studies in human subjects (e.g., Talmadge *et al.*, 1999; Knight and Kemp, 2000, 2001; Kalluri and Shera, 2001; Konrad-Martin *et al.*, 2001; Dhar *et al.*, 2005) have supported the two-source model of DPOAE generation synthesized by Shera and Guinan (1999) for low- to moderate-level primaries. The present results from human subjects also confirm that a steep phase gradient DPOAE component can be extracted with ITs near f_{dp}. Presumably, this emission corresponds to the reflection component arising from randomly distributed inhomogene-

ities on the BM in the f_{dp} region. However, significant vertical-phase banding remained in the DPOAE L/P maps that was shown to be associated with steep phase gradients. By placing the IT 0.33 oct above f_2, most, if not all, of this remaining vertical banding, along with a significant component associated with wave-fixed phase behavior, was extracted [Fig. 2(f)]. It is noteworthy that the high-frequency IT did not alter the pattern of DPOAE fine structure at wide f_2/f_1 ratios, which is due to the interaction of the f_2 and f_{dp} emission components. This finding suggests that the IT acted by removing or modifying components basal to the f_2 that were responsible for the remaining place-fixed phase behavior. These results are in agreement with evidence that other emissions, such as the $2f_2$-f_1 DPOAE, can be generated basal to the f_2 place (e.g., Martin *et al.*, 1998). The fact that placing the IT 0.33 above f_2 strongly influenced the $2f_2$-f_1 DPOAE supports the earlier findings for this emission. These remaining steep phase gradient emissions also cannot be attributed to suppression of multiple internal reflections, since Dhar *et al.* (2002) demonstrated that the presence of such reflections interact to significantly alter DPOAE fine structure, which, in this case, remained unchanged when these emissions were removed by the IT. It is possible that these basal source contributions could somehow be reflected from f_{dp}. In this latter situation, the high-frequency IT at 0.33 oct above f_2 eliminates the basal source and consequently the reflected component from f_{dp} associated with vertical-phase behavior. However, the observations that this component cannot be affected by the IT at f_{dp} [Figs. 1(c) and 1(d)] and that the fine structure remains unchanged [Fig. 2(c)] in the presence of the high-frequency IT seem to support the notion that this particular DPOAE component associated with vertical-phase banding comes from a region basal to f_2.

Overall, the present findings suggest that the two-source model of DPOAE generation in humans may be limited to situations where the primary tones are kept at low to moderate levels. As primary-tone levels are increased to higher levels, as in the present case, it appears that other emission components are generated basal to f_2 for the case of the $2f_1$-f_2 DPOAE, and even well basal to the emission place for the upper sideband $2f_2$-f_1 DPOAE. These basal components have phase properties that have previously been attributed strictly to distortion or reflection emissions generated at f_2 or f_{dp}, respectively. Knight and Kemp (2000) obtained DPOAE L/P maps in two human subjects in response to 70-dB SPL primary-tone levels that were very similar to the ones illustrated here. They subsequently attempted to extract the place-fixed and wave-fixed components (Knight and Kemp, 2001) using temporal windowing and inverse fast Fourier transform (IFFT) techniques. The resulting outcomes led these authors to conclude that both emission types were widely distributed, a finding that is contradictory to the current results. Recently, Dhar *et al.* (2005) also discovered over a very restricted frequency range of 400 Hz that the reflection component dominated at narrow ratios and low primary-tone levels. Interestingly, this outcome was not observed over the large frequency span encompassed by the current DPOAE L/P maps. It is important to emphasize that the present results were obtained using IT techniques that directly remove emission components by the suppressive effects of a third tone in a nonlinear system.

The other more frequently employed strategy of using time windowing and IFFT filtering assumes that only two components are to be separated, and that this approach can be utilized, because their corresponding phase behaviors result in unique and separable latencies in the pseudotime domain. It appears likely that under the conditions used in the present study, these methods would also extract two components, but the assumption that one of the components arose entirely from f_{dp} and the other from f_2 would be incorrect. Of course, this circumstance represents an important issue that needs to be tested by comparing the distribution of components that can be extracted with the two techniques at these higher primary-tone levels. Overall, if other investigators can confirm these findings, then the commonly accepted two-source model of DPOAE generation in humans may require further qualifications based upon primary-tone levels.

Acknowledgments

This work was supported in part by the NIH (Grant No. DC000613) and the VA Rehabilitation, Research and Development Service (Grant Nos. C449R and C6212L). The authors thank Alisa

Nelson-Miller for technical assistance and three anonymous reviewers for their insightful comments and suggestions.

References and links

Dhar, S., Long, G. R., Talmadge, C. L., and Tubis, A. (**2005**), "The effect of stimulus-frequency ratio on distortion product otoacoustic emission components," J. Acoust. Soc. Am. **117**, 3766–3776.

Dhar, S., Talmadge, C. L., Long, G. R., and Tubis, A. (**2002**), "Multiple internal reflections in the cochlea and their effect on DPOAE fine structure," J. Acoust. Soc. Am. **112**, 2882–2897.

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.

Kim, D. O. (**1980**), "Cochlear mechanics: Implications of electrophysiological and acoustical observations," Hear. Res. **2**, 297–317.

Knight, R. D., and Kemp, D. T. (**2000**), "Indications of different distortion product otoacoustic emission mechanisms from a detailed f1,f2 area study," J. Acoust. Soc. Am. **107**, 457–473.

Knight, R. D., and Kemp, D. T. (**2001**), "Wave and place fixed DPOAE maps of the human ear," J. Acoust. Soc. Am. **109**, 1513–1525.

Konrad-Martin, D., Neely, S. T., Keefe, D. H., Dorn, P. A., and Gorga, M. P. (**2001**), "Sources of distortion product otoacoustic emissions revealed by suppression experiments and inverse fast Fourier transforms in normal ears," J. Acoust. Soc. Am. **109**, 2862–2879.

Martin, G. K., Jassir, D., Stagner, B. B., Whitehead, M. L., and Lonsbury-Martin, B. L. (**1998**), "Locus of generation for the $2f_1$-f_2 vs $2f_2$-f_1 distortion-product otoacoustic emissions in normal-hearing humans revealed by suppression tuning, onset latencies, and amplitude correlations," J. Acoust. Soc. Am. **103**, 1957–1971.

Shera, C. A., and Guinan, J. J., Jr. (**1999**), "Evoked otoacoustic emissions arise by two fundamentally different mechanisms: A taxonomy for mammalian OAEs," J. Acoust. Soc. Am. **105**, 782–798.

Stagner, B. B., Meinke, D. K., Lonsbury-Martin, B. L., and Martin, G. K. (**2007**), "Group delay contour plots derived from DPOAE level/phase maps in normal hearing and noise-damaged humans," Assoc. Res. Otolaryngol. Abstr. **18**, 13.

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.

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.

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.

Adaptation to frozen babble in spoken word recognition

Robert Albert Felty
Department of Psychological and Brain Sciences, Indiana University, 1101 E. 10th Street, Bloomington, Indiana 47405
robfelty@indiana.edu

Adam Buchwald
Department of Psychological and Brain Sciences, Indiana University, 1101 E. 10th Street, Bloomington, Indiana 47405 and Department of Speech-Language Pathology and Audiology, 665 Broadway, Suite 910, Steinhardt School of Culture, Education and Human Development, New York University, New York, New York 10012
buchwald@nyu.edu

David B. Pisoni
Department of Psychological and Brain Sciences, Indiana University, 1101 E. 10th Street, Bloomington, Indiana 47405
pisoni@indiana.edu

Abstract: Previous research has shown that listeners can adapt to particular samples of noise, a phenomenon known as "frozen noise" [Langhans and Kohlrausch, J. Acoust. Soc. Am. **91**, 3456–3470 (1992)]. However, no studies have reported a similar effect for multi-talker babble. The results of this study comparing open-set word recognition in multi-talker babble showed that listeners are significantly more accurate when the babble is fixed than when the babble is random. This documents the effect the authors refer to as "frozen babble."

PACS numbers: 43.71.Sy, 43.72.Dv, 43.71.Gv, 43.71.Es [JH]
Date Received: September 8, 2008 **Date Accepted:** December 17, 2008

1. Introduction

Previous studies have shown that listeners can adapt to particular repeated samples of identical noise, a phenomenon known as "frozen noise." For example, Langhans and Kohlrausch (1992) reported that the threshold for listeners to detect the presence of signals presented in frozen noise is significantly lower than for signals presented in random noise. However, no studies have reported such effects for multi-talker babble, a form of noise that is being used more in studies of speech perception and spoken word recognition due to its high level of ecological validity (e.g., Killion *et al.*, 2004; Cutler *et al.*, 2004; Wilson, 2003). In this paper, we report a subset of data from a larger study, in which a change in our methodology allows us to compare spoken word recognition performance of words mixed with a fixed segment of babble to spoken word recognition of words mixed with a random segment of babble.

2. Method

2.1 Materials

The stimulus list consisted of 1428 English words chosen from the Hoosier Mental Lexicon (HML; Nusbaum *et al.*, 1984), designed to be a representative sample of the entire English lexicon. To create a representative sample, the list was constructed such that it did not differ statistically from either the HML or the CELEX (Baayen *et al.*, 1993) on the following features: (1) number of phonemes, (2) number of syllables, (3) syllable structure, (4) initial phoneme, and (5) lexical frequency.

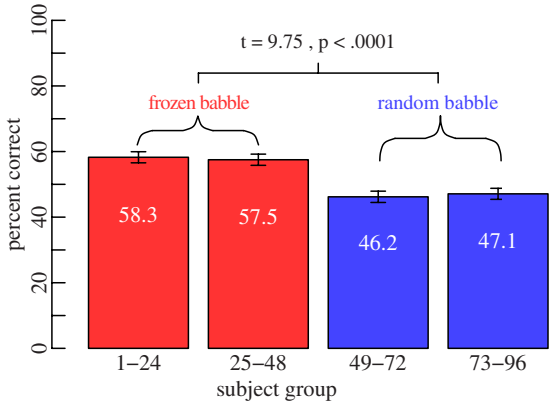


Fig. 1. (Color online) Percent correct of fixed and random babble groups.

Digital audio recordings of each word were created from the production of a male speaker of American English in an IAC sound-proof booth at a sampling rate of 22.05 KHz. Six-talker babble (three male and three female speakers) from the Connected Speech Test (Cox *et al.*, 1987) was added to the stimuli at three different signal-to-noise ratios (S/N): 0, 5, and 10 dB. The signal was centrally embedded in the babble, with a leading and trailing 420 ms of babble. The S/N ratio for each token was determined by comparing the rms average amplitude of the signal file with the babble file.

2.2 Procedure

The stimuli were presented to 96 native English-speaking undergraduates from Indiana University over Beyer-Dynamic D-210 headphones at 77 dB SPL. Each listener heard only one-quarter of the stimuli (357). One-third of the stimuli were presented at each S/N and were fully randomized such than no listener heard the same words at the same S/N. The experiment was self-paced and responses were typed on a keyboard.

2.3 Fixed versus random babble

After running the first 48 listeners, two changes in the methodology were made. The first change involved a switch from using a fixed portion of babble to a random portion of babble. That is, the stimuli presented to the first 48 listeners used a segment of multi-talker babble which always began at a fixed point. In contrast, the stimuli for the remaining 48 listeners were mixed with randomly selected segments of multi-talker babble.

In addition to the fixed versus random babble difference, a slightly different leveling procedure was used for the stimuli presented to the final 48 listeners. The level of the stimuli with fixed babble was equated before mixing in the multi-talker babble, which had the effect that the overall level of the stimuli increased as S/N decreased. Alternatively, the random babble stimuli were releveled after mixing in the babble, so that the average rms amplitude of all the stimuli was equal.

3. Results and discussion

Figure 1 shows the mean accuracy rates for listeners in the frozen and random babble conditions. The listeners in the random babble condition were significantly less accurate (mean =48.0, SD=0.303) on the word recognition task than the listeners in the frozen babble condition (mean=57.7, SD=0.307; $t=9.75$, $p<0.0001$). To determine whether these differences were due to random subject factors, the listeners in each condition were split in half and the two groups were compared. No significant difference was found between the two subgroups in either condition.

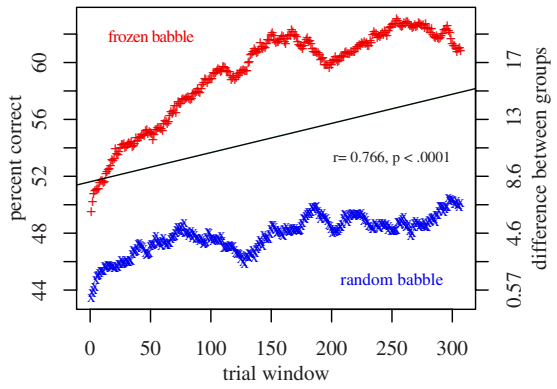


Fig. 2. (Color online) Learning rate for fixed and random babble. Each point corresponds to the mean percent correct for all subjects in the respective condition over a 50 trial window starting with trials 1–50 and ending with trials 308–357. The left axis shows percent correct. The line shows the least-squares fit to the difference in percent correct between the two groups for each 50 trial window and is represented by the right axis.

The significant difference in word recognition accuracy between the listeners in the fixed babble condition and the random babble condition is consistent with the claim that the listeners in the former condition adapted to the frozen babble. However, it remains possible that the difference was related to the releveling of the stimuli.[1] To address this, we examined changes in accuracy over the course of the experiment. If the accuracy difference comes from listeners adapting to the frozen babble, we should see an improvement over the course of the experiment (as they become more familiar with the noise pattern). Note that it is common for listeners to improve over the course of an experiment as they become more familiar with the task. It is likely that the listeners in the random babble condition will also show some learning, but not as much as the listeners in the fixed babble condition. If listeners in the fixed babble condition show a steeper learning curve than those in the random babble condition, we can conclude that the difference in accuracy is not due to the way the stimuli were leveled, but rather to the difference between fixed and random babble.

Figure 2 displays the accuracy for subjects in each condition over a moving 50-trial window. The first point represents trials 1–50, the second point 2–51, and so on. To determine whether these learning rates were significantly different, the frozen babble values were subtracted from the random babble values, and a Pearson's r correlation test was performed between these differences and the trial window. If the learning rates are the same, then there should be no correlation (as the difference should be a horizontal line). However, a significant positive correlation indicates that the frozen babble group shows a steeper learning rate. This analysis revealed a strong positive correlation ($r=0.766$; $p<0.001$), consistent with the claim that the difference in accuracy shown in Fig. 1 is an example of the frozen noise phenomenon.

In order to determine whether the frozen noise phenomenon can be changed based on the S/N ratio in the stimuli, we also analyzed the data at each S/N ratio. Analysis of the learning rate between the fixed and random babble groups was significant at each S/N ratio, as shown in Fig. 3. In addition, learning rate was computed for each listener as the slope of the least-squares fit regression line to the moving window data for each listener. A 2×3 ANOVA was carried out with learning rate as the dependent variable, babble type (fixed versus random) as between subjects factor, and S/N (0, 5, and 10 dB) as within subjects factor. The ANOVA showed babble type to be a significant factor (fixed=0.0333, random=0.0122, $F=7.4284$, $p<0.01$), but neither S/N ($F=1.0649$, $p>0.3$) nor the S/N by babble type interaction ($F<1$) was significant.

4. Conclusions

Our results indicate that the frozen noise phenomenon affects listeners who listen to stimuli mixed with the same set of multi-talker babble. Although this outcome is expected given the

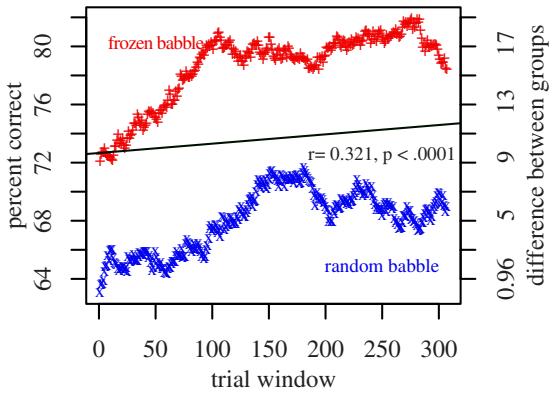


Fig. 3. (Color online) Learning rate for fixed and random babble by S/N. The axes are the same as in Fig. 2 but broken down for each S/N used.

literature on frozen noise, this has not been previously reported for multi-talker babble, which has been used in a number of studies in recent years. Some of these studies have used frozen babble (e.g., Cutler *et al.*, 2004; Engen and Bradlow, 2007),[2] while others have used random

babble (e.g., Killion *et al.*, 2004; Wilson, 2003). Depending upon the research questions being investigated, the use of frozen babble may be desired. It is our hope that this finding will aid researchers in designing future experiments using stimuli mixed with multi-talker babble.

[1]A recent study by Engen (2007) found that releveling stimuli of different S/N ratios had little effect. Nevertheless, this possibility will be considered here.
[2]Note that Engen and Bradlow (2007) repeated the same segment of babble in their six-talker babble condition, while they alternated randomly between four different segments of babble in the two-talker babble condition.

Baayen, H. R., Piepenbrock, R., and Rijn, H. (**1993**). "The CELEX lexical database," (CD-ROM) (Linguistics Data Consortium, University of Pennsylvania, Philadelphia).
Cox, R. M., Alexander, G. C., and Gilmore, C. (**1987**). "Development of the connected speech test (cst)," Ear Hear. **8**, 119S–126S.
Cutler, A., Weber, 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.
Engen, K. J. V. (**2007**). "A methodological note on signal-to-noise ratios in speech research," J. Acoust. Soc. Am. **122**, 2994.
Engen, K. J. V., and Bradlow, A. R. (**2007**). "Sentence recognition in native- and foreign-language multi-talker background noise," J. Acoust. Soc. Am. **121**, 519–526.
Killion, M. C., Niquette, P. A., and Gudmundsen, G. I. (**2004**). "Development of a quick speech-in-noise test for measuring signal-to-noise ratio loss in normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **116**, 2395–2405.
Langhans, A., and Kohlrausch, A. (**1992**). "Differences in auditory performance between monaural and diotic conditions. I. Masked thresholds in frozen noise," J. Acoust. Soc. Am. **91**, 3456–3470.
Nusbaum, H. C., Pisoni, D. B., and Davis, C. K. (**1984**), "Sizing up the hoosier mental lexicon: Measuring the familiarity of 20,000 words," Research on Speech Perception Progress Report 10, Speech Research Laboratory, Psychology Department, Indiana University, Bloomington.
Wilson, R. H. (**2003**). "Development of a speech-in-multitalker-babble paradigm to assess word-recognition performance," J. Am. Acad. Audiol **14**, 453–470.

 [DOI: 10.1121/1.3062148] Published Online 5 February 2009

Effect of a competing instrument on melodic contour identification by cochlear implant users

John J. Galvin III, Qian-Jie Fu, and Sandra I. Oba
Department of Communication and Auditory Neuroscience, House Ear Institute, Los Angeles, California 90057
jgalvin@hei.org, qfu@hei.org, soba@hei.org

Abstract: Cochlear implant (CI) users have limited access to fundamental frequency (F0) and timbre cues, which are needed to segregate competing voices and/or musical instruments. In the present study, CI users' melodic contour identification was measured for three target instruments in the presence of a masker instrument; the F0 of the masker was varied relative to the target instruments. Mean CI performance significantly declined in the presence of the masker, while mean normal-hearing performance was largely unaffected. However, the most musically experienced CI users were able to make use of timbre and F0 differences between instruments.

PACS numbers: 43.66.Ts, 43.66.Jh, 43.66.Hg, 43.75.Cd [DD]
Date Received: September 8, 2008 **Date Accepted:** December 10, 2008

1. Introduction

In music, timbre cues are important for segregating multiple instruments presented in a polyphonic context. When multiple instruments are played simultaneously or non-simultaneously, timbre cues can be used to track the melodic components within a piece of music. Due to the limited spectral resolution currently available in cochlear implants (CIs), CI users often have great difficulty in understanding, perceiving, and appreciating music, especially large ensemble performances (e.g., Looi *et al.*, 2008). Melodic pitch perception is difficult for CI users, even with a single instrument (e.g., Guerts and Wouters, 2004; Kong *et al.*, 2004). Melodic contour identification (MCI) has been used to quantify listeners' melodic pitch perception. Galvin *et al.* (2007) found that CI users' MCI performance was quite varied, depending on musical experience before and after implantation, and was generally poorer than that of normal-hearing (NH) listeners. Galvin *et al.* (2008) also showed that timbre cues can significantly affect CI users' MCI performance, suggesting that melodies might sound quite different (in terms of pitch sequences), depending on which instrument is playing. In general, as musical listening tasks become more complex, CI users experience greater difficulty. Musically experienced CI users may encounter less difficulty, and training has been shown to improve CI users' timbre perception (Gfeller *et al.*, 2002) and MCI performance (Galvin *et al.*, 2007).

However, even musically experienced CI users can have great difficulty with polyphonic music. Looi *et al.* (2008) showed that instrument identification was much better with solo instruments than with small or large ensembles. Due to the lack of spectro-temporal fine structure cues, CI listeners have difficulty using timbre cues to segregate instruments, resulting in distorted or confusing melody lines. Currently, it is not well known how CI listeners' melodic pitch perception may be influenced by a competing instrument. In the present study, MCI was measured for different instruments (organ, violin, and piano), with or without a competing instrument (piano). The masking contour was always flat (no change in pitch). Four masker notes (A3, A4, A5, and A6) were tested to observe whether listeners could use F0 and timbre differences between the masker and target instruments to identify the target contour.

2. Methods

Seven CI and seven NH subjects participated in the present experiment. All CI subjects participated in previous MCI studies (Galvin *et al.*, 2007, 2008), and thus were familiar with the MCI task and instrument stimuli. CI subject demographics are shown in Table 1. Note that subjects

Table 1. CI subject demographics.

Subject	Gender	Age	CI experience (years)	Device/strategy
S1	M	50	1	Freedom/ACE
S2	F	63	4	N24/ACE
S3	M	56	17	N22/SPEAK
S4	M	49	15	N22/SPEAK
S5	M	77	11	N22/SPEAK
S6	M	74	8	CII/Fidelity 120
S7	M	60	16	N22/SPEAK

S1 and S2 had greater music experience before and after implantation, compared with the other CI subjects. The mean age for NH subjects was 37.9 years (range: 20–52 years). Two of the NH subjects were active musicians, and five of the NH subjects had some musical instruction (e.g., piano, cello, organ, or drum lessons) during childhood. All NH subjects participated in the previous study by Galvin *et al.* (2008), and thus were experienced with the MCI task and instrument stimuli. All subjects were paid for their participation, and all provided informed consent before participating in the experiment.

Target stimuli consisted of nine five-note melodic contours (rising, flat, falling, flat-rising, falling-rising, rising-flat, falling-flat, rising-falling, and flat-falling), similar to those used in the previous MCI studies (Galvin *et al.*, 2007, 2008). All notes in the contours were generated according to $f=2^{x/12}f_{\mathrm{ref}}$, where f is the frequency of the target note, x is the number of semitones relative to the root note, and f_{ref} is the frequency of the root note (A3, the lowest note of the contour). The spacing between successive notes in each contour was varied to be one, two, three, four, or five semitones. Across all target contours and interval conditions, the F0 range was 220–698 Hz. Each note was 250 ms in duration, and the interval between notes was 50 ms. The contours were played by sampled instruments with synthesis (Roland Sound Canvas with Microsoft Wavetable synthesis). The three target instruments were organ, violin, and piano. The three instruments were selected to represent a range of spectral and temporal complexities. For example, the organ has a relatively simple spectrum with little attack, the violin has a more complex spectrum with a slow attack, and the piano has the most complex spectrum with a sharp attack. In the previous study by Galvin *et al.* (2008), CI users' MCI performance was best with the organ and poorest with the piano. Baseline performance for the three target instruments was measured with no masker. The masker stimuli consisted of the flat contour played by the piano, i.e., each note was the same. To test the effect of F0 overlap between the target and masker contours, the F0 of the masker was varied to be either A3 (220 Hz), A4 (440 Hz), A5 (880 Hz), or A6 (1720 Hz). The masker and target contours were normalized to have the same long term rms amplitude (65 dB).

Polyphonic music often involves multiple instruments that interleave and overlap in time, and listeners may use temporal offsets between instruments to track different melodic components. When played simultaneously by multiple sound sources, it is more difficult to track the melodic components. In the present study, the masker and target were presented simultaneously; the onset, duration, and offset for each note of the target and masker were the same.

Figure 1 shows electrodograms for the target A3 rising contour (five semitones between notes) for the organ and piano, with or without the piano masker; audio examples of these stimuli are given in Mm. 1. The electrodograms were generated using custom software that implemented the default parameter settings of the Advanced Combination Encoder (ACE) (Vandali *et al.*, 2000) strategy typically implemented in Nucleus-24 and Freedom implant devices, i.e., frequency allocation Table 6, 900 Hz/channel, 8 maxima, etc. With no masker, the rising contour can be easily observed, as the F0s and harmonics span most of the electrode array; note that the stimulation pattern is more spectrally dense and temporally complex for the

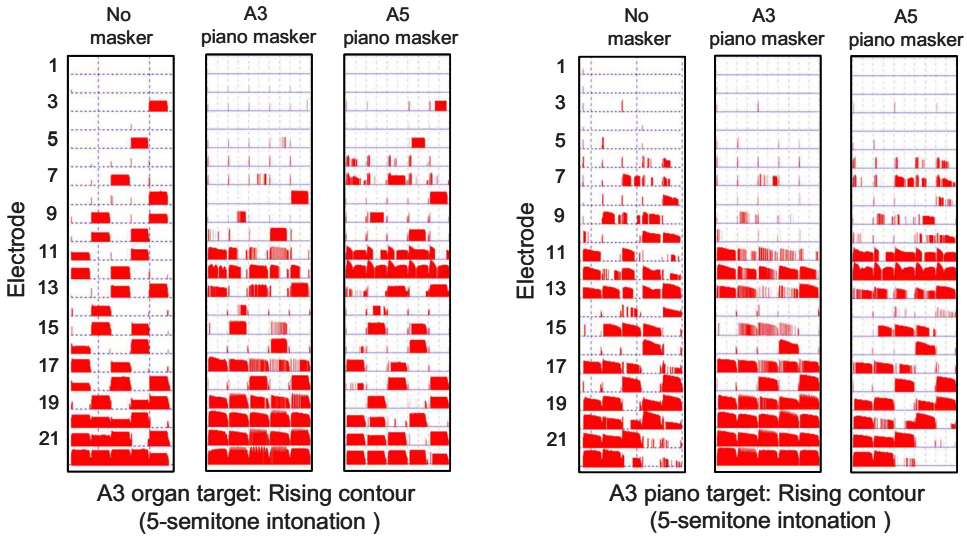


Fig. 1. (Color online) Left panels: electrodograms for the A3 rising contour (five semitones between notes) played by the organ with no masker (left), the A3 piano masker, and the A5 piano masker. The piano masker was the flat contour. The left axis shows the electrode number. Right panels: similar electrodograms for the A3 piano target rising contour.

piano than for the organ. With the A3 piano masker, the stimulation pattern is denser, as the eight spectral maxima are shared between the masker and target instruments; again, the stimulation pattern is denser for the piano than for the organ. With the A5 piano masker, the stimulation pattern is less dense, and the rising contour can be observed across the apical channels; again, the stimulation pattern remains denser and more complex for the piano than for the organ. Note that these stimulation patterns could be different with different speech parameter settings (e.g., coding strategy, acoustic frequency allocation, etc.).

[Mm. 1. Audio example of the target A3 rising contour (5 semitones between notes) for the organ and piano, with or without the piano masker. This is a file of type wav (546 Kb).]

For both NH and CI subjects, stimuli were presented acoustically via a single loudspeaker (Tannoy Reveal) at 70 dBA. Testing was conducted in a sound-treated booth (IAC) with subjects directly facing the loudspeaker. CI subjects were tested using their clinically assigned speech processors. CI subjects were asked to use their everyday sensitivity and volume settings; once set, subjects were asked not to change these settings. NH subjects listened to unprocessed signals. The different target/masker conditions were tested independently and randomized across subjects; each condition was tested three times. During each test block, a target stimulus was randomly selected (without replacement) from among the 45 stimuli (9 contours $\times$ 5 intonations). Subjects were asked to click on one of the nine response choices shown onscreen, after which the next target stimulus was presented. Subjects were instructed to guess if they were unsure. Subjects were explicitly instructed that they would hear two simultaneous contours, one of which would always be the flat piano masker, and that the pitch of the masker might change from test to test. Subjects were allowed to repeat each stimulus up to three times; however, no preview or trial-by-trial feedback was provided.

3. Results

Figure 2 shows mean MCI performance for CI and NH subjects, as a function of target instrument. For CI subjects, baseline performance with no masker was 79.1% correct for the organ, 73.3% correct for the violin, and 60.5% correct for the piano. A two-way repeated-measures analysis of variance (RM ANOVA) showed that CI performance was significantly affected by

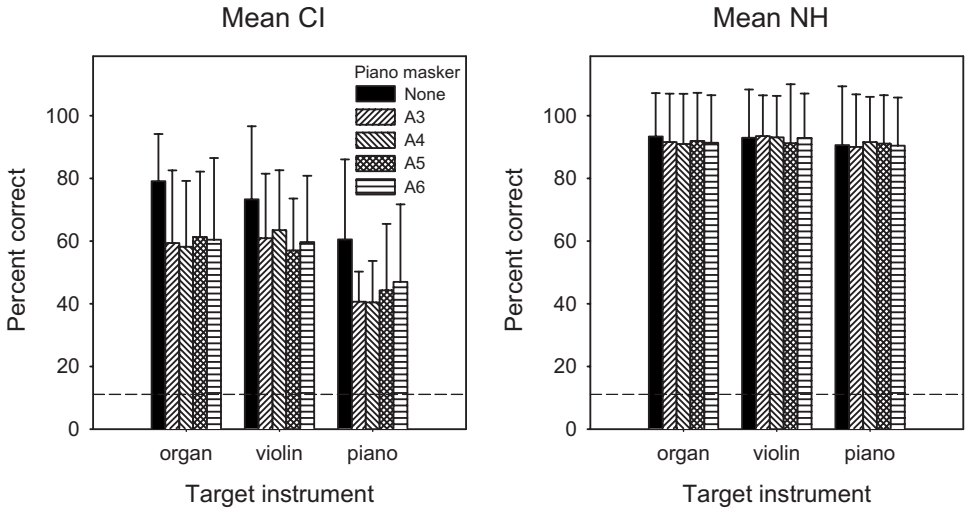


Fig. 2. Mean MCI performance across CI subjects (left) and NH subjects (right) for the different piano masker F0s, as a function of target instrument. The error bars show one standard deviation of the mean, and the dashed line shows chance performance level (11.1% correct).

target instrument [$F(2,48)=10.02$, $p=0.003$] and masker note [$F(4,48)=8.03$, $p<0.001$]. *Post hoc* Bonferroni t-tests showed that performance was significantly better without the piano masker ($p<0.05$), and that there was no significant difference among masker note conditions. *Post hoc* Bonferroni t-tests also showed that masked MCI performance was not significantly affected by the target instrument, except for the A4 piano masker (violin > piano; $p>0.05$). For NH subjects, mean performance with no masker was 93.3% correct for the organ, 93.0% correct for the violin, and 90.6% correct for the piano. A two-way RM ANOVA showed no significant effects for target instrument [$F(2,48)=3.83$, $p=0.052$] or masker note [$F(4,48)=0.49$, $p=0.743$].

While mean CI users' masked MCI performance was largely unaffected by differences in target instrument timbre or by differences in masker note, individual CI subjects exhibited markedly different patterns of results. Figure 3 shows individual CI subjects' MCI performance, as a function of target instrument. Subject S1's baseline performance was >97% correct for all three target instruments. While the piano masker generally reduced performance, S1 was sensitive to target instrument timbre (i.e., better performance when masker and target instruments were different) and masker note (i.e., better performance as the F0 separation between masker and target was increased). Subject S2 also exhibited some sensitivity to target instrument timbre and masker note. S2's performance with the organ was virtually unchanged by the presence of the piano masker. For the violin and piano targets, increasing the F0 separation between the masker and target improved performance. The pattern of results was different for subject S4. While S4 was somewhat sensitive to target instrument timbre, increasing the F0 separation between masker and target reduced performance. For subjects S3 and S5, the effects of instrument timbre and masker note were inconsistent. Subjects S6 and S7 exhibited little sensitivity to target instrument timbre or masker note.

4. Discussion

In general, CI subjects generally had great difficulty in perceiving the target melodic contours in the presence of a competing simultaneous instrument masker. Most CI subjects were unable to make use of timbre and/or pitch differences between the masker and target stimuli. This deficit in masked MCI performance is in agreement with previous studies showing poorer instrument identification in the context of small or large instrument ensembles (Looi *et al.*, 2008). The

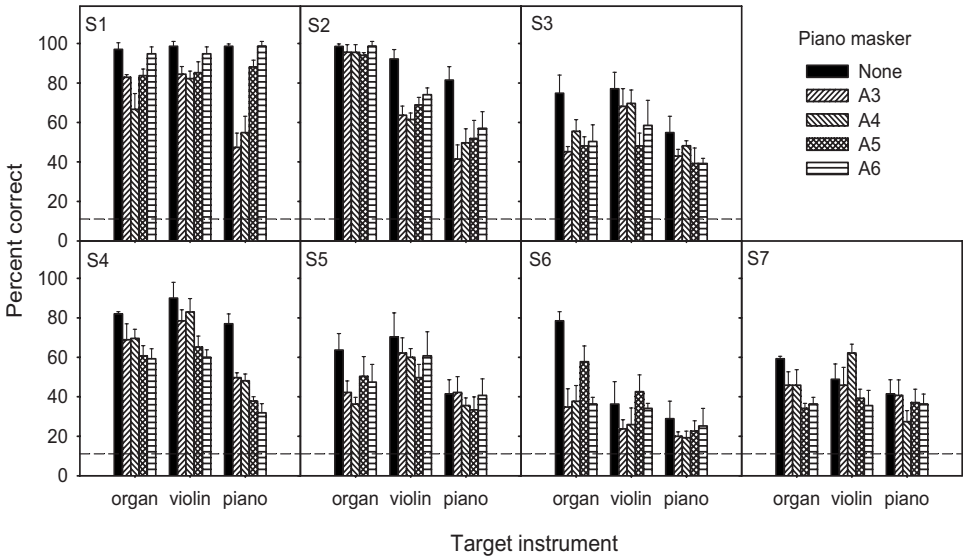


Fig. 3. MCI performance for individual CI subjects for the different piano masker F0s, as a function of target instrument. The error bars show one standard deviation of the mean, and the dashed line shows chance performance level (11.1% correct).

mean CI results are also in agreement with Stickney *et al.* (2007), who showed that CI listeners were unable to make use of F0 differences between talkers to recognize speech in the presence of a competing talker. Note that the MCI task in the present experiment required listeners to attend exclusively to pitch cues, as opposed to a speech recognition task in which pitch cues are used to segregate talkers and access lexically meaningful information. Given the weak F0 coding in speech processors, CI users must make use of spectral envelope differences and/or periodicity cues to track changes in pitch. Because of the relatively coarse spectral resolution, competing speakers or instruments cannot be segregated, even with fairly large F0 differences.

Interestingly, some CI subjects (S1 and S2) were able to consistently use timbre and F0 differences to identify the target contours. These subjects were the most musically experienced, which may have contributed to their better attention to timbre and pitch cues. Subject S1 exhibited a somewhat predictable pattern, i.e., masked MCI performance improved as the timbre and/or F0 differences were increased between the target and the masker. S2 exhibited a similar pattern (albeit poorer overall performance) with the violin and piano target instruments. For the remaining less musically experienced CI subjects, there were no clear effects for target instrument timbre and/or masker F0. For example, performance worsened for S4 as the masker F0 was increased, suggesting that the highest spectral component dominated the percept. For subject S7, the presence of the competing masker only marginally reduced the already low MCI scores. The different devices and processing strategies used by CI subjects may have also contributed to differences in individual performance, although there are too few subjects to fairly compare across CI devices. Despite differences among devices and strategies, contemporary CIs do not provide the spectro-temporal fine structure cue to support complex pitch perception and timbre discrimination. As such, musically experienced CI users may be better able to use the limited spectral and temporal envelope cues for music perception. Previous studies also show that targeted training improved CI users' music perception (Galvin *et al.*, 2007; Gfeller *et al.*, 2002), even though CI subjects had years of experience with their clinical devices and strategies. Until CIs can provide fine structure cues, music experience and/or training may provide the greatest advantage for music perception with electric hearing.

Acknowledgments

The authors would like to thank all the subjects who participated in this study. The authors would also like to thank Dr. Diana Deutsch and two anonymous reviewers, as well as David Landsberger and Bob Shannon, for insightful comments on this work. This work was supported by NIH Grant No. DC004993.

References and links

Galvin, J., Fu, Q.-J., and Nogaki, G. (**2007**). "Melodic contour identification by cochlear implant listeners," Ear Hear. **28**, 302–319.

Galvin, J., Fu, Q.-J., and Oba, S. (**2008**). "Effect of instrument timbre on melodic contour identification by cochlear implant users," J. Acoust. Soc. Am. **124**, EL189–EL195.

Geurts, L., and Wouters, J. (**2004**). "Better place coding of the fundamental frequency in cochlear implants," J. Acoust. Soc. Am. **115**, 844–852.

Gfeller, K., Witt, S., Adamek, M., Mehr, M., Rogers, J., Stordahl, J., and Ringgenberg, S. (**2002**). "Effects of training on timbre recognition and appraisal by postlingually deafened cochlear implant recipients," J. Am. Acad. Audiol **13**, 132–145.

Kong, Y.-Y., Cruz, R., Jones, J., and Zeng, F.-G. (**2004**). "Music perception with temporal cues in acoustic and electric hearing," Ear Hear. **25**, 173–185.

Looi, V., McDermott, H., McKay, C., and Hickson, L. (**2008**). "Music perception of cochlear implant users compared with that of hearing aid users," Ear Hear. **29**, 421–434.

Stickney, G., Assmann, P., Chang, J., and Zeng, F.-G. (**2007**). "Effects of cochlear implant processing and fundamental frequency on the intelligibility of competing sentences," J. Acoust. Soc. Am. **122**, 1069–1078.

Vandali, A. E., Whitford, L. A., Plant, K. L., and Clark, G. M. (**2000**). "Speech perception as a function of electrical stimulation rate: Using the Nucleus 24 cochlear implant system," Ear Hear. **21**, 608–624.

 [DOI: 10.1121/1.3057860] Published Online 5 February 2009

Investigation of four distinct glottal configurations in classical singing—A pilot study

Christian T. Herbst
Laboratory of Biophysics, Department of Experimental Physics, Faculty of Science, Palacký University Olomouc, tř. Svobody 26, 77146 Olomouc, Czech Republic, and Tölzer Knabenchor, Herbert-von-Karajan-Platz 2, 5020 Salzburg, Austria
herbst@ccrma.stanford.edu

Sten Ternström
Department of Speech, Music and Hearing, School of Computer Science and Communication, Kungliga Tekniska Högskolan, Lindstedtsvägen 24, SE-100 44 Stockholm, Sweden
stern@kth.se

Jan G. Švec
Laboratory of Biophysics, Department of Experimental Physics, Faculty of Science, Palacký University Olomouc, tř. Svobody 26, 77146 Olomouc, Czech Republic
svecjan@vol.cz

Abstract: This study investigates four qualities of singing voice in a classically trained baritone: "naïve falsetto," "countertenor falsetto," "lyrical chest" and "full chest." Laryngeal configuration and vocal fold behavior in these qualities were studied using laryngeal videostroboscopy, videokymography, electroglottography, and sound spectrography. The data suggest that the four voice qualities were produced by independently manipulating mainly two laryngeal parameters: (1) the adduction of the arytenoid cartilages and (2) the thickening of the vocal folds. An independent control of the posterior adductory muscles versus the vocalis muscle is considered to be the physiological basis for achieving these singing voice qualities.

PACS numbers: 43.75.Rs, 43.70.Gr [TR]
Date Received: June 26, 2008 **Date Accepted:** November 25, 2008

1. Introduction

The ability to control voice quality in singing is crucial for a singer, yet the information on the specific mechanisms used in singing has been unsatisfactory. Probably the most controversial topic of singing voice quality from the historical perspective has been that of voice registers (Titze, 2000). The two main registers of the singing voice, particularly in males, have been recognized to be the chest and falsetto register. Most singers, however, are capable of blending the registers and produce more than just two distinct voice qualities. Information on how these qualities can be achieved has been largely insufficient.

This study investigates a classically trained baritone who demonstrated the ability to produce four distinct voice qualities, which he considered to be the "building blocks" of his singing technique. He called those four qualities as Type A "*Naïve singer's falsetto*," Type B "*Countertenor falsetto*," Type C "*Lyrical style*," and Type D "*Full chest*." Acoustic recordings of these four phonation types are demonstrated in the audio file Mm. 1. Based on his self-perception, the subject claimed that these four voice qualities are created with four different laryngeal configurations. The specific goal of this study was to investigate these phonations and the laryngeal adjustments in these four phonation types. The more general goal was to establish better understanding of the laryngeal adjustment strategies that are used to control the voice quality in singing.

[Mm. 1. Acoustic recordings of the four phonation types A–D (naïve falsetto, counter tenor falsetto, lyrical chest, full chest). This is a file of type .wav (1.1 MB).]

2. Material and methods

The investigated subject was a baritone with a university degree in voice pedagogy and 15 years of experience in classical singing (the author C.H.). He produced eight to ten sustained phonations in each of the four phonation qualities at a fundamental frequency of 294 Hz (tone D4). This frequency was at the second passaggio of the baritone, where both chest and falsetto phonations were possible. Vowel /i/, which allows examination through rigid laryngoscopy, was chosen for all phonatory tasks.

The adjustment and vibration of the vocal folds was observed with a rigid endoscope using two alternative techniques—laryngeal videostroboscopy (Bless *et al.*, 1987; Baken and Orlikoff, 2000) and videokymography (Švec and Schutte, 1996). The specific audiovisual equipment used was identical to the one used in the study of Švec *et al.* (2008). The microphone and electroglottographic signals were recorded simultaneously with the video signals and stored in the two audio channels of the final digital video file. The video signals were digitized using a Pinnacle DV 500 System set to full PAL mode (25 images/s, 720 × 525 pixels resolution) and stored as AVI files using the Intel Indeo 5.0 codec.

The subject repetitively produced each phonation type while the position and focus of the endoscope was adjusted to obtain videokymographic images of the best quality at the place of maximum vibration amplitude of the vocal folds. After satisfactory videokymographic images were obtained from each phonation types, the videokymographic camera was replaced by the standard color camera and the continuous light was replaced by stroboscopic light. Then the subject repeated the phonations for all the types and strobovideolaryngoscopic recordings were done.

From laryngeal videostroboscopy two images were extracted, both at maximum glottal closure and maximum glottal opening to document the laryngeal configuration as well as the vibration mode of the vocal folds. From videokymographic recordings, representative phonations were extracted from the stable portion in the middle of each video sample; no voicing onsets and offsets were included in the analyzed data. For the sustained notes sung in the four distinct phonation types, the following criteria had to apply for data to be selected for further evaluation: (1) Phonation was stable. (2) The minimum duration exceeded one second. (3) The subject confirmed that in his opinion the phonation did represent the attempted phonation type. (4) Videokymographic images showed sharp and clear glottal contours and the scan line was located approximately at the place of the maximum vibration amplitude of the vocal folds. In this way three to five representative video samples per phonation type were selected out of a total of 115 videokymographically recorded phonations, resulting in a total of 15 samples.

The videokymographic closed quotient (CQ_{VKG}) was determined using the formula $CQ_{VKG}=t_c/T_0$ where t_c is the duration of the closed phase and T_0 is the duration of the vibratory cycle. The durations were measured by manually counting the number of pixels for open and closed phase. The pixel spacing corresponded to a time interval of 128 μs, or about 26.5 pixels per period of the fundamental frequency. For each of the 15 samples of stable phonation, the CQ_{VKG} was calculated from four consecutive glottal cycles at two or three (in short or long recordings, respectively) different instances equally distributed over the duration of the phonatory sample. The data points were separated by at least eight video frames (320 ms). CQ_{VKG} readings were then averaged per phonation type.

The microphone signal was low-pass filtered with a cutoff-frequency of 8 kHz (to remove an artifact noise around 11 kHz produced by the light source during videokymographic recordings) and the sampling rate was reduced to 16 kHz. The alpha ratio was calculated as the ratio of the high-band energy (1000–5000 Hz) and the low-band energy (up to 1000 Hz) of the acoustic signal, expressed in dB (Löfqvist and Mandersson, 1987). Typical alpha ratio values are negative numbers, which increase (i.e., become less negative) as the spectrum of the signal becomes flatter (i.e., the signal has stronger high-frequency components).

The electroglottographic signal was used to obtain the electroglottographic contact quotient (CQ_{EGG}), which was calculated by a criterion level method with a threshold level of 25% as has been used, e.g., by Orlikoff (1991) or Herbst and Ternström (2006). Since the elec-

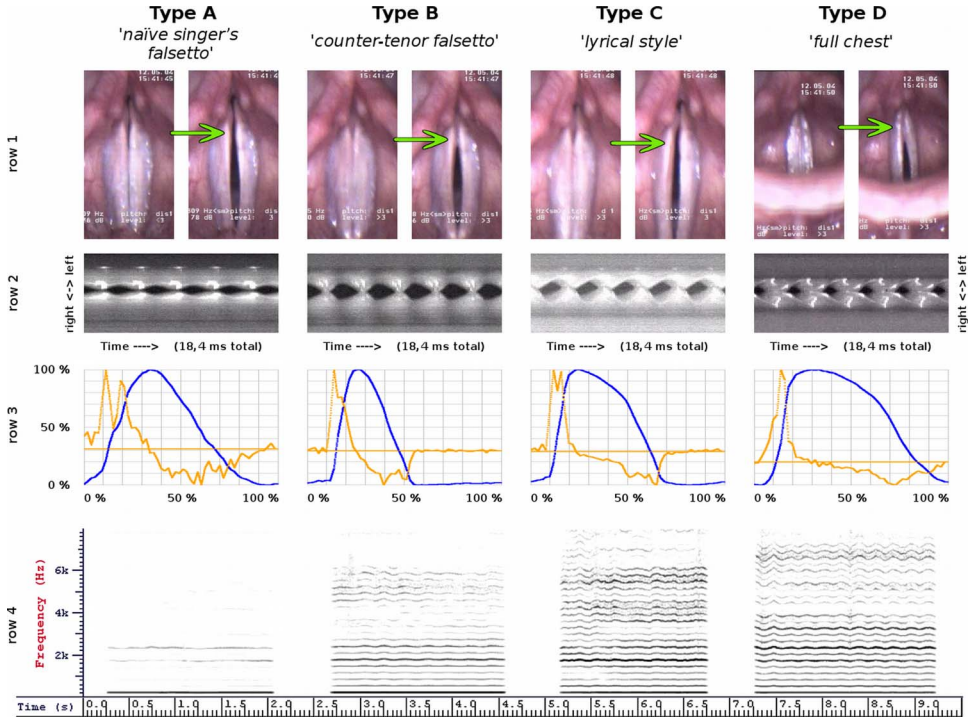


Fig. 1. (Color online) Images and signals documenting the vocal fold vibration in the four phonation types A–D. From top to bottom: Row 1: Pairs of videostroboscopic images at the phases of maximum vocal fold contact and maximum glottal opening. The arrows point at the position of the vocal processes of the arytenoid cartilages which are apart in types A and C and closed in types B and D. Row 2: Typical videokymographic images at the place of maximal vibration amplitude of the vocal folds. Row 3: Typical EGG signals (dark/blue) and their first derivative (light/orange online) for phonation types A–D, normalized both in amplitude and time. The *x*-axis represents normalized glottal cycle duration, and the *y*-axis shows EGG and DEGG signal amplitude, locally normalized per glottal cycle. Row 4: Spectrogram of sustained notes (taken from the audio channel of the videokymographic recordings) produced with phonation types A, B, C, and D consecutively. The videostroboscopic recordings of the four phonation types are documented in the media file below.

troglottograph used had no automatic gain controller, it was also possible to evaluate the strength of the EGG signal, which was quantified by computing the time-varying rms value of the signal, using a window duration of 125 ms. The values were expressed on a dB scale with an arbitrary reference value.

3. Results

Examples of the videostroboscopic recordings of the four phonation types are given in the video file Mm. 2.

> [Mm. 2. Videostroboscopic recordings of the four phonation types A–D (naïve falsetto, counter tenor falsetto, lyrical chest, full chest). This is a file of type .avi (7.4 MB).]

The laryngeal adjustments in the four types of phonations are shown in the top section of Fig. 1. The images reveal distinct adjustments of the laryngeal structures and vibratory features of the vocal folds for each phonation type. The most distinct differences were seen (1) at the posterior, cartilaginous part of glottis, which was varying between slightly open (types A and C) and closed (types B and D); (2) at the vocal processes of the arytenoid cartilages (marked by arrows in the stroboscopic images of Fig. 1, row 1), which were in some cases vibrating with the vocal folds (types A and C) and in other cases pressed together and not vibrating (type D; in phonation type B, the vocal processes were intermittently pressed together

Table 1. Glottal configurations and vibratory features, as seen in laryngeal videostroboscopy for phonation types A–D.

	Type A ("naïve singer's falsetto")	Type B ("countertenor falsetto")	Type C ("lyrical style")	Type D ("full chest")
Posterior glottis	slightly abducted	closed	barely closed	closed
Vocal processes vibration	yes[a]	intermittent[b]	yes[a]	no[c]
Mucosal waves extent	very short[d]	very short[d]	medium	long

[a]Vibrating together with the focal folds.
[b]Vocal processes alternated between slightly vibrating and pressed together.
[c]Vocal processes pressed together.
[d]Seen only in posterior part of vocal folds.

or vibrating with the vocal folds); and (3) in the mucosal waves on the vocal folds, the extent of which varied from long (type D) to very short (types A and B). The specific findings for each of the four phonation types are summarized in Table 1.

The videokymographic images revealed distinct differences in the vibratory pattern of the vocal folds among the four phonation types (Fig. 1, row 2). Considering the categorization of vibration characteristics as specified by Švec *et al.* (2007), the most prominent differences were recognized in (1) duration of closed phase, which increased when going from type A to type D; (2) roundedness of the lateral peaks, which successively sharpened when going from type A to type D; and (3) the extent of laterally travelling mucosal waves (only short mucosal waves for types A and B in contrast to medium/long mucosal waves for types C and D, respectively). The specific findings for each of the four phonation types are summarized in Table 2.

For acoustic analysis, the audio signal was extracted from the same 15 video samples that were considered for CQ_{VKG} analysis. Spectral analysis showed an increasing energy content in high frequency partials from type A to D. For phonation types A, B, and C this corresponded to a steady increase of alpha ratio values (Table 3). The types C and D showed similar alpha ratios but different distribution of spectral maxima: type C showed more energy in the region of 3500–6000 Hz, whereas the D type had more energy in the region of 2300 to 3500 Hz (see Fig. 1, row 4).

The EGG signals showed an increase of vocal fold contact duration when going from type B through C to D. This was reflected in the increase of the calculated average CQ_{EGG} (Table 3). The EGG waveform for phonation type A had a much smaller amplitude than the other types

Table 2. Videokymographic findings for phonation types A–D. For closed quotient mean values and standard deviations are listed; the numbers in the parentheses give the ranges of the values observed.

Phonation type	No. of phonations/ measurements	Closed quotient (CQ_{VKG})	Lateral peaks shape	Mucosal waves extent
A	3/32	**0**	rounded	none/barely visible
B	4/40	**0.28** ± 0.05 *[0.2, 0.37]*	rather rounded	none/barely visible
C	3/36	**0.47** ± 0.05 *[0.37, 0.57]*	rather sharp	medium
D	5/60	**0.69** ± 0.04 *[0.63, 0.78]*	sharp	long

Table 3. Acoustic and electroglottographic signal characteristics for phonation types A–D: alpha ratio, SPL values, CQ_{EGG}, and averaged EGG signal strength (expressed in dB with arbitrary reference). All computed data are indicated as mean values with standard deviation and the extreme values in the parentheses.

		Acoustic signal		Electroglottography	
Phonation type	No. of phonations	Alpha ratio (dB)	SPL@30 cm (dB *re* 20 μPa)	EGG signal strength (dB)	CQ_{EGG}
A	3	**−16.3** ± 0.8 *[−16.78, −15.46]*	**80** ± 1.8 *[77, 81]*	**−26** ± 1.3 *[−27, −24]*	**N/A**[a]
B	4	**−9.2** ± 1.6 *[−10.4, −6.3]*	**88** ± 0.9 *[87, 89]*	**−18** ± 0.3 *[−18, −18]*	**0.31** ± 0.01 *[0.30, 0.33]*
C	3	**0.3** ± 1.6 *[−1.4, 1.7]*	**84** ± 2.2 *[81, 86]*	**−16** ± 0.6 *[−17, −15]*	**0.48** ± 0.07 *[0.40, 0.59]*
D	5	**−1.1** ± 1.7 *[−2.5, 1.9]*	**82** ± 2.1 *[79, 84]*	**−17** ± 1.7 *[−20, −15]*	**0.65** ± 0.01 *[0.64, 0.66]*

[a]See text.

and a quasi-sinusoidal shape, which suggested that there was no full contact of the vocal folds. Thus, we concluded, in accordance with Herbst and Ternström (2006), that it is not appropriate to calculate a CQ_{EGG} based on that kind of signal.

4. Discussion

The data supported the original hypothesis that the four singing types are produced with four different laryngeal adjustments. The differences were directly observed with VKG and laryngeal videostroboscopy and were also reflected in the electroglottographic and microphone signals. The most remarkable factor for distinguishing the types A from B and C from D was found to be the configuration of the posterior glottis and the position of the vocal processes. In types B and D the posterior glottis was fully adducted and the vocal processes were mostly pressed together, thus actively shortening the vibrating part of the vocal folds to only the membranous part. On the other hand, in the A and C types the posterior glottis was slightly abducted and the vocal processes were participating in the vibration of the vocal folds (Table 1).

The singer indicated that the phonation types A and B were produced in falsetto voice whereas types C and D were of chest register quality. It has been known that chest versus falsetto control is physiologically achieved mostly by the vocalis muscle (Hirano, 1974; Titze, 2000). Activity of the vocalis muscle thickens the body of the vocal fold and slackens the vocal fold cover, resulting in vibration with larger vertical phase differences and more pronounced mucosal waves (Hirano, 1974; Titze, 2000). Generally, these characteristics are expected to be reflected in the increased sharpness of the lateral peaks in the videokymographic images of the vocal folds, as well as in the increased extent of the mucosal waves on the upper vocal fold surface (Švec *et al.*, 2008). Indeed, the videokymographic images of type A and B phonations show more rounded lateral peaks and shorter laterally traveling mucosal waves than the types C and D (Table 2) thus objectively supporting the singer's assumption that types A and B belong rather to falsetto register, whereas types C and D rather belong to chest register.

Based on these findings, it may be appropriate to call the phonation types A and B "abducted falsetto" and "adducted falsetto" [referred to as "open-chink falsetto" and "closed-chink falsetto" by Rubin and Hirt (1960)], whereas the types C and D can be seen as "abducted chest" and "adducted chest" registers, respectively. The differences in the closed phase seen in Table 2 can be explained as a consequence of these adjustments.

While register control is physiologically achieved mostly by the vocalis muscle (Hirano, 1974; Titze, 2000), posterior glottal adduction is known to be regulated by laryngeal adductors, such as posterior cricoarytenoid (PCA) and interarytenoid (IA) muscles (Titze, 2000). The present results suggest that in singing, the adduction of vocal processes is actively

used as a parameter in voice quality control, which is independent and separate from that of the (chest-falsetto) register control by the vocalis muscle. While the posterior glottal adduction has been recognized as an independent parameter in voice production (e.g., Titze, 2000), the use of this parameter separately from the chest-falsetto control has not, to our knowledge, been documented in singing before. An independent control and flexibility of the adductory muscles appears to be an important factor in allowing the production of different singing qualities.

This pilot study is limited in that it documents the laryngeal behavior laryngoscopically in only a single subject. However, it is not too far-fetched to conjecture that the independent control of the vocal fold body tension (through vocalis muscle) and posterior glottal adduction (through PCA and IA muscles) plays an important role in controlling singing voice quality in general. In classical music different voice qualities are expected for modulating the timbre within a musical piece, in order to enhance the expression of the artistic performance. It is conceivable that the lack of independent control of these two parameters may lead to problems in singing students who fail in singing specific styles. In future studies we plan to address this hypothesis further using a larger number of singer subjects.

5. Conclusions

The results of this study showed that the four phonation types were acoustically distinct and that they were indeed produced with different laryngeal settings. These settings could be explained by the independent manipulation of mainly two laryngeal parameters: (1) *the thickening of the vocal folds* and (2) *the adduction of the posterior glottis*. These two physiologic parameters represent two physiologically distinct types of glottal adduction: membranous adduction (adjustable by thyroarytenoid-vocalis muscles) and cartilaginous adduction (adjustable by cricoarytenoid and interarytenoid muscles). The two types of glottal adduction should be separated from each other when studying different voice qualities in singing.

Acknowledgments

The data for this study were obtained in 2004 during the authors' stays at the Department of Speech, Music and Hearing, KTH, Stockholm. C. Herbst's stay was supported by the Erasmus Student Exchange Programme of the European Commission, and J. Švec's stay was supported by an individual grant from the Wenner-Gren Foundation. In 2008, the study was supported by the Grant Agency of the Czech Republic, Project GAČR 101/08/1155. The authors thank Hans Larsson at the Department of Logopedics and Phoniatrics, Karolinska University Hospital at Huddinge, Stockholm, for his help in acquiring the laryngoscopic recordings.

References and links

Baken, R. J., and Orlikoff, R. F. (**2000**). *Clinical Measurement of Speech and Voice*, 2nd ed. (Singular Publishing Group, San Diego, CA).
Bless, D. M., Hirano, M., and Feder, R. J. (**1987**). "Videostroboscopic Evaluation of the Larynx," Ear Nose Throat J. **66**(7), 289–296.
Herbst, C., and Ternström, S. (**2006**). "A Comparison of Different Methods to Measure the EGG Contact Quotient," Logoped. Phoniatr. Vocol. **31**(3), 126–138.
Hirano, M. (**1974**). "Morphological Structure of the Vocal Cord as a Vibrator and its Variations," Folia Phoniatr. **26**(2), 89–94.
Löfqvist, A., and Mandersson, B. (**1987**). "Long-time average spectrum of speech and voice analysis," Folia Phoniatr. **39**(5), 221–229.
Orlikoff, R. F. (**1991**). "Assessment of the Dynamics of Vocal Fold Contact From the Electroglottogram: Data From Normal Male Subjects," J. Voice **34**(5), 1066–1072.
Rubin, H. J., and Hirt, C. C. (**1960**). "The falsetto: A high speed cinematographic study," Laryngoscope **70**, 1305–1324.
Švec, J., and Schutte, H. K. (**1996**). "Videokymography: High-speed line scanning of vocal fold vibration," J. Voice **10**(2), 201–205.
Švec, J., Šram, F., and Schutte, H. K. (**2007**). "Videokymography in Voice Disorders: What to Look For?," Ann. Otol. Rhinol. Laryngol. **116**(3), 172–180.
Švec, J., Sundberg, J., and Hertegård, S. (**2008**). "Three registers in an untrained female singer analyzed by videokymography, strobolaryngoscopy and sound spectrography," J. Acoust. Soc. Am. **123**(1), 347–353.
Titze, I. R. (**2000**). *Principles of Voice Production* (National Center for Voice and Speech, Iowa City, IA).

LETTERS TO THE EDITOR

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

Contribution of very low amplitude-modulation rates to intelligibility in a competing-speech task (L)[a)]

Christian Füllgrabe,[b)] Michael A. Stone, and Brian C. J. Moore
Department of Experimental Psychology, University of Cambridge, Downing Street, Cambridge CB2 3EB, England

(Received 30 July 2008; revised 16 December 2008; accepted 24 December 2008)

It is generally agreed that the slow fluctuations in the envelope of speech in different spectral channels carry critical information for intelligibility. Previous studies in which amplitude modulation (AM) was selectively removed from the speech signal showed that modulation rates between 4 and 16 Hz are most important, and that rates falling outside this range contribute little or not at all to speech intelligibility. The present study investigated the role of very low (<4 Hz) AM rates in the ability to identify sentences in an interfering background talker. The mixture was processed through a noise vocoder. The depth of AM with rates below 4, 1.3, or 0.4 Hz was reduced using a multi-channel envelope compressor with a high compression ratio. Data obtained using nine normal-hearing listeners demonstrate that low-rate AM, in the range 0.4–4 Hz, contributes to the intelligibility of relatively long speech utterances, at least for adverse listening conditions in which background noise is present and listeners are forced to rely on envelope cues in a few spectral channels. [DOI: 10.1121/1.3075591]

PACS number(s): 43.66.Mk, 43.71.Sy, 43.71.Es [RLF] Pages: 1277–1280

I. INTRODUCTION AND RATIONALE

The waveform of speech shows slow fluctuations in its global amplitude over time, resulting from articulatory movements; the most prominent envelope fluctuations occur at rates around 3–4 Hz, which correspond roughly to the average syllabic rate of speech (e.g., Houtgast and Steeneken, 1985). Evidence for the role of temporal-envelope cues in speech identification comes from cochlear-implant users, whose auditory nerves are electrically stimulated by pulse trains amplitude modulated by the speech envelope extracted from several spectral channels. Many implant users achieve good identification of speech in quiet (e.g., Wilson *et al.*, 1991). For normal-hearing listeners, studies using vocoder-processed stimuli (e.g., Shannon *et al.*, 1995) have demonstrated that the presence of low-rate (below ~50 Hz) amplitude modulation (AM) in only a few spectral channels is sufficient for excellent speech intelligibility in quiet, despite degraded spectral and temporal fine structure cues.

The relative importance of AM at low modulation rates was investigated in several psychoacoustical studies (e.g., Drullman *et al.*, 1994a, 1994b; Arai *et al.*, 1999). The work of Drullman *et al.* is of particular interest because the speech-reception threshold in noise (SRT; defined as the speech-to-noise ratio yielding 50% correct identification) of normal-hearing listeners was measured for fairly long sentences, processed using the following stages: (i) bandpass filtering into multiple channels, (ii) extraction of the Hilbert envelope of each channel, (iii) envelope manipulation by either low- or high-pass filtering at different cutoff frequencies (ranging from 0 to 64 Hz), and (iv) modulation of the fine structure in the source channel by the processed envelope. The results indicated that SRTs did not significantly worsen when AM with rates above 16 Hz or below 4 Hz was filtered out.

The claim that modulation rates below 4 Hz do not contribute to speech identification was recently questioned by Stone and Moore (2008). They used normal-hearing listeners to investigate the effect of multi-channel amplitude compression on the intelligibility of noise-vocoded sentences presented with a competing talker background. The target and background speech were combined and bandpass filtered before applying a compressor to the envelope extracted from each analysis channel. The "speed" of gain variation of the compressor was increased from "slow" (of the order of several hundred milliseconds) to "fast" (of the order of tens of milliseconds) by shortening both the attack and release times. As previously shown (Stone and Moore, 1992, 2003),

[a)]Work presented at the Acoustics '08 meeting, Paris, France, June 2008.
[b)]Electronic mail: cf277@cam.ac.uk

the compressor reduced the modulation depth of the "lower" AM rates only, and an increase in compressor speed led to the additional reduction of AM depth at "higher" rates. Stone and Moore (2008) used a moderate compression ratio (CR) similar to those commonly used in assistive hearing devices, so the compressor only partially removed AM at low rates. As expected, identification scores improved as the compressor speed decreased. Unexpectedly, for an 8-channel, but not for a 12-channel, vocoder/compressor, identification was worse than that obtained in a condition with no compression, even when the compressor only reduced the modulation depth for AM rates below 2 Hz. This result appears to suggest that AM rates below 2 Hz are important for intelligibility, although it is not clear why this would be the case for an 8-channel, but not a 12-channel, vocoder/compressor. Possibly, very low AM rates are only important when spectral information is very limited.

The present study was designed to investigate more closely the contribution of low-rate AM to speech-in-speech intelligibility for a range of spectral resolutions. The study was similar to that of Stone and Moore (2008), but there were three important differences. First, a larger CR was used to provide more effective and precise removal of temporal-envelope cues at low AM rates. Second, compressors with lower speeds were included, so as to investigate in more detail the influence of very low AM rates on speech intelligibility. Third, we used vocoders/compressors with three different numbers of channels, 6, 9, and 15, so as to span the range used by Stone and Moore, and to answer the question of whether the importance of AM at very low rates depends on the number of spectral channels.

II. LISTENERS, STIMULI, PROCESSING, AND PROCEDURE

Nine naive native-English speaking students (aged 22–28 years; four females) with normal hearing (<20 dB HL at all audiometric frequencies between 0.125 and 8 kHz) were paid for their participation. Their task was to identify sentences presented in an interfering-speech background. The target and background were both processed to remove all temporal fine structure information and to systematically vary the amount of spectral information; this forced the listener to use mainly temporal-envelope cues in a few spectral channels.

The target sentences were taken from the IEEE corpus (IEEE, 1969) and spoken by a male speaker of British standard English. The background utterances consisted of recordings of a different male speaker of British standard English reading prose passages. The two signals were mixed together with the background starting 1 s before and ending about 1 s after the target sentence. The target-to-background ratio (TBR) was either +10, +8, or +5 dB depending on the number of analysis/synthesis channels of the subsequent vocoder processing; these TBRs were chosen to avoid floor and ceiling effects in performance. The mixture was then processed through an N-channel noise vocoder incorporating an N-channel compressor; both systems were implemented in MATLAB. First, the signal was bandpass filtered into $N=6$, 9, or 15 spectral channels. Channels were equally spaced on an ERB_N-number scale (Glasberg and Moore, 1990). The corner frequencies of the filters were (i) 100, 331, 725, 1396, 2538, 4485, and 7800 Hz for the 6-channel vocoder; (ii) 100, 240, 440, 725, 1132, 1712, 2539, 3718, 5401, and 7800 Hz for the 9-channel vocoder; and (iii) 100, 178, 275, 394, 542, 725, 952, 1232, 1579, 2008, 2539, 3196, 4009, 5015, 6260, and 7800 Hz for the 15-channel vocoder. The envelope was extracted from each channel signal by half-wave rectification and low-pass filtering at a cutoff frequency of 45 Hz (initial filter slope=−18 dB/oct), to preserve envelope but not periodicity cues related to the speakers' fundamental frequency (Stone *et al.*, 2008). The extracted channel envelope was then compressed with a very high nominal CR of 50 to remove any change in amplitude for quasi-static input signals. The envelope compressor (Stone and Moore, 2003) used a two-pole Bessel-derived low-pass filter, producing near symmetric attack and release times to minimize distortion of the shape of the temporal envelope: values of 2449, 774, and 245 ms were chosen to affect selectively different ranges of low-rate AM.

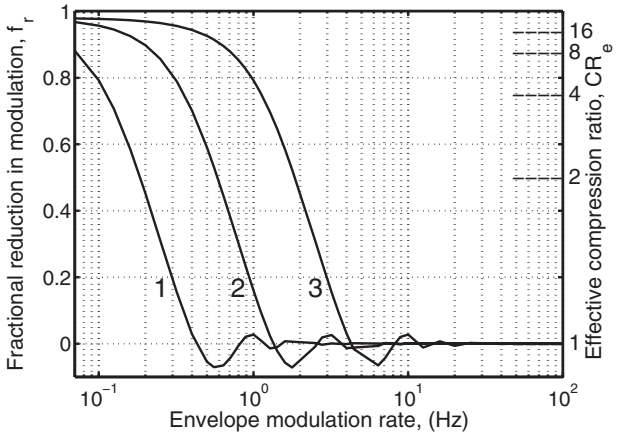


FIG. 1. Effect of different types of compressors on envelope modulation, expressed as the effective compression ratio (CR_e; right ordinate) and the fractional reduction in modulation (f_r; left ordinate) as a function of the input envelope modulation rate in Hz. Lines labeled "1" to "3" represent the effects of the three compressors used in the present study.

The performance of the compressor in terms of effective CR (CR_e; Stone and Moore, 1992, 2003) as a function of AM rate (for an input modulation depth of 10 dB) for the three pairs of attack and release times is shown by the functions labeled "1" to "3" in Fig. 1 (right ordinate). A more easily interpretable measure is the fractional reduction in modulation (f_r; left ordinate), which indicates the relative amount of modulation removed by the compressor (varying from 0 to 1) (Stone and Moore, 2003). The f_r plots show that compressors "1," "2," and "3" selectively reduced the modulation depth of AM with rates below approximately 0.4, 1.3, and 4 Hz, respectively. Although compressor 1 only reduced the modulation depth for very low AM rates, in some channels the gain changed by up to 4 dB during the course of a sentence, and this change occurred over a time of about 1 s. A low compression threshold of −15 dB relative to the channel root-mean-square value was used to ensure that the compressor was active during the presentation of the target speech. After compression, the channel envelope was used to modu-

late a white noise, which was then bandlimited by the same bandpass filter as used during the analysis stage. Finally, the N channel signals were combined. All signal processing was performed off-line. For stimulus presentation, the signals were generated via a high-quality sound card in a PC, passed through a mixing desk, and delivered diotically at 68 dB SPL through Sennheiser HD580 headphones. The listener was seated in a sound-attenuating chamber. For further details of the speech signals, the implementation of the noise-vocoder and compression systems, and the testing equipment, the reader is referred to experiment 2 in Stone and Moore (2008).

Since naive listeners initially show large learning effects in their ability to identify vocoded speech (e.g., Stone and Moore, 2003), the first session (90 min) was entirely dedicated to interactive training with feedback, using 168 target sentences presented using increasingly difficult conditions of channel number and TBR [for details, see Stone and Moore (2008)]. The second session (120 min), completed on average 2 days after the first, started with another 40 training sentences. During the subsequent experiment proper, identification performance (% words correct) was assessed for the nine experimental conditions (3 vocoders $\times$ 3 compression speeds). These were presented in a counterbalanced order across listeners (Latin-square design), using two ten-sentence lists per condition. To reduce the intra-subject variability due to list differences and potential learning effects, the measurement was repeated after a 15-min rest period, using a different Latin-square design and new sentence lists. After each sentence presentation, listeners were instructed to report verbally as many words as possible, and, if necessary, to guess. Scoring was by key words.

III. RESULTS

Within each Latin-square design, the time-ordered averaged scores were computed to check for possible training effects. No such effects were found. Mean identification scores for each experimental condition are shown in Fig. 2. The bottom and top axes show the compression speed and the equivalent highest affected modulation rate, respectively. Different symbols denote the three combinations of N and TBR.

The use of different TBRs for each N makes a direct comparison across N inappropriate, but was generally successful in yielding mid-range mean and individual performance, ranging from 34%–53% and 13%–87% correct, respectively. For all three combinations of N and TBR, scores increased as compression speed decreased from "3" to "1," that is, as the range of affected AM rates was narrowed from 0–4 Hz to 0–0.4 Hz. On average, performance improved by 8 and then 5 percentage points as the highest affected modulation rate was lowered from 4 to 1.3, and then to 0.4 Hz. Separate within-subject analyses of variance with factor "compression speed" were computed on the arcsine-transformed identification scores for each combination of N and TBR. These revealed a significant main effect [all $F(2,16) > 12.5$, all $p \leq 0.002$] and linear contrast [all $F(1,8) > 20.8$, all $p \leq 0.002$]. Subsequent *post hoc* pairwise

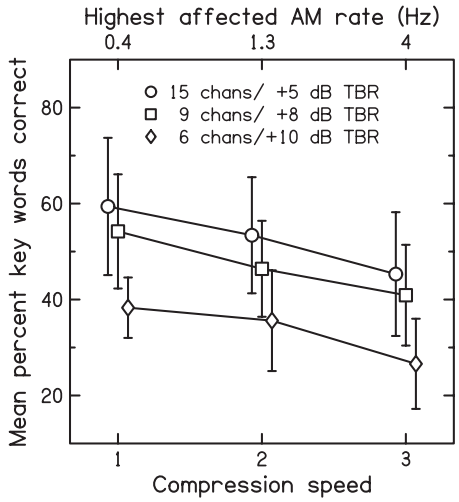


FIG. 2. Mean identification scores as a function of compression speed (bottom axis). The highest modulation rate (in Hz) affected by a given compression speed is indicated by the top axis. The number of processing channels, N, and the target-to-background ratio (TBR) are given in the key. Error bars show $\pm$ one standard deviation about the mean across listeners. Data points are horizontally displaced for better visibility.

comparisons (based on LSD tests) showed that, for each combination of N and TBR with one exception, identification improved significantly as the compressor speed was reduced (all $p < 0.02$); only the difference between scores for speeds "1" and "2" with the 6-channel vocoder did not reach significance ($p = 0.27$).

IV. SUMMARY AND DISCUSSION

We evaluated the contribution of AM at very low rates (<4 Hz) to speech intelligibility by using multi-channel compression with different compression speeds and a very high CR to effectively reduce the modulation depth of the AM at rates below a certain value. For the adverse listening condition simulated here (i.e., in the presence of an interfering talker, reduced spectral cues, and no temporal fine structure cues), it was found that

(1) independently of spectral resolution (ranging from 6 to 15 channels), envelope fluctuations in the range from 4 to 1.3 Hz contributed significantly to speech intelligibility;
(2) even envelope fluctuations between 1.3 and 0.4 Hz significantly aided speech intelligibility for the two higher spectral resolutions.

These results contrast with those of Stone and Moore (2008), who found that the reduction of AM depth for rates below 2 Hz had a significant effect on speech identification for an 8-channel, but not for a 12-channel, noise vocoder/compressor. Our finding of an effect for channel numbers of 9 and 15 may reflect the greater CR used in this study.

A possible reason why our results differ from previous results is that in some earlier studies (Drullman *et al.*, 1994b) the temporal fine structure was left intact. This can lead to the regeneration of (physically removed) temporal-envelope cues by the peripheral auditory system when the bandpass analysis channels are broad relative to the auditory filters.

However, Drullman *et al.* found similar results when using 1-oct-wide and 1/4-oct-wide channels, which suggests that regeneration of low-rate AM is not responsible for the difference between our results and theirs. Alternatively, the presence of temporal fine structure cues may have made low-rate envelope cues redundant; for example, listeners may have used temporal fine structure cues to perceptually segregate the target speech and background noise. In the present study, where no such cues were available, listeners may have partly relied on the low-rate AM, corresponding to the onset, offset, and general envelope shape of the utterance, to aid segregation. Such cues may be particularly useful when the onsets of the target and background are asynchronous, as in the present study. It may also be the case that low-rate AM cues are more important when the background is fluctuating (as in our study) than when it is steady (as in the studies of Drullman *et al.*). When compression is applied to a mixture of sounds, the envelopes of the sounds become partially correlated, due to the time-varying gain applied to the mixture; this effect has been called cross modulation (Stone and Moore, 2008). The cross-modulation effect produced by compression probably plays a greater role when speech is presented in a fluctuating background than when it is presented in a steady background.

To gain some insight into the possible role of difficulties in segregating the target and background, we analyzed the scores to determine the proportion of sentences that were identified completely correctly for each condition. One might argue that, as difficulties in segregation decrease, the proportion of such sentences would increase. Consistent with this idea, for 9 and 15 channels, the proportion of completely correct sentences decreased with increasing compression speed. However, when proportion correct scores were analyzed as a function of position within the sentences, there was no clear difference in the pattern of results across compression speeds, although for 6 channels, the proportion correct was lower for the first two words than for later words. Overall, these results suggest that at least part of the limitation in the ability to identify the target speech was produced by masking of the target speech rather than by the failure to segregate the target from the background.

Since we used vocoded rather than intact signals, the extent to which our results are applicable to intact speech perceived by normal-hearing listeners remains unclear. However, our results are directly relevant to listeners with moderate hearing impairment and cochlear implantees, since both groups are characterized by an inability to take advantage of the temporal fine structure of speech (Lorenzi *et al.*, 2006) and therefore rely mainly on temporal-envelope cues.

Finally, in future studies of the influence of low-rate AM on speech intelligibility, it might be prudent to avoid using brief materials [e.g., syllables or words (Arai *et al.*, 1999)]. A role for very low AM rates probably can only be demonstrated when speech materials with relatively long durations are used, such as the, on average, 2.2-s-long sentences of the present study. Indeed, measures of the ability to understand and recall entire passages might reveal an even larger contribution of low-rate AM to speech-in-speech identification.

ACKNOWLEDGMENTS

Author CF was supported by a Marie Curie Intra-European Fellowship and a Wolfson College Junior Research Fellowship. Authors MAS and BCJM were supported by the MRC, UK. We thank the three anonymous reviewers and the editor Dr. Freyman for helpful comments.

Arai, T., Pavel, M., Hermansky, H., and Avendano, C. (**1999**). "Syllable intelligibility for temporally filtered LPC cepstral trajectories," J. Acoust. Soc. Am. **105**, 2783–2791.

Drullman, R., Festen, J. M., and Plomp, R. (**1994a**). "Effect of temporal envelope smearing on speech reception," J. Acoust. Soc. Am. **95**, 1053–1064.

Drullman, R., Festen, J. M., and Plomp, R. (**1994b**). "Effect of reducing slow temporal modulations on speech reception," J. Acoust. Soc. Am. **95**, 2670–2680.

Glasberg, B. R., and Moore, B. C. J. (**1990**). "Derivation of auditory filter shapes from notched-noise data," Hear. Res. **47**, 103–138.

Houtgast, T., and Steeneken, H. J. M. (**1985**). "A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria," J. Acoust. Soc. Am. **77**, 1069–1077.

IEEE (**1969**). *Recommended Practice for Speech Quality Measurements* (Institute of Electrical and Electronic Engineers, New York).

Lorenzi, C., Gilbert, G., Carn, C., Garnier, S., and Moore, B. C. J. (**2006**). "Speech perception problems of the hearing impaired reflect inability to use temporal fine structure," Proc. Natl. Acad. Sci. U.S.A. **103**, 18866–18869.

Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). "Speech recognition with primarily temporal cues," Science **270**, 303–304.

Stone, M. A., and Moore, B. C. J. (**1992**). "Syllabic compression: Effective compression ratios for signals modulated at different rates," Br. J. Audiol. **26**, 351–361.

Stone, M. A., and Moore, B. C. J. (**2003**). "Effect of the speed of a single-channel dynamic range compressor on intelligibility in a competing speech task," J. Acoust. Soc. Am. **114**, 1023–1034.

Stone, M. A., and Moore, B. C. J. (**2008**). "Effects of spectro-temporal modulation changes produced by multi-channel compression on intelligibility in a competing-speech task," J. Acoust. Soc. Am. **123**, 1063–1076.

Stone, M. A., Füllgrabe, C., and Moore, B. C. J. (**2008**). "Benefit of high-rate envelope cues in vocoder processing: Effect of number of channels and spectral region," J. Acoust. Soc. Am. **124**, 2272–2282.

Wilson, B. S., Finley, C. C., Lawson, D. T., Wolford, R. D., Eddington, D. K., and Rabinowitz, W. M. (**1991**). "Better speech recognition with cochlear implants," Nature (London) **18**, 236–238.

A multi-rate decay model to predict energy-based acoustic parameters in churches (L)

Francesco Martellotta[a)]
Dipartimento di Architettura e Urbanistica, Politecnico di Bari, via Orabona 4, I70125 Bari, Italy

(Received 12 September 2008; revised 16 December 2008; accepted 16 December 2008)

Multi-rate decays are sometimes observed in room acoustics, appearing when markedly different volumes are coupled together and resulting in nonlinear decay curves. Such behavior appears in several churches at the very beginning of the decay process, although in conditions which cannot be explicitly referred to as coupling phenomena. Consequently, multi-rate exponential decays may be suitable to model energy distribution in this group of buildings, providing a more elegant and easily applicable set of equations in place of a previously defined "linear" model, used to adapt Barron's revised theory. The paper shows that the multi-rate approach ensures ease of calculation, without significant loss in accuracy in predicting energy-based acoustic parameters.
 [DOI: 10.1121/1.3075568]

PACS number(s): 43.55.Br, 43.55.Gx [NX] Pages: 1281–1284

I. INTRODUCTION

One of the most powerful instruments to predict acoustical parameters for the purposes of architectural acoustics is the "revised theory" proposed by Barron and Lee,[1] which, assuming that the reflected sound cannot arrive earlier than direct sound, makes both parameters decrease as a function of the distance. Unfortunately, several measurements carried out in a widespread group of worship places[2–5] show that reflected sound levels generally fall below those predicted by the revised theory. However, despite some fluctuations, the acoustic parameters under investigation are well related to source-receiver distance, allowing the development of prediction equations based on simple regression models, or leading to modifications of the revised theory in order to fit the measured data better.[2–5]

The model proposed by Cirillo and Martellotta[4] for Apulian-Romanesque churches and later generalized to a wider group of churches[5] assumes that the early reflected energy varies linearly within a time interval proportional to the source-receiver distance, after which the reverberant sound follows a purely exponential decay. As observed by Zamarreno *et al.*,[3] this "linear" assumption, despite the relatively good accuracy the model provides, results in a certain "roughness" of the formulation and in a calculation complexity mostly due to the variable length of the early-interval and to the different treatment required to integrate the early and late part of the curve.

This intrinsic limit may be tackled by means of a multi-rate exponential decay, which greatly simplifies the calculations. Multi-rate decays are not unusual in room acoustics, being mostly related to coupled-volume phenomena. However, using powerful Bayesian estimation tools,[6] multi-rate behavior can also be detected in the early part of sound decays measured in large reverberant churches, appearing in spaces where coupling effects may reasonably be excluded. An interesting aspect that can be frequently observed in churches (Fig. 1) is that as the distance from the source grows, the early part of the decay changes from a concave to a convex shape, resulting from the combination of the main decay process (with a time constant equal to reverberation time T) with a secondary exponential decay (with a time constant that is smaller than T). Transition from concave to convex shape corresponds to a change from positive to negative in the sign of the secondary decay.

According to the classical theory of coupled volumes,[7] such "negative" part always appears when the impulsive sound source and the receiver are in different subspaces, as the initial energy in the receiver room must be initially zero. However, in churches this phenomenon also appears when source and receiver are in the same subspace and becomes more evident (with some exceptions) when the source-receiver distance grows, resulting in an increase of the time constant of the secondary decay and in a decrease of its magnitude towards negative values. This seems to suggest that a single large volume behaves as the addition of several smaller volumes joined together through apertures, even though the latter are often large enough to ensure a high level of coupling, therefore contradicting the basic hypothesis of coupled-volume theory, which assumes weak coupling. More refined models of coupled spaces[8] might help to better explain such behavior, but the violation of one of the most important assumptions, together with the magnitude of the phenomenon, generally circumscribed to the very early part of the decay, suggests that the slower energy buildup is unlikely to depend on acoustic coupling (even though the latter might contribute to the effect in some circumstances). More likely, as already observed in Ref. 5, decorations, pillars, aisles, and side chapels scatter, hinder, or simply delay (after weakening) the early reflections, modifying both their energy and their time growth.

Even though a detailed explanation of the above experimental observations requires more detailed analysis, the results support the idea that a combination of exponentials might provide a better description of the sound propagation

[a)]Electronic mail: f.martellotta@poliba.it

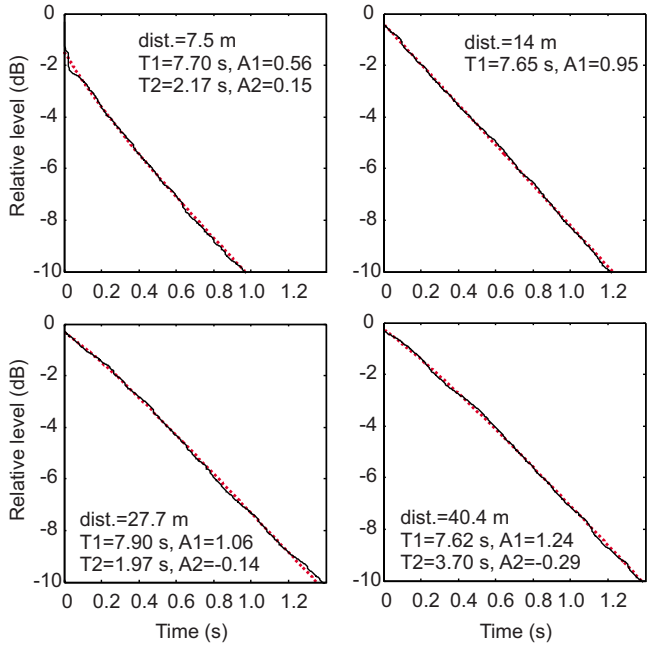


FIG. 1. (Color online) Early decay curves measured in the Basilica of San Lorenzo in Florence at different distances from the source. Dotted curve represents the Bayesian model calculated according to Ref. 6.

in churches. As further evidence, the paper aims at demonstrating that this approach fits very well to the model proposed by Cirillo and Martellotta,[5] allowing, at the same time, a more elegant mathematical formulation of the same model and greater ease in its application, without any loss in prediction accuracy.

II. OVERVIEW OF PREVIOUS MODELS

According to the classical theory of sound propagation in enclosed rooms, if the absorption is uniformly distributed and if the sound field is diffuse, the relative energy density of the reflected sound (expressed with reference to the energy density of an impulsive sound at a distance of 10 m) is a function of reverberation time T and room volume V:

$$g(t) = (13.8 \times 3\ 1200/V)e^{-13.8/T}\ (\mathrm{s}^{-1}). \tag{1}$$

Integration from time t_0 to infinity provides the following dimensionless equation:

$$i(t_0) = (31\ 200T/V)e^{-13.8t_0/T}, \tag{2}$$

allowing a simple and straightforward calculation of all the energy-based acoustic parameters. However, beyond the reverberation radius the classical theory predicts negligible variations as a function of distance compared to the actual measured values.

Barron and Lee[1] assumed that the direct sound is followed by linear level decay at a rate corresponding to the reverberation time. During a decay the instantaneous level of the late decaying sound is uniform throughout the space, so that decay traces for all receiver positions are superimposed. The time $t=0$ corresponds to the time the signal is emitted from the source, therefore the direct sound reaches a point at a distance r from the source after a time $t_D=r/c$ (c being sound velocity in air). In this way the integrated energy decreases when the source-receiver distance increases, while the early/late reflected energy ratio remains constant. The integrated value for the reflected sound level is assumed to be, at $t=0$, equal to the value predicted by the classical theory.

In order to predict the total sound-pressure level and the clarity index, the sound energy is divided into three components:[1] the direct sound (d), the early reflected sound (from 0 to 80 ms, E_0^{80}), and the late reflected sound (from 80 ms to infinity, E_{80}^{∞}). From Eq. (2), the corresponding energies of each of them become

$$d(r) = 100/r^2, \tag{3}$$

$$E_0^{80}(r) = (31\ 200T/V)e^{-0.04r/T}(1 - e^{-1.11/T}), \tag{4}$$

$$E_{80}^{\infty}(r) = (31\ 200T/V)e^{-0.04r/T}e^{-1.11/T}. \tag{5}$$

So the relative sound level is given by

$$G(r) = 10\log(d + E_0^{80} + E_{80}^{\infty}), \tag{6}$$

and the clarity index is given by

$$C_{80}(r) = 10\log[(d + E_0^{80})/E_{80}^{\infty}]. \tag{7}$$

The center time is the first order momentum of the acoustic energy and, taking into account that for a purely exponential decay the distance of the center of gravity from the origin (i.e. the starting point of the decay t_D) is independent of r and is equal to $T/13.8$, it is given by

$$T_s(r) = T(E_0^{80} + E_{80}^{\infty})/[13.8(d + E_0^{80} + E_{80}^{\infty})]. \tag{8}$$

The analysis of the results initially observed in a group of Romanesque churches[4] and then confirmed by measurements in a larger set of Italian churches[5] showed that the basic hypothesis of the revised theory, namely the uniformity of the reverberant sound field throughout the space, was generally satisfied. However, the time at which the decay began to be linear was later, the farther the measurement position was from the source. Furthermore, at points near the source the early reflections were stronger than the ideal classical reverberant field [as given by Eq. (2)], while, conversely, when the distance from the source grew, the early reflections became weaker (Fig. 1).

In order to fit with these observations two modifications were introduced. The first one was to assume the reverberant sound field to be uniform, as it is in Barron's theory, but that linear level decay starts with a certain delay (t_R) after the arrival of the direct sound. The measurements showed that this delay was proportional to the source-receiver distance, therefore, in general, it could be written as $t_R=\rho r$ or, taking into account arrival time of direct sound, as $t_R=kt_D$, where $k=\rho c$. The k coefficient (as well as its distance equivalent ρ) depended on the room characteristics and it was demonstrated that it could be expressed as a function of architectural features of each church, assuming integer values from 1 to 3 growing with church complexity (Table III in Ref. 5).

The second modification was to schematize the early reflected sound arriving between the direct sound and the reverberant sound field as a continuous linear function varying from an initial value (at the time t_D), proportional

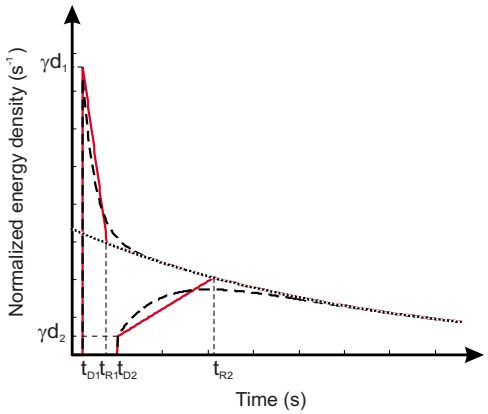


FIG. 2. (Color online) Shapes of the energy decay curve obtained with the linear model (—) and with the multi-rate exponential model (---) observed at different distances from the source. Dotted curve represents classical exponential decay.

through a factor γ to the energy of the direct sound, and a final value (at time t_D+t_R), equal to the energy of the reverberant field at the same time (Fig. 2). The factor γ was assumed to be equal to $(1-\alpha)(1-s)/\Delta\tau$, where α and s were, respectively, the mean absorption coefficient and the mean scattering coefficient of the room surfaces, while $\Delta\tau$ was the time for sound to travel the mean free path (MFP), equal to $4V/cS$, where S is the total surface area. In this way factor γ transformed the energy of the direct sound into the energy of the first specular reflection distributed over time $\Delta\tau$. The estimation of the mean scattering coefficient was simplified by assigning values varying from 0.2 to 0.8 as a function of the mean characteristics of the area surrounding the sound source (Table IV in Ref. 5).

III. THE MULTI-RATE DECAY MODEL

A multi-rate decay process is the result of a linear combination of decays having different time constants and different weights,[7] so its general equation may be written as

$$g'(t) = A_1 e^{-13.8t/T_1} + A_2 e^{-13.8t/T_2} + \cdots + A_n e^{-13.8t/T_n}, \quad (9)$$

where T_i are the different reverberation times (which in coupled volume problems depend on the characteristics of subvolumes), and A_i are the linear parameters which depend both on the characteristics of subspaces and on the initial conditions of the problem (i.e., source and receiver positions).

For the purposes of this paper a simple two-rate model is investigated. The first decay coincides with the classical exponential decay as revised by Barron, so that $T_1=T$ and $A_1(r)=(13.8\times 31\,200/V)e^{-0.04r/T}$. The second needs to be adapted in order to fit the "linear" part of the previous model, so the decay constant T_2 must be proportional to t_R in order to have a slower buildup as the distance grows (while remaining much smaller than T). The choice of the proportionality factor between T_2 and t_R is crucial in order to obtain the best agreement between the energy of the early and late reflections calculated with the linear function and with the exponential function. The best results are obtained by using $T_2(r)=6.9t_R=6.9\rho r$, which is a convenient value because the center of gravity of the second exponential function falls in the middle of the t_R interval. The linear parameter A_2 must be defined in order to let g' have an initial value equal to γd. Therefore,

$$A_2(r) = \gamma d - A_1(r) = 100\gamma/r^2 - (13.8 \times 31\,200/V)e^{-0.04r/T} \quad (10)$$

and may assume negative values (as observed in Fig. 1), with a limit equal to $-(13.8\times 31\,200/V)e^{-0.04r/T}$.

Taking the origin of times at the direct sound arrival and expressing all the parameters, except A_2, as a function of r, the general equation may be written as

$$g'(r,t) = (13.8 \times 31\,200/V)e^{-0.04r/T}e^{-13.8t/T} + A_2 e^{-2t/\rho r}, \quad (11)$$

and Eq. (3) may be rewritten as

$$i'(t,r) = (31\,200T/V)e^{-0.04r/T}e^{-13.8t/T} + (A_2\rho r/2)e^{-2t/\rho r}. \quad (12)$$

Consequently, Eqs. (4) and (5) may be rewritten as follows:

$$E_0'^{80}(r) = E_0^{80} + (A_2\rho r/2)(1 - e^{-0.16/\rho r}), \quad (13)$$

$$E_{80}'^{\infty}(r) = E_{80}^{\infty} + (A_2\rho r/2)e^{-0.16/\rho r}. \quad (14)$$

In this way G and C_{80} may be calculated by first determining A_2 according to Eq. (10), and then replacing E with E' in Eqs. (6) and (7), while T_s needs to be calculated with the following equation in order to account for the different position of the centers of gravity of the two exponential functions:

$$T_s(r) = [T \cdot E_0^{\infty}/13.8 + (\rho r/2)(E_0'^{\infty} - E_0^{\infty})]/(d + E_0'^{\infty}), \quad (15)$$

where $E_0'^{\infty}=E_0'^{80}+E_{80}'^{\infty}$ and $E_0^{\infty}=E_0^{80}+E_{80}^{\infty}$.

The main drawback of the old formulation[5] was the use of two different functions for early and late energy, defined on a time interval (t_R) which varied as a function of the distance. Consequently, the equations to calculate the integrated energy changed according to the relative value of t_R compared with the time integration limit. The new approach combines two exponential functions defined on the same time interval, providing a fixed set of equations for a given integration limit and, thanks to their similarity with Barron's original formulation, allows a simpler and straightforward application.

Replacing the linear function with a double-rate decay obviously changes the energy distribution as a function of time. Figure 2 compares the new shape of the energy curve with the old one, showing that the multi-rate model underestimates the early energy at points close to the source and overestimates it at farther points. At later time, approximately after time t_R, the energy follows an opposite behavior. This is better explained in Fig. 3, showing that the total reflected energy is practically the same for both models. The early sound (with reference to the conventional 80 ms interval) appears the same at points close to the source, but at farther points the exponential model gives higher values than the linear one. The late sound shows an opposite behavior but the difference is generally smaller as a consequence of

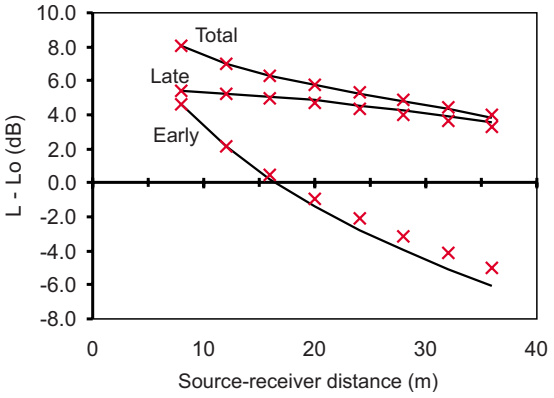


FIG. 3. (Color online) Plot of early, late, and total reflected energy (time limit 80 ms, and no direct sound) calculated with linear model (—) and with the multi-rate exponential model (×) as a function of distance from the source. Parameters used in the model were $V=25\,000$ m^3, $T=4$ s, $S=8000$ m^2, $k=2$, $s=0.4$.

the compensation appearing between parts above and below the energy predicted by the linear model. Consequently, the new model is expected to predict strength with the same accuracy of the linear model, while it is likely to overestimate clarity and slightly underestimate center time (especially at farthest points). However, the effect of these differences in terms of prediction accuracy needs to be investigated by comparison with measured values.

Taking into account exactly the same sample of churches used to generalize the original model (see Ref. 5 for plans, architectural description, and choice of k and s coefficients), the multi-rate revision was applied, using the same parameters, in order to validate its predictive accuracy. Then, for each church, the values of G, C_{80}, and T_s were calculated for every source-receiver combination using both the modified and the multi-rate theory. The results of the calculations were then compared with the data measured in each church by determining the mean rms error between measured and predicted values for each source-receiver combination.

Table I summarizes the results of the comparisons showing, as expected, a slight loss of accuracy for clarity (of about 0.1 dB), while both G and T_s show the same average performance. Taking into account individual results, the new model shows the highest loss of accuracy for C_{80} in the Basilica of St. Nicholas in Bari and in the Gesù in Rome, with rms errors respectively increased by 0.4 and 0.7 dB. However, in the latter church the old model also gave an high rms error (2.5 dB), suggesting a substantial lack of early reflections (and possibly of direct sound) due to the grazing incidence above the pews distributed along the nave, which cannot be accounted for by both models. Improved C_{80} prediction was observed in those churches, such as San Lorenzo and the Concattedrale, where the pattern of the reflections arriving within the first 80 ms is richer. T_s shows the best improvements for Lucera and the Gesù, where the measured values become more stable at farthest points, reaching a sort of plateau, possibly because of reflections from the back walls. Finally, it can be observed that in some cases an increase in the error on C_{80} corresponds to a decrease on T_s, suggesting that the actual reflected energy distribution probably lies in between the two models.

TABLE I. Summary of rms errors between measured and theoretical values of sound strength, clarity, and center-time at 1 kHz frequency band. Theoretical values are calculated according to the old "linear" model and the new "exponential" model.

	G (dB)		C_{80} (dB)		T_s (ms)	
Church	lin	exp	lin	exp	lin	exp
S. Sabina Basilica, Rome	0.7	0.7	1.4	1.5	21	22
St. Apoll. in Cl., Ravenna	0.4	0.4	1.0	1.1	14	15
Modena Cathedral	0.7	0.7	1.2	1.5	31	31
St. Nicholas, Bari	0.4	0.4	1.1	1.5	24	27
St. Petronius, Bologna	0.8	0.8	2.5	2.7	27	28
Lucera Cathedral	0.4	0.5	1.3	1.7	31	25
San Lorenzo, Florence	0.5	0.6	1.3	1.1	25	23
Gesù, Rome	0.7	0.6	2.5	3.2	37	32
St. Luca & Martina, Rome	0.5	0.5	0.9	1.1	18	17
Martina Franca	0.6	0.7	1.2	1.4	27	26
Concattedrale, Taranto	0.5	0.5	1.1	0.7	14	18
Riola	0.4	0.4	0.8	0.9	23	23
Mean	*0.6*	*0.6*	*1.4*	*1.5*	*24*	*24*

IV. CONCLUSIONS

The use of a multi-rate decay model to adapt Barron's revised theory to churches was investigated as a refinement of a previous linear model. Prediction formulas proved to be considerably simpler without affecting the theoretical background (so that all the parameters previously defined as a function of the architectural characteristics of the church may still be used). On average, prediction accuracy remained unchanged, with only a small increase of 0.1 dB in the rms error calculated for C_{80}. Further work is needed to investigate the possibility to extend the model to other types of spaces which present similar features and to validate its accuracy in predicting early decay time variations as a function of distance.

[1]M. Barron and L. J. Lee, "Energy relations in concert auditoriums. I," J. Acoust. Soc. Am. **84**(2), 618–628 (1988).
[2]J. J. Sendra, T. Zamarreño, and J. Navarro, "Acoustics in churches," in *Computational Acoustics in Architecture*, edited by J. J. Sendra (Computational Mechanics, Southampton, 1999), pp. 133–177.
[3]T. Zamarreño, S. Giròn, and M. Galindo, "Acoustic energy relations in Mudejar-Gothic churches," J. Acoust. Soc. Am. **121**(1), 234–250 (2006).
[4]E. Cirillo and F. Martellotta, "An improved model to predict energy-based acoustic parameters in Apulian-Romanesque churches," Appl. Acoust. **64**, 1–23 (2003).
[5]E. Cirillo and F. Martellotta, "Sound propagation and energy relations in churches," J. Acoust. Soc. Am. **118**(1), 232–248 (2005).
[6]N. Xiang and T. Jasa, "Evaluation of decay times in coupled spaces: An efficient search algorithm within the Bayesian framework," J. Acoust. Soc. Am. **120**(6), 3744–3749 (2006).
[7]L. Cremer, and H. A. Muller, *Principles and Applications of Room Acoustics, Volume 1* (Applied Science, London, 1982).
[8]J. E. Summers, R. R. Torres, and Y. Shimizu, "Statistical-acoustics models of energy decay in systems of coupled rooms and their relation to geometrical acoustics," J. Acoust. Soc. Am. **116**(2), 958–969 (2004).

An investigation of Rubens flame tube resonances[a)]

Michael D. Gardner[b)] and Kent L. Gee
Department of Physics and Astronomy, Brigham Young University, N283 Eyring Science Center, Provo, Utah 84602

Gordon Dix
JBL Professional, 8500 Balboa Boulevard, Northridge, California 91329

(Received 29 July 2008; revised 12 December 2008; accepted 30 December 2008)

The Rubens flame tube is a century-old teaching demonstration that allows observers to visualize acoustic standing wave behavior [H. Rubens and O. Krigar-Menzel, (1905). Ann. Phys. **17**, 149–164]. Flammable gas inside the tube flows through holes drilled along the top, and flames are then lit above. The tube is closed at one end and driven with a loudspeaker at the other end. When the tube is driven at one of its resonance frequencies, flames form a visual standing wave pattern as they vary in height according to the pressure amplitude in the tube. Although the basic performance of the tube has been explained [G. Ficken and C. Stephenson, (1979). Phys. Teach. **17**, 306–310], this paper discusses a previously unreported characteristic of the tube: a shift of the tube's resonance frequencies away from those predicted by simple introductory physics. Results from an equivalent circuit model of the tube and agreement between experiments and the model suggest that the shift is caused by the presence of the holes. For teachers and educators seeking to better understand and explain the tube to students, this article serves as a resource regarding the basic phenomena affecting the behavior of the tube. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3075608]

PACS number(s): 43.10.Sv, 43.20.Ks [VWS] Pages: 1285–1292

I. INTRODUCTION

A. History

In 1905, German physicists Rubens and Krigar-Menzel[1] discovered a way to demonstrate acoustic standing waves visually using what they referred to as a "flame tube." One hundred 2-mm diameter holes were drilled across the top of a round brass tube that was 4 m long and 8 cm in diameter. The tube was filled with coal gas and then flames were lit from the gas exiting through the holes on top. The tube was closed at both ends and driven at one of the ends with a tuning fork in a box. At resonances of the tube, the standing wave was seen in the flames above the tube with the flame height correlating with the pressure amplitude inside the tube.

B. Classroom demonstration

Because the Rubens flame tube provides an exciting visual representation of sound waves, it naturally serves well as a teaching demonstration in the classroom setting of introductory physics or acoustics. When teaching about sound waves, it is common to talk about harmonically related resonances in pipes. The Rubens tube is suitable for a discussion of resonances because the resonances are easily seen: The flame height variation increases dramatically as resonance is reached. Because the tube was developed as a visual demonstration of a simple behavior, it has been used to foster student learning of basic physics principles. For example, resonance-induced patterns in the flame have been used to explore basic relationships between frequency, wavelength, and sound speed. Via examples on the Internet, we are aware of students using the flame tube to deduce the speed of sound inside the pipe by measuring the distance between two peaks in the flame pattern and assuming that to correspond to a half-wavelength (e.g., see Fig. 1). With information regarding the wavelength and the driving frequency, a student then calculates the sound speed.

C. Previous research

Although this demonstration is more than 100 years old, few studies have been published on the behavior of the Rubens flame tube. This is likely due to the fact that the tube is meant to demonstrate relatively simple physics; consequently, few have taken the time to study its behavior in detail. The most notable exception is Ficken and Stephenson,[2] who drove their flame tube with a directly coupled loudspeaker and showed that flame maxima occur at pressure nodes in the tube and flame minima at pressure antinodes. They explained this result using Bernoulli's equation, which indicates that the time-averaged mass flow rate of the gas is greatest at the pressure nodes. However, they showed that for low static gas pressures inside the tube or high acoustic amplitudes, the effect reverses, such that the flame minima occur at the pressure nodes and the flame maxima occur at the pressure antinodes. This phenomenon has not been fully explained and may be the subject of future study.

In addition to the work of Ficken and Stephenson,[2] other short studies have been performed on aspects of the tube's

[a)]Portions of this work were presented at the 153rd meeting of the Acoustical Society of America in Salt Lake City, UT.
[b)]Electronic mail: gardnerfrance@gmail.com

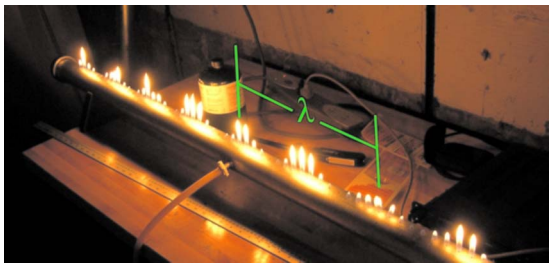


FIG. 1. (Color online) Hypothetical measurement of a wavelength in the tube using flame peaks produced by the standing wave pattern.

behavior. For example, Jihui and Wang[3] discussed the relationship between flame height and pressure in the tube, while Spagna[4] researched the behavior of the flame itself. He tried to determine the phase relationship of the flame's flicker relative to the loudspeaker response, but found that his results were somewhat inconclusive. A direct extension of Rubens and Krigar-Menzel's work[1] is that of Daw,[5,6] who published articles detailing the construction and performance of square and circular flame tables that are used for visualizing two-dimensional modal patterns.

D. Motivation

Although prior studies have addressed some aspects of the flame tube's performance, none of the investigations have discussed the relationship between tube resonance frequencies. Why might the neglect of this point be significant? Perhaps the answer is best illustrated by a direct example. During presentation of the seemingly simple physics of the tube in a department seminar at Brigham Young University, it was noted by one of the authors that the resonance frequencies of the flame tube used in the demonstration were not harmonically related as was expected from a simple explanation of the tube physics. Rather, the frequencies for the lower modes appeared to be shifted upwards and the cause for this shift was not obvious to him or to those in attendance. Anecdotally, in speaking with teachers at other institutions, we have discovered that others have observed similar phenomena in demonstrating the flame tube in their classes. Because the Rubens flame tube has been intended as a teaching tool to engage student interest in introductory classes, the shift in resonance frequency partially negates its effectiveness.

The results of the study presented in the remainder of this paper indicate that the modal frequency shift is due to the presence of the holes themselves, despite their small size relative to the dimensions of the tube. The holes create a Helmholtz-type resonance in the response of the tube that then causes the modal frequencies to be shifted upward. A Helmholtz resonator consists of a volume of gas in the bulk of a container acting as an acoustic compliance and a mass of air in the neck of the container acting as an acoustic mass. In the case of the flame tube, the volume of gas or acoustic compliance is the tube interior and the acoustic masses of air in each of the drilled holes create many Helmholtz resonators in parallel along the tube. These resonators notably affect the behavior of the tube as is shown through modeling and measuring tube behavior.

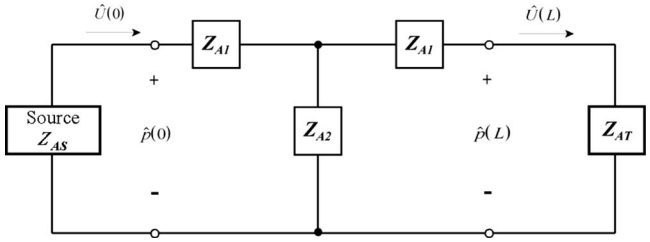


FIG. 2. T-network with source and termination (Z_{AT}).

II. METHODS

A. Tube modeling

In order to better understand the tube and to quickly test parameter changes, an equivalent circuit model of the tube was developed as a direct adaptation of a similar model previously developed by Dix[7] for different purposes. Modeling of tubes with holes has been done before, notably by Plitnik and Strong[8] and Keefe,[9] in the application of musical instrument modeling. A summary of the equivalent circuit theory, based on Keefe's[9] development, used in modeling the tube follows.

For lumped-element systems (i.e., systems where all dimensions are small compared to wavelength), equivalent circuits can be used to calculate quantities such as volume velocity and pressure inside the tube as a function of frequency. In our case, we have chosen an impedance analog, where the acoustic pressure ($\hat{p}$) corresponds to voltage in the circuit, and the volume velocity ($\hat{U}$) corresponds to current. Because the tube's length is much greater than the cross-sectional dimensions and greater than some of the wavelengths of interest, a waveguide circuit is used to account for changing acoustic parameters along the longer dimension. Waveguide circuits translate impedances from a termination to an input location as[10]

$$Z_{\mathrm{AI}} = \left(\frac{\rho_0 c}{S}\right)\frac{Z_{\mathrm{AT}} + j(\rho_0 c/S)\tan(kL)}{\rho_0 c/S + jZ_{\mathrm{AT}}\tan(kL)}, \tag{1}$$

where Z_{AI} is the desired acoustic impedance at the input location, c is the speed of sound, S is the cross-sectional area of interest, k is the wave number, ρ_0 is the fluid density, and Z_{AT} is the acoustic termination impedance to be translated a distance L down the tube. A waveguide circuit can also account for variable conditions along the length of the waveguide, or in our case, the presence of holes along the top.

In a waveguide circuit, two terms correspond to an arbitrary source and an arbitrary termination. The three other impedances make up the "T-network" in Fig. 2.

By equating the input impedance of this circuit to the impedance "translation" theorem in Eq. (1), one finds the series impedance terms Z_{A1} (acoustic impedance) to be

$$Z_{A1} = j\left(\frac{\rho_0 c}{S}\right)\tan\left(\frac{kL}{2}\right) \tag{2}$$

and the shunt impedance term Z_{A2} to be

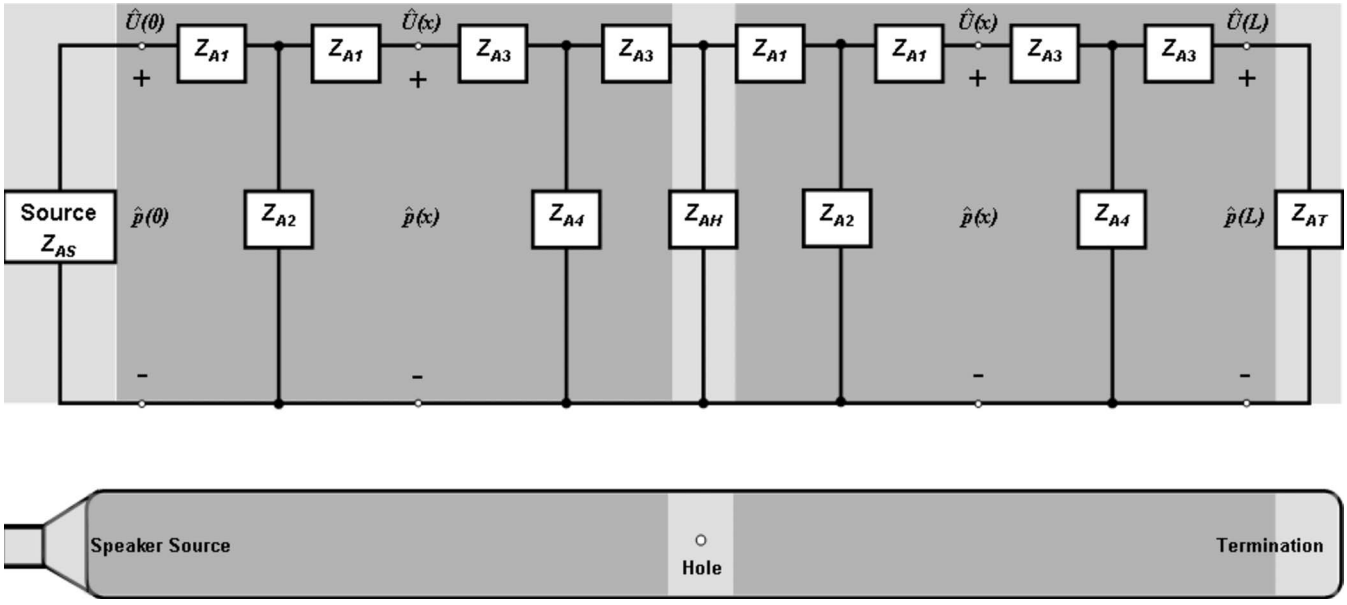


FIG. 3. Waveguide circuit diagram for a simple tube with source, termination, and one hole (Z_{AH}) with appropriate shading showing the impedance elements' corresponding physical objects.

$$Z_{A2} = -j\left(\frac{\rho_0 c}{S}\right)\csc(kL). \tag{3}$$

This circuit, comprised of Fig. 2 and Eqs. (2) and (3), describes pressure and volume velocity at the source and at the termination but not in between. Although the circuit serves as the foundation for the rest of the flame tube model, additions are needed to calculate the pressure along the length of the tube and to incorporate the presence of the holes.

To acquire acoustic quantities at points between the source and termination, a modification of the waveguide circuit is required. By coupling two T-network circuits together, the pressure or volume velocity can be obtained for any position x by using the voltage (pressure) drop at the junction of the circuits. The impedances Z_{A1} and Z_{A2} remain the same, but now there are two more impedance quantities to solve for in the second T-network, namely,

$$Z_{A3} = j\left(\frac{\rho_0 c}{S}\right)\tan\left[\frac{k(L-x)}{2}\right] \tag{4}$$

and

$$Z_{A4} = -j\left(\frac{\rho_0 c}{S}\right)\csc[k(L-x)]. \tag{5}$$

By allowing x, which in this case represents the distance from the source end of the tube, to vary in Eqs. (4) and (5), the pressure or volume velocity may be calculated along the length of the tube.

To incorporate the holes at the top of the tube, only one major change is needed in the model. These holes represent a change in impedance that can be accounted for by taking multiple T-networks and juxtaposing them together with shunt terms for each of the holes. The sound pressure and volume velocity can then be modeled at the source, the termination, or any of the holes. To calculate the pressure or volume velocity at any point along the tube, two T-networks are coupled together between each hole impedance, source, or termination impedance, schematically shown for one hole in Fig. 3. Note that in Eqs. (4) and (5) the definition of x is changed to mean either the distance from the source or from the hole immediately to the left, whichever is closer.

The equivalent circuit theory is used to make a code that calculates the pressure along the length of the tube across a span of frequencies. The internal impedance of the source (the loudspeaker) is included and the volume velocity at the face is obtained using the Thiele–Small parameters. The cavity, a series of 2 cm gaps of varying diameter, is also included. The fluid properties are included as well, i.e., propane in the tube and air in the source cavity and loudspeaker enclosure. The frequencies at which the pressure amplitude is a local maximum at the termination (a rigid cap) are considered resonance frequencies. We numerically calculate the pressure at every point along the tube at all the frequencies of interest by multiplying the impedance and the volume velocity together as

$$\hat{p}(x) = \hat{U}(x)Z(x). \tag{6}$$

B. Experiment setup

To study the impact of the holes on the shift in tube resonance frequencies and to compare against the equivalent circuit model, we constructed two flame tubes. The two galvanized-steel tubes are 1.524 m long with a 2.6 cm radius and have a 4 mm wall thickness. Both tubes have 60 deburred holes drilled in the top, each 2.2 cm apart, that begin 12 cm from the source end of the tube. The two tubes differ only in the size of the holes for the flames; one tube has 0.92 mm radius holes and the other has 0.46 mm radius holes, or half the size of the larger holes. Drilled in the center-side of each tube are 9.5 mm holes each with a short section of pipe used for the gas intake. A small hole was drilled in the

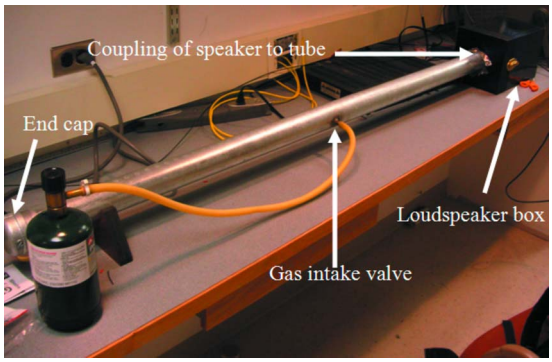


FIG. 4. (Color online) Flame tube setup.

galvanized-steel end cap (termination) for the tube to accommodate the 6.35-mm (0.25-in.) GRAS Type 1 microphone used to monitor the acoustic pressures at the termination end of the tube. The entire flame tube setup is displayed in Fig. 4.

To drive the tube, an enclosed loudspeaker was coupled directly to the tube at the other end as shown in Fig. 5. A 10.5 cm diameter driver, whose measured Thiele–Small parameters are provided in Table I, was placed in a medium-density fiberboard enclosure sealed with putty. The driver radiates into the tube through a 2 cm deep cavity that is 15.3 cm in diameter. This 15.3 cm section is the largest cross-sectional dimension in the setup, which limits our one-dimensional circuit model to 1018 Hz in propane. Although not visible in Fig. 5, plastic wrap approximately 15 μm thick was inserted at the drive end of the tube to prevent propane from leaking into the loudspeaker enclosure from the tube. Figure 6 shows the large-holed flame tube operating, the mode with four pressure antinodes and three pressure nodes being clearly shown.

In order to compare the model and physical flame tube results, acoustical measurements are needed inside the tube. This is accomplished by inserting a microphone in the termination end of the tube. Because of the rigidity of the galvanized-steel end cap, pressure antinodes occur at this location at resonance. With white noise driving the loudspeaker, the frequency response is measured at the termination. This technique was used to measure the frequency response for both the small- and large-holed tubes with the

FIG. 5. (Color online) Close-up of the coupling of the speaker to the tube with relevant dimensions.

TABLE I. List of Thiele–Small parameters for the loudspeaker (Ref. 7).

Parameter	Value	Units
R_e	5.42	Ω
L_1	0.152	mH
L_2	0.488	mH
R_2	16.5	Ω
Q_{MS}	3.03	⋯
C_{MS}	482.8	μm/N
M_{MS}	9.66	g
F_s	73.7	Hz
Bl	6.59	Tm

flames lit. These measured frequency responses become the benchmark for comparison between the modeled and measured tubes.

III. RESULTS

A. Equivalent circuit model

The first results to discuss are those of the modeled tube. In order to quickly change model parameters and easily visualize the results, a MATLAB®-based graphical user interface was developed. A user inputs into the model the relevant tube dimensions (tube radius, hole spacing, number, size, etc.), the speed of sound inside the tube, the density of the gas, and the relevant source model parameters. The model then outputs a graph of the magnitude of the pressure along the tube at whatever frequencies are chosen. Example results can be seen in Fig. 7 for the large-holed tube. The figure shows relative sound pressure level in the tube (white indicates greater level) as a function of frequency and position in the tube. Because propane is denser than air, we assume that the tube eventually completely fills with propane as the air is forced out through the holes. Based on the average of temperature measurements inside the tube during operation, the modeled propane was calculated to have a sound speed of 256 m/s and a density of 1.62 $\mathrm{kg/m^3}$.

Figure 7 shows the sound pressure level variation inside the large-holed tube for a number of scenarios and it is worthwhile to discuss those in some detail. First, although Figs. 7(a) and 7(c) are for the tube modeled with a constant volume velocity source, rather than the loudspeaker, all results shown include the loudspeaker cavity for the sake of consistency. In Fig. 7(a), which neglects both the holes and the loudspeaker, the resonance frequencies (white lines) are harmonically related and have a high quality factor. Figure 7(b) extends the results in Fig. 7(a) by introducing the effects

FIG. 6. (Color online) Flame tube in operation.

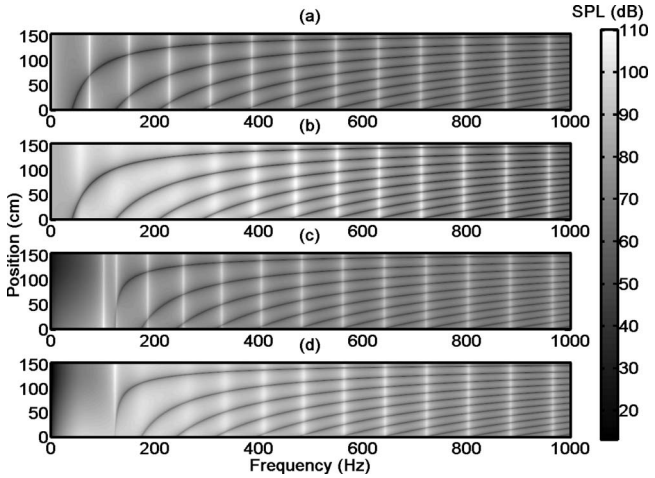


FIG. 7. Graph of sound pressure level inside large-holed tube from the speaker end (0 cm) to the cap end (150 cm) looking at frequencies from 0 to 1000 Hz. (a) The graph is modeled *without* the source or holes but *with* the cavity. (b) *Without* the holes but *with* the source cavity and speaker. (c) *With* holes and cavity *without* the speaker. (d) *With* holes, cavity, and speaker.

of the speaker. The speaker alters the resonance frequencies and greatly decreases the quality factor of the lowest several modes. This is expected because the speaker has its own resonance at 73.7 Hz and introduces significant damping to the system. Figure 7(c) includes the effects of the holes with an ideal volume velocity source. Here, there are two effects to consider. First, we observe an additional resonance of the tube where the pressure level does not vary spatially. This is the Helmholtz resonance of the tube. Second, the natural frequencies of the first several modes are significantly shifted upward in frequency from the no-holed case in Fig. 7(a).

By including the loudspeaker and the holes in Fig. 7(d), the model comes closest to matching the actual experiment. Here, the presence of the coupled loudspeaker and the holes both reduces the quality factor of the lowest resonances and significantly shifts the frequencies of first several modes upward. It is clear from comparing Figs. 7(c) and 7(d) that the dominant cause of the overall resonance frequency shift above the first few low-quality-factor modes is due to the presence of the holes, not the loudspeaker. It is also worth noting that the interaction between the holes and the loudspeaker in Fig. 7(d) actually removes the Helmholtz mode from the response.

Before moving to the comparison of the modeled and physical flame tubes, a further illustration of the effect of the holes is worthwhile. To remove the effect of the real loudspeaker, we assume that our small- and large-holed tubes are driven with a constant volume velocity source (again, keeping the source cavity that is present in the physical tube). The resonance frequencies for the first nine modes are shown in Table II, along with the resonance frequencies for the tube without holes and the percent shift upward predicted. A comparison of the last two columns of Table II demonstrates that the resonance frequencies of the large-holed flame tube are significantly more affected by the holes than the small-holed tube. Using the large-holed tube as an example, examination of the spacing of the resonance frequencies themselves illustrates the possible difficulty encountered in trying to illustrate "simple physics" in a classroom setting. For the approximation that the tube is a simple closed-closed pipe, the resonance frequencies of the driven tube should be harmonically related. The average spacing of the resonance frequencies of the large-holed pipe is around 75 Hz, but the frequencies themselves are not multiples of 75 Hz and Fig. 7(c) reveals that there is no predicted resonance behavior below 100 Hz.

B. Comparison of predicted and observed responses

The results in Table II were meant to examine the impact of the holes on the tube by themselves. The real system, however, will be impacted by the presence of the loudspeaker, as was modeled in Fig. 7(d). To compare the observed to the predicted modal frequencies, the frequency response magnitude at the termination position was plotted in decibels for both the model and as measured by the microphone at the end cap of the physical tube. The resonance frequency results for the large-holed tube with propane are displayed in Fig. 8. Again, the emphasis is on the frequencies, so the observed and predicted responses have been deliberately offset from each other. Although the agreement between the relative amplitudes for the first several modes is

TABLE II. Model-predicted resonance frequencies in propane for the three cases of the tube *with* the cavity: small holes, large holes, and no holes, including the percent shift from no holes. The tubes are driven by ideal constant volume velocity sources.

Mode No.	Small hole frequency (Hz) Helmholtz res. 60 Hz	Large hole frequency (Hz) Helmholtz res. 102 Hz	No hole frequency (Hz)	Small hole % shift	Large hole % shift
1	92	127	75	22.7	69.3
2	162	187	151	7.28	23.8
3	236	255	228	3.51	11.8
4	314	329	307	2.28	7.17
5	392	405	387	1.29	4.65
6	472	483	467	1.07	3.43
7	552	562	548	0.73	2.55
8	643	641	630	0.48	1.75
9	726	722	711	0.42	1.55

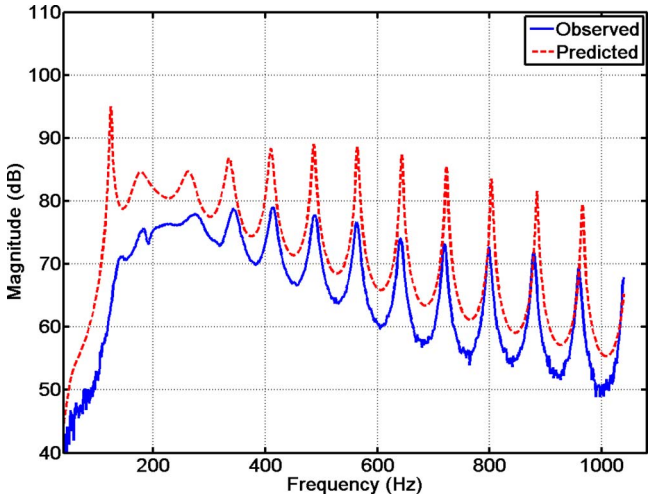


FIG. 8. (Color online) Frequency response at the cap end of tube comparing the observed response (tube in operation) to the model predicted response in large-holed tube.

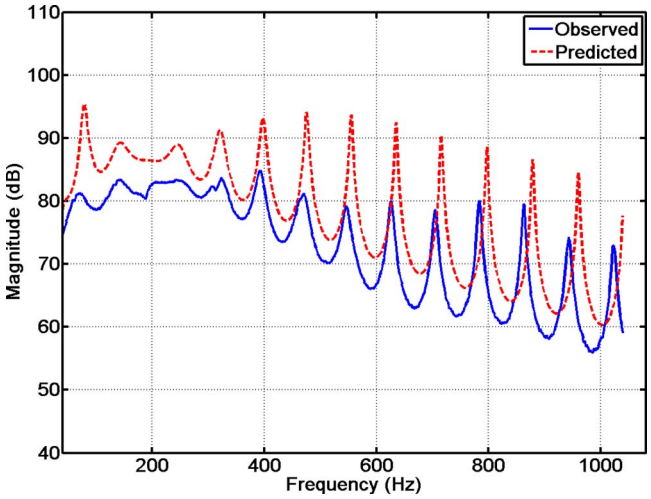


FIG. 9. (Color online) Frequency response at the cap end of tube comparing the observed response (tube in operation) to the model predicted response in small-holed tube.

not very well modeled, especially for the fundamental, the overall trend in the predicted and observed frequency shifts is readily observable. Similar results may be seen for the frequency response of the small-holed tube in Fig. 9. As with the large tube, there are some relative maxima and minima in the measured response that are absent from the predicted data below 300 Hz. Nevertheless, the measured and predicted responses both show the predicted upward shift in resonance frequency.

Table III quantifies the data in Figs. 8 and 9, showing both predicted and observed resonance frequencies for the small- and large-holed tubes. It also gives percentage error. The greatest error between predicted and measured frequencies is for the fundamental mode for both tubes. For the large-holed tube, the error is 17%, whereas for the small-holed tube, the error is 10%. Beyond the fundamental mode, the percentage error for all higher modes is less than 5% and is less than 2% in most cases. Hence, by including terms that account for the impedance of the holes and then calculating the resonance frequencies, the model accurately depicts the behavior of the tube and shows that the holes create the shift in resonance frequency. Based on altering various parameters within the model, we believe dominant source of error to be uncertainty in our estimate of the sound speed in the tube. Although a constant sound speed was used in the model, the sound speed, proportional to the square root of temperature, varies within the actual tube because the tube is hotter in the middle (where the inlet is) and cooler on the ends.

IV. DISCUSSION

Based on the results of the equivalent circuit model, it is clear that the holes are a primary cause of the shift in resonance frequencies. The fact that the lower modes are impacted more significantly than the higher modes can be qualitatively explained by considering the volume of gas in the hole as an acoustic mass, M_A, which is the actual mass of the gas divided by the square of the cross-sectional area of the hole. The acoustic impedance of the hole may then be written[11] as

$$\frac{\hat{p}}{\hat{U}} = j\omega M_A = j\omega M_M / S^2, \tag{7}$$

where M_M is the (mechanical) mass of air inside the hole, including assumed flanged end corrections,[10] and S is the cross-sectional area of the hole. Because the acoustic imped-

TABLE III. Predicted and observed resonance frequencies for both tubes and the error between the observed and the predicted. Note that there is no clearly observed distinction between the Helmholtz and the first mode.

Mode No.	Small hole res. frequencies (Hz)		Large hole res. frequencies (Hz)		Small hole % error	Large hole % error
	Observed	Predicted	Observed	Predicted		
1	71	79	145	124	10	17
2	142	143	183	178	0.7	2.8
3	242	246	274	263	1.6	4.2
4	323	321	342	335	0.6	2.1
5	393	398	413	410	1.0	0.7
6	470	475	488	486	1.1	0.4
7	546	555	563	564	1.6	0.2
8	626	635	640	643	1.4	0.5
9	705	716	720	723	1.5	0.4

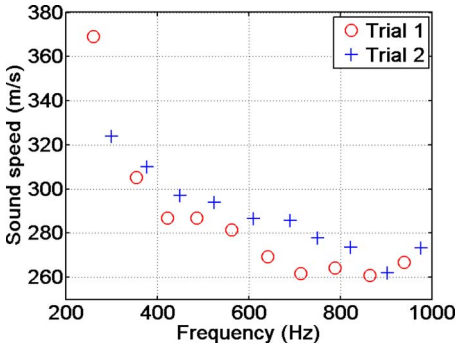


FIG. 10. (Color online) Calculated sound speed versus frequency using $c=\lambda f$ in large-holed tube. The distance between two flame peaks is used as a measurement of λ and f is the frequency input into the tube from the signal generator.

ance in the hole increases as a function of frequency, the holes play less of a role at high frequencies than at low frequencies and the tube begins to look like the no-hole case.

This same reasoning can be used to explain why the large-holed tube resonances are affected more than those of the small-holed tube. For a hole of constant length (the tube thickness in this case), M_M increases linearly with volume and consequently, as a function of S. However, as shown in Eq. (7), as the diameter of the hole is made larger, the acoustic impedance is found to be inversely proportional to S. Therefore, the large-holed tube resonances are more affected than those of the small-holed tube because the greater cross-sectional area of the hole reduces M_A and, therefore, the acoustic impedance of the hole. This may be useful information for one designing a tube: By making the holes as small as possible or the tube as thick as possible while maintaining sufficient mass flow rate of the flammable gas, the resonance frequency shift can be minimized.

It also must be noted that the loudspeaker affects the symmetry of the pressure response across the tube. The constant pressure resonance now decreases in amplitude away from the cap end, and symmetric modal responses are not expected. This occurs because the loudspeaker creates a boundary condition, which is dissimilar to that of a rigid cap: The tube is not "closed-closed." The loudspeaker's effect decreases with increasing frequency, because as its impedance increases as a function of frequency, it begins to act less like a compliant source and more like a closed end. In summary, we learn that the combined effect of the holes and loudspeaker shifts and smears the resonance frequencies, which are no longer harmonically related, and creates pressure asymmetry across the tube.

One implication of the results of this study is worth discussing. An example of how the wavelength of sound inside the flame tube could be measured via the flame pattern was shown in Fig. 1. Using this information and the drive frequency, one can use $c=\lambda f$ to calculate the sound speed inside the pipe. As a consequence of the shift in resonance frequencies due to the presence of the holes, this sound speed calculation practice will prove erroneous. Figure 10 shows the calculated sound speed based on measurements of the distance between flame maxima in the large-holed tube. For both trials, there is an apparent decrease in sound speed as a function of frequency, which is physically untenable. The sound speed predicted by $c=\lambda f$, especially for the lower modes, is too high to be accurate, taking into account the gas inside the tube and its temperature. The sound speeds for the higher frequencies are closer to the actual sound speed, which we estimated to be about 256 m/s. Therefore, the flame peak distance is not a reliable source for the wavelength in the equation $c=\lambda f$. Information previously available on the Internet indicated that others have observed this apparently dispersive sound speed, but have not attributed it to the presence of the holes and the shift in resonance frequency. These results were confirmed using the equivalent circuit model, where one can readily observe that the distance between pressure maxima is greater than a half-wavelength. This results in an erroneous sound speed calculation for the lower modes. For the higher modes, however, the distance between adjacent pressure maxima approaches a half-wavelength, which again indicates that the holes play a decreasing role at higher frequencies.

V. CONCLUSION

The Rubens flame tube serves well as a classroom demonstration, but calculating resonance frequencies or sound speeds is not a straightforward exercise of basic acoustics. Depending on how the tube is built, the phenomena observed here may or may not be strongly present. For example, smaller and fewer holes will decrease the resonance frequency shift and will allow an instructor to demonstrate and discuss the simpler physics in a quantitative fashion with greater accuracy. However, if the holes are too small or too far apart, this could compromise the effectiveness of the demonstration. In a more advanced setting, the tube could be used as a demonstration of parallel impedances or an example of acoustic masses, where it might be viewed as initially counterintuitive by students that the holes actually play less of a role at high frequencies where they look large relative to a wavelength. There are likely other uses for the flame tube, and instructors can take advantage of its complicated nature to teach students that the simple explanations of a physical system are often approximations that neglect potentially richer and important phenomena.

ACKNOWLEDGMENTS

This research was partially funded by an undergraduate grant from the Office of Research and Creative Activities at Brigham Young University. Jarom Giraud is also gratefully acknowledged for his preliminary work on this project.

[1]H. Rubens and O. Krigar-Menzel, "Flammenröhre für akustishe beobachtungen (Flame tube for acoustical observations)," Ann. Phys. **322**, 149–164 (1905).
[2]G. Ficken and C. Stephenson, "Rubens flame-tube demonstration," Phys. Teach. **17**, 306–310 (1979).
[3]D. Jihui and C. T. P. Wang, "Demonstration of longitudinal standing waves in a pipe revisited," Am. J. Phys. **53**, 1110–1112 (1985).
[4]G. Spagna, Jr., "Rubens flame tube demonstration: A closer look at the flames," Am. J. Phys. **51**, 848–850 (1983).
[5]H. Daw, "A two-dimensional flame table," Am. J. Phys. **55**, 733–737

(1987).
[6]H. Daw, "The normal mode structure on the two-dimensional flame table," Am. J. Phys. **56**, 913–915 (1988).
[7]G. R. Dix, "Development and comparison of highly directional loudspeakers," MS thesis, Department of Physics and Astronomy, Brigham Young University, Provo, UT, 2006.
[8]G. R. Plitnik and W. J. Strong, "Numerical method for calculating input impedances of the oboe," J. Acoust. Soc. Am. **65**, 816–825 (1979).
[9]D. H. Keefe, "Theory of the single woodwind tone hole," J. Acoust. Soc. Am. **72**, 676–687 (1982).
[10]L. E. Kinsler, A. R. Frey, A. B. Coppens, and J. V. Sanders, *Fundamentals of Acoustics* (Wiley, New York, 2000).
[11]L. L. Beranek, *Acoustics* (Acoustical Society of America, Woodbury, NY, 1996).

Two-dimensional wave propagation in an elastic half-space with quadratic nonlinearity: A numerical study

Sebastian Küchler[a)]
Institute for System Dynamics, University of Stuttgart, 70569 Stuttgart, Germany

Thomas Meurer[b)]
Automation and Control Institute, Vienna University of Technology, 1040 Vienna, Austria

Laurence J. Jacobs[c)]
School of Civil and Environmental Engineering, GW Woodruff School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332

Jianmin Qu[d)]
GW Woodruff School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332

(Received 6 March 2008; revised 18 December 2008; accepted 28 December 2008)

This study investigates two-dimensional wave propagation in an elastic half-space with quadratic nonlinearity. The problem is formulated as a hyperbolic system of conservation laws, which is solved numerically using a semi-discrete central scheme. These numerical results are then analyzed in the frequency domain to interpret the nonlinear effects, specifically the excitation of higher-order harmonics. To quantify and compare the nonlinearity of different materials, a new parameter is introduced, which is similar to the acoustic nonlinearity parameter β for one-dimensional longitudinal waves. By using this new parameter, it is found that the nonlinear effects of a material depend on the point of observation in the half-space, both the angle and the distance to the excitation source. Furthermore it is illustrated that the third-order elastic constants have a linear effect on the acoustic nonlinearity of a material.
 [DOI: 10.1121/1.3075597]

PACS number(s): 43.25.Dc, 43.25.Ba, 43.35.Cg [OAS] Pages: 1293–1301

I. INTRODUCTION

Most of the existing literature on waves in nonlinear solids only considers one-dimensional plane wave motion, and does not examine the multi-dimensional interaction between different types of waves. A classic problem in this area that has not been solved is the nonlinear version of Lamb's problem, namely, wave propagation in an elastic half-space with quadratic nonlinearity due to a time-dependent line-load on the surface. This nonlinear Lamb problem requires the solution of nonlinear hyperbolic partial differential equations that only allow closed form solutions in a few special cases. Hence, the following considers a second-order semi-discrete central scheme[1] in order to provide a detailed numerical resolution and analysis of the nonlinear effects. For this, the library CENTPACK (Ref. 2) is used with several enhancements required to deal with the semi-infinite domain.

Based on the numerical results for the nonlinear Lamb problem under consideration, the influence of the material nonlinearity is evaluated in the frequency domain. Here the excitation of higher-order harmonics can be observed and a detailed parameter study reveals the angular dependency of the nonlinear interactions, i.e., the relationship between the interacting waves and the position in the half-space with respect to the location of the source. The nonlinear behavior of a material is typically presented in terms of β, which is an intrinsic material parameter developed for one-dimensional longitudinal waves. This β can be presented in terms of the ratio of the amplitude of the second harmonic frequency to the square of the magnitude of the fundamental frequency.[3,4] In order to quantify the nonlinear effects in the current two-dimensional problem, a new parameter κ_r is introduced, which is similar to the acoustic nonlinearity parameter β for longitudinal waves. By using this κ_r parameter, it is possible to show that a nonlinear stress-strain relationship causes acoustic nonlinearities that are directly influenced by the interaction among different types of waves and the third-order elastic constants.

Before closing this section, it is worth mentioning that nonlinear ultrasonic techniques have attracted significant attention in recent years. In particular, the technique of second harmonic generation has been used to correlate fatigue damage in metallic materials.[5–10] It has been found that the acoustic nonlinearity parameter β correlates very well with the cumulative damage in several high temperature alloys subjected to cyclic load. It is well known in literature that the acoustic nonlinearity parameter β may come from two sources, namely, lattice anharmonicity and dislocations. It is believed that the increase in β during metal fatigue is mainly due to the increased plastic deformation, a manifestation of dislocation dynamics. This work, however, only considers

[a)]Electronic mail: sebastian.kuechler@isys.uni-stuttgart.de
[b)]Electronic mail: meurer@acin.tuwien.ac.at
[c)]Electronic mail: laurence.jacobs@coe.gatech.edu
[d)]Electronic mail: jianmin.qu@me.gatech.edu

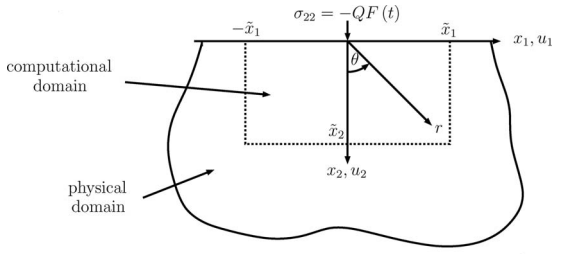


FIG. 1. Half-space at $x_3=0$ with a load applied normal to the surface together with the semi-infinite physical domain and the finite computational domain with numerical boundaries at $\pm\tilde{x}_1$ and $\tilde{x}_2$.

the effect of lattice anharmonicity on the nonlinear acoustic parameter. In other words, the half-space considered here is assumed to be elastic with quadratic nonlinearity so that lattice anharmonicity is represented by the third-order elastic constants.

The paper is organized as follows. In Sec. II the considered nonlinear Lamb problem is introduced. This is followed in Sec. III by a reformulation of the problem as a system of conservation laws and the introduction of the applied numerical solution approach, which is based on an appropriate high-resolution central scheme. The obtained numerical results are analyzed and evaluated in Sec. IV where in addition a detailed parameter study is performed. Some final remarks conclude the paper.

II. MODEL FOR WAVE PROPAGATION IN AN ELASTIC HALF-SPACE WITH QUADRATIC NONLINEARITY (THE NONLINEAR LAMB PROBLEM)

The model under consideration consists of two coupled hyperbolic partial differential equations, which describe the wave propagation in an elastic half-space with quadratic nonlinearity. For this, consider a Cartesian coordinate system (x_1,x_2,x_3) attached to an elastic half-space subjected to a time-dependent line-load applied at $x_1=x_2=0$, $-\infty<x_3<\infty$; see Fig. 1. Symmetry of the problem leads to a two-dimensional motion governed by the following second-order hyperbolic system of partial differential equations:

$$\frac{\partial\sigma_{11}}{\partial x_1}+\frac{\partial\sigma_{12}}{\partial x_2}=\rho\ddot{u}_1,$$

$$\frac{\partial\sigma_{12}}{\partial x_1}+\frac{\partial\sigma_{22}}{\partial x_2}=\rho\ddot{u}_2, \tag{1}$$

where ρ denotes the density of the material, $u_1=u_1\ (x_1,x_2,t)$ and $u_2=u_2\ (x_1,x_2,t)$ denote the displacement components in the x_1- and x_2-directions, respectively, and $\sigma_{ij}=\sigma_{ij}\ (x_1,x_2,t)$ denote the stress components. The boundary conditions for the Lamb problem depicted in Fig. 1 are given by

$$\sigma_{22}(x_1,0,t)=-QF(t)\delta(x_1),$$

$$\sigma_{12}(x_1,0,t)=0, \tag{2}$$

where $\delta(x_1)$ is the delta function representing the line-load at $x_3=0$ and $F(t)$ is the temporal input signal with the input amplitude Q. The half-space is assumed initially at rest such that the displacements and particle velocities are zero at time zero.

The derivation of the nonlinear stress-strain relationship can be found in literature.[11–14] Since this research considers only small strain deformations, the terms of the geometrical nonlinearity are neglected. The resulting constitutive equations are given by

$$\sigma_{11}=(\lambda+2\mu)u_{1,1}+\lambda u_{2,2}+(l+2m)u_{1,1}^2+lu_{2,2}(2u_{1,1}+u_{2,2})+\frac{m}{2}(u_{1,2}+u_{2,1})^2, \tag{3}$$

$$\sigma_{22}=(\lambda+2\mu)u_{2,2}+\lambda u_{1,1}+(l+2m)u_{2,2}^2+lu_{1,1}(2u_{2,2}+u_{1,1})+\frac{m}{2}(u_{1,2}+u_{2,1})^2, \tag{4}$$

$$\sigma_{12}=(u_{1,2}+u_{2,1})(\mu+mu_{1,1}+mu_{2,2}), \tag{5}$$

where λ and μ are second-order elastic constants, l and m are third-order elastic constants, and $u_{i,j}=\partial u_i/\partial x_j$.

III. NUMERICAL SOLUTION

In view of the numerical analysis, the considered nonlinear Lamb problem can be reformulated into a hyperbolic system of conservation laws being defined on a two-dimensional domain $(x_1,x_2)\in(-\infty,\infty)\times(0,\infty)$, i.e.,

$$\frac{\partial}{\partial t}\mathbf{q}+\frac{\partial}{\partial x_1}\mathbf{f}(\mathbf{q})+\frac{\partial}{\partial x_2}\mathbf{g}(\mathbf{q})=\mathbf{0}, \tag{6}$$

where $\mathbf{q}=\mathbf{q}(x_1,x_2,t)$.

This can be achieved by introducing the state vector $\mathbf{q}$ according to

$$\mathbf{q}(x_1,x_2,t)=\begin{bmatrix}q^{(1)}(x_1,x_2,t)\\ q^{(2)}(x_1,x_2,t)\\ q^{(3)}(x_1,x_2,t)\\ q^{(4)}(x_1,x_2,t)\\ q^{(5)}(x_1,x_2,t)\end{bmatrix}=\begin{bmatrix}\dot{u}_1\\ \dot{u}_2\\ u_{1,1}\\ u_{2,2}\\ u_{1,2}+u_{2,1}\end{bmatrix}, \tag{7}$$

with the two flux vectors

$$\mathbf{f}(x_1,x_2,t)=\begin{bmatrix}-\frac{1}{\rho}\sigma_{11}(q^{(3)},q^{(4)},q^{(5)})\\ -\frac{1}{\rho}\sigma_{12}(q^{(3)},q^{(4)},q^{(5)})\\ -q^{(1)}\\ 0\\ -q^{(2)}\end{bmatrix},$$

$$\mathbf{g}(x_1,x_2,t) = \begin{bmatrix} -\frac{1}{\rho}\sigma_{12}(q^{(3)},q^{(4)},q^{(5)}) \\ -\frac{1}{\rho}\sigma_{22}(q^{(3)},q^{(4)},q^{(5)}) \\ 0 \\ -q^{(2)} \\ -q^{(1)} \end{bmatrix}. \tag{8}$$

As can be easily deduced from Eq. (8), this formulation holds similarly for general constitutive equations, which might also involve rate-dependent behavior such as hysteresis.[15]

For a linear stress-strain relationship [$l=m=0$ in Eqs. (3)–(5)] both Jacobian matrices $\partial\mathbf{f}/\partial\mathbf{q}$ and $\partial\mathbf{g}/\partial\mathbf{q}$ of the flux vectors according to Eq. (8) have the same eigenvalues at $\nu_1=0$, $\nu_{2,3}=\pm c_l$, and $\nu_{4,5}=\pm c_t$, where c_l and c_t denote the longitudinal and the transverse phase velocity, respectively. It can also be shown that all linear combinations of the Jacobian matrices of the flux vectors have real eigenvalues. Hence, the linear system of conservation laws according to Eq. (6) is hyperbolic[16–18] and can be solved by many numerical schemes for hyperbolic systems of conservation laws.

For a nonlinear stress-strain relationship ($l\neq 0$ and $m\neq 0$), the hyperbolicity of Eq. (6) is not guaranteed in all regions, as there exist some regions, where the Jacobian matrices $\partial\mathbf{f}/\partial\mathbf{q}$ and $\partial\mathbf{g}/\partial\mathbf{q}$ may have complex eigenvalues. Since many numerical schemes are based on the hyperbolicity of a system and do not consider a change in the characteristic of a system, it is important to limit the system in the hyperbolic region, i.e., the eigenvalues must be real. Note that the eigenvalues depend only on $q^{(3)}$, $q^{(4)}$, and $q^{(5)}$ and thus on the spatial derivatives of $u_1(x_1,x_2,t)$ and $u_2(x_1,x_2,t)$. Hence the hyperbolicity can be guaranteed by limiting the input amplitude Q of the excitation signal at the boundary (see Fig. 1). This does not add additional constraints since the geometrical nonlinearities are neglected as previously described—the strains are assumed to be small.

Many numerical solution algorithms[1,19] have been developed and successfully applied in computational fluid dynamics for systems of conservation laws according to Eq. (6). In the following, for the numerical solution to the wave propagation problem, a high-resolution second-order central scheme is considered as proposed by Kurganov and Tadmor.[1] A brief description of this scheme providing the essential details required for the subsequent analysis can be found in the Appendix. The resulting system of ordinary differential equations (ODEs) is solved via an explicit second-order Runge–Kutta method (modified Euler method).[20] The complete numerical solution algorithm is implemented using the library CENTPACK (Ref. 2) with several enhancements required to deal with the semi-infinite domain $(x_1,x_2)\in(-\infty,\infty)\times(0,\infty)$.

One key factor for numerical simulations of real physical problems is the implementation of the boundary conditions since they have an immense influence on the solution of the problem. The boundary conditions in this research are implemented using the so-called ghost-cell method,[21] which has been integrated into CENTPACK. For this, the physical domain is extended by suitable ghost-cells, which are required in order to evaluate the flux vectors along the boundaries to determine the inflow (or outflow) originating from the free surface as required for the numerical solution (cf. Appendix). This avoids spurious oscillations in the numerical solutions that are due to the fact that the numerical evaluation of the flux vectors according to Eq. (A2) is based on an averaging process involving flux limiters given in Eq. (A5). Once integrated into the scheme, the state values in the ghost-cells are prescribed at the beginning of each integration step depending on the respective boundary conditions according to Eq. (2) and the interior (physical) solution of the previous time step. Figure 2 shows the extension of the free surface at $x_2=0$ of the elastic half-space. For the implementation of the ghost-cell method, it is required that the computational domain being a finite part of the physical domain (see Fig. 1) ends exactly at the physical boundary. Thus the grid of the computational domain used for the numerical simulation in this research is

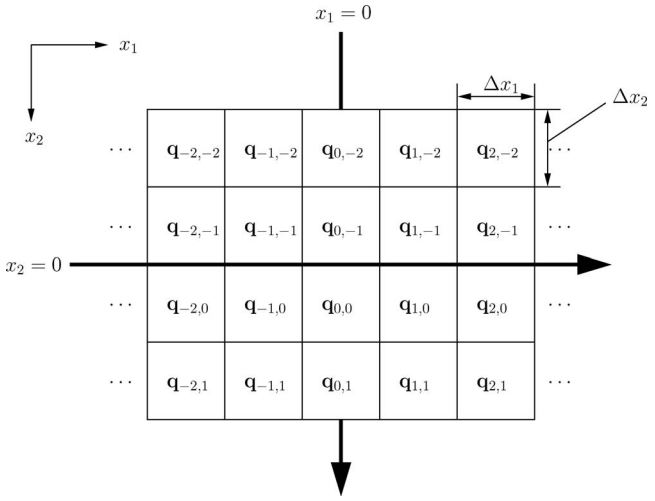


FIG. 2. Extension of the computational domain by ghost-cells at $x_2=0$.

$$x_{1_j}=j\Delta x_1 \quad \text{and} \quad x_{2_k}=\frac{\Delta x_2}{2}+k\Delta x_2; \quad j\in\mathbb{Z}, \quad k\in\mathbb{N}, \tag{9}$$

where Δx_1 and Δx_2 denote the spatial discretizations in the x_1- and x_2-directions. This grid choice guarantees the ending of the computational domain exactly at the free surface, as the discretized cell (x_{1_j},x_{2_k}) covers the computational domain corresponding to $[x_{1_j}-\Delta x_1/2,x_{1_j}+\Delta x_1/2]\times[x_{2_k}-\Delta x_2/2,x_{2_k}+\Delta x_2/2]$. Furthermore the choice according to Eq. (9) allows the application of the line-load exactly at $x_1=0$, since x_{1_j} approximates the solution exactly at $x_1=0$ for $j=0$.

The values in the ghost-cells ($k=-1,-2$) have to ensure that the free surface boundary conditions according to Eq. (2) are satisfied. Since the particle velocities $q^{(1)}=\dot{u}_1$ and $q^{(2)}=\dot{u}_2$ are not influenced directly by the boundary conditions, they are symmetrically extrapolated to the ghost-cells, which yields

$$q^{(1)}_{j,-k}=q^{(1)}_{j,k-1},$$

$$q^{(2)}_{j,-k} = q^{(2)}_{j,k-1}, \tag{10}$$

where $j \in \mathbb{Z}$, $k=1,2$, and $q^{(i)}_{j,k}=q^{(i)}_{j,k}(j\Delta x_1, \Delta x_2/2+k\Delta x_2, t)$. The same holds for $\sigma_{11}(q^{(3)},q^{(4)},q^{(5)})$, while $\sigma_{22}(q^{(3)},q^{(4)},q^{(5)})$ and $\sigma_{12}(q^{(3)},q^{(4)},q^{(5)})$ are influenced directly by the free surface boundary conditions according to Eq. (2) such that they have to be considered differently. Figure 2 shows that the values exactly on the physical boundary are given by linear interpolation between the ghost-cells and the interior cells for $k=0,1$. Since the values on the physical boundary for $\sigma_{22}(q^{(3)},q^{(4)},q^{(5)})$ and $\sigma_{12}(q^{(3)},q^{(4)},q^{(5)})$ are given by the boundary conditions, the required values for $q^{(3)}$, $q^{(4)}$, and $q^{(5)}$ in the ghost-cells are obtained together with the symmetrical extrapolation of $\sigma_{11}(q^{(3)},q^{(4)},q^{(5)})$ from a set of equations

$$\begin{aligned}
&\sigma_{11}(q^{(3)}_{j,-k},q^{(4)}_{j,-k},q^{(5)}_{j,-k}) - \sigma_{11}(q^{(3)}_{j,k-1},q^{(4)}_{j,k-1},q^{(5)}_{j,k-1}) = 0,\\
&\sigma_{22}(q^{(3)}_{j,-k},q^{(4)}_{j,-k},q^{(5)}_{j,-k}) + \sigma_{22}(q^{(3)}_{j,k-1},q^{(4)}_{j,k-1},q^{(5)}_{j,k-1})\\
&\quad = 2QF(t)\delta(x_{1_j}),\\
&\sigma_{12}(q^{(3)}_{j,-k},q^{(4)}_{j,-k},q^{(5)}_{j,-k}) + \sigma_{12}(q^{(3)}_{j,k-1},q^{(4)}_{j,k-1},q^{(5)}_{j,k-1}) = 0,
\end{aligned} \tag{11}$$

with $j \in \mathbb{Z}$, $k=1,2$, and $q^{(i)}_{j,k}=q^{(i)}_{j,k}(j\Delta x_1, \Delta x_2/2+k\Delta x_2, t)$.

For a linear elastic half-space ($l=0$, $m=0$), the solution to the set of equations according to Eq. (11) can be calculated explicitly. This is no longer true in general for a nonlinear elastic half-space ($l\neq 0$, $m\neq 0$). Hence, in this research, the Newton–Raphson method is used to approximate the solution to the set of equations according to Eq. (11).

The convergence of the Newton–Raphson method to the correct physical solution depends highly on the initial guess. Since it is known that the solution in a nonlinear elastic half-space is mostly driven by the linear parts,[22] the initial guess for solving the nonlinear system of equations according to Eq. (11) using the Newton–Raphson method is obtained from the solution to the linear system of equations ($l=0$, $m=0$).

Note that the problem at hand has a semi-infinite solution domain (the physical domain), but the numerical solution requires a finite solution domain (the computational domain). Hence, one has to avoid the appearance of reflections at the artificial boundaries at $\pm\tilde{x}_1$ and $\tilde{x}_2$ (see Fig. 1), which is done by introducing transparent boundary conditions. In this research, these transparent boundary conditions are also implemented via the ghost-cell method with the values in the ghost-cells being calculated by zero-order extrapolation.

As there only exists an analytical solution to the linear Lamb problem with $l=m=0$,[23–25] the accuracy of the proposed numerical solution procedure is first evaluated based on the linear scenario. For example, Fig. 3 shows the numerical solution and the exact analytical one for the displacement

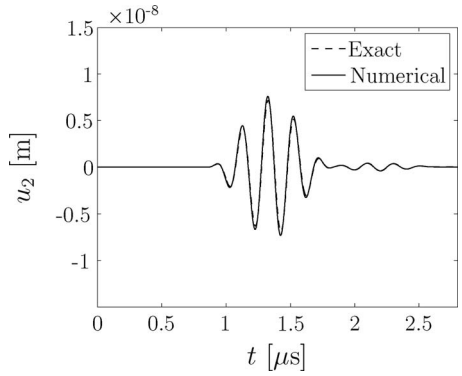


FIG. 3. Numerical solution of displacement u_2 at $r=5.2$ mm and $\theta=0°$ and exact solution for the original linear Lamb problem.

in the x_2-direction at $r=5.2$ mm and $\theta=0°$ (see Fig. 1) for the original linear Lamb problem. The input signal $F(t)$ used in this simulation is given by

$$F(t) = \begin{cases} \frac{1}{2}\sin(2\pi f_f t)\left(1-\cos\left(2\pi\frac{t}{t_f}\right)\right) & \text{if } t \le t_f \\ 0 & \text{otherwise,} \end{cases} \tag{12}$$

with $f_f=5$ MHz and $t_f=1$ μs, and the input amplitude is chosen to $Q=10^4$ N/m. It can be seen that the numerical solution is in good agreement with the exact analytical one. A detailed evaluation of the presented numerical solution procedure is given in Ref. 26, where it is shown that the solution procedure used in this research offers a high-resolution solution for wave propagation in both linear and nonlinear elastic half-spaces. The free surface boundary conditions according to Eq. (2) implemented via the ghost-cell method are satisfied for wave propagation in both linear and nonlinear elastic half-spaces.[26,27]

IV. NUMERICAL RESULTS AND DISCUSSIONS

Using the numerical solution technique discussed in Sec. III, the nonlinear Lamb problem is solved. The material parameters used correspond to Aluminum D54S (Ref. 28) and are summarized in Table I together with the parameters of the applied line load [cf. Eq. (12)]. The mesh density parameters used here are listed in Table II, where the parameters J and K denote the number of grid cells in the x_1- and x_2-directions, respectively. In order to ensure convergence, the numerical time step Δt must be chosen in order to fulfill the algorithm-dependent Courant–Friedrichs–Levy (CFL)-condition according to Eq. (A7). This CFL-condition guarantees that the analytical domain of dependence of the solution at a given point is completely included in the numerical domain of dependence of the difference scheme at the same

TABLE I. Material and input parameters.

ρ (kg/m^3)	λ (N/m^2)	μ (N/m^2)	l (N/m^2)	m (N/m^2)	f_f (MHz)	t_f (μs)	Q (N/m)
2719	4.91×10^{10}	2.6×10^{10}	-38.75×10^{10}	-35.8×10^{10}	5	1	10^4

TABLE II. Simulation parameters.

t_{sim} (μs)	Δx_1 (m)	Δx_2 (m)	J	K	$\tilde{x}_1$ (mm)	$\tilde{x}_2$ (mm)
2.8	9×10^{-6}	9×10^{-6}	2501	1250	±11.25	11.25

point.[16] A convergence analysis of the applied numerical scheme for the one-dimensional problem was performed Ref. 26 by refining the spatial resolution of the grid in both directions, i.e., increasing the number of cells in the computational domain. It is thereby shown that a further increase in the spatial discretization does not alter the obtained time and frequency domain signals, which confirms the convergence of the applied scheme. Hence, the simulation parameters summarized in Table II are used subsequently. The numerical solutions will be presented in the time domain first, followed by their analysis in the frequency domain.

Since the displacements are not obtained directly from the numerical scheme, particle velocities are used in the following presentation in order to study the nonlinear effects using a direct signal from the numerical scheme. In addition, the particle velocities obtained in Cartesian coordinates from the numerical scheme are transformed into polar coordinates to better reveal the physics of the Lamb problem. The transformation into polar coordinates is given by

$$\dot{u}_r = \dot{u}_1 \cos(\theta) - \dot{u}_2 \sin(\theta),$$

$$\dot{u}_\theta = \dot{u}_1 \sin(\theta) + \dot{u}_2 \cos(\theta). \tag{13}$$

A. Time domain

A typical output of the numerical solution to the particle velocity is shown in Fig. 4. For comparison purposes, solutions to the corresponding linear problem ($l=m=0$) are also shown in the same figure. Clearly, there is very little observable difference between the solutions to the linear and nonlinear problems. Such similarity also implies that the input signal strength Q is small enough that the nonlinear system remains hyperbolic.

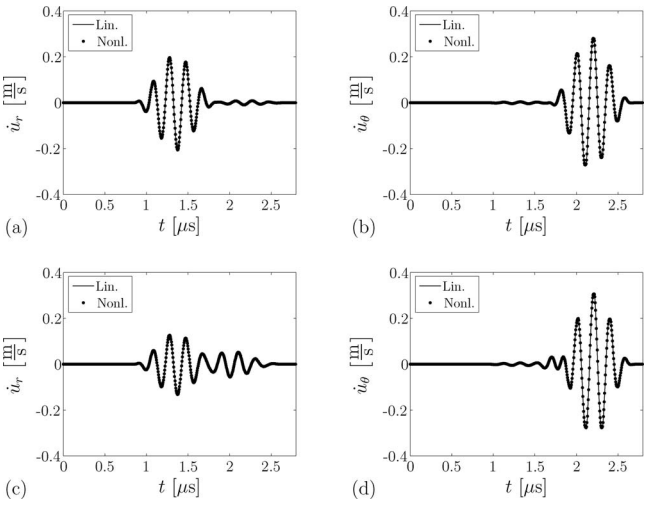


FIG. 4. Numerical solution of the particle velocities $\dot{u}_r$ (left) and $\dot{u}_\theta$ (right) in linear and nonlinear media at $r=5.2$ mm and $\theta=30°$ [(a) and (b)] and $r=5.2$ mm and $\theta=60°$ [(c) and (d)].

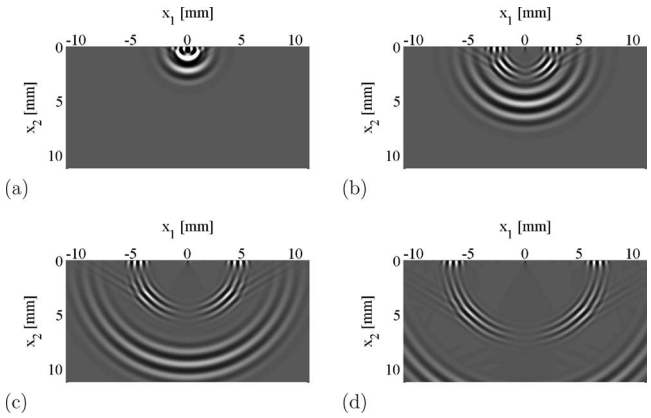


FIG. 5. Propagating wave fronts of σ_{22} in a nonlinear elastic half-space at different fixed times: (a) $t_1=0.7$ μs, (b) $t_2=1.4$ μs, (c) $t_3=2.1$ μs, and (d) $t_4=2.8$ μs.

As in the linear Lamb problem, the free surface of the half-space generates a head-wave that propagates between the longitudinal and shear wave fronts. For the material parameters used, the phase velocities of the longitudinal and shear waves are $c_l=6098$ m/s and $c_t=3092$ m/s, respectively. This leads to the angle $\theta_c=30.47°$, so that in the region $\theta<\theta_c$, only the longitudinal and shear waves are present. This is clearly seen in Fig. 5, where the amplitude of the stress component σ_{22} is plotted for four different instants in time for the nonlinear Lamb problem.

At the very beginning ($t=0.7$ μs), the stress wave is primarily producing pressure normal to the surface near the loading point, see Fig. 5(a). At a later time [Fig. 5(b)], the longitudinal wave is gradually separating from the transverse wave, and the head-wave initiating from the free surface is formed to bridge the longitudinal and transverse wave fronts. This becomes more clear in Figs. 5(c) and 5(d). Because of the cylindrical nature of the wave front, the amplitude of σ_{22} is highest at $\theta=0°$ for the longitudinal wave and gradually decreases for increasing θ. This is best demonstrated in Fig. 5(c). On the other hand, it is just the opposite for the transverse wave, for it induces no σ_{22} at $\theta=0°$, see Fig. 5(d). It is also seen that in the region $\theta>\theta_c$, the wave field is much more complicated due to the interaction between the head and the transverse waves. Figure 5(d) shows additionally that the introduced transparent boundary conditions do not produce any artificial reflections in the simulation results.

B. Frequency domain

To better observe the effect of material nonlinearity, the frequency spectra of the time harmonic signals can be obtained by using the fast Fourier transform (FFT). This section analyzes the frequency spectra of the longitudinal time domain signals. Clearly, this is possible only after a certain propagation distance when the longitudinal wave is completely separated from the transverse wave. To extract the longitudinal wave signal from the rest of the signals a Hann (Hanning) window is applied over the longitudinal wave part

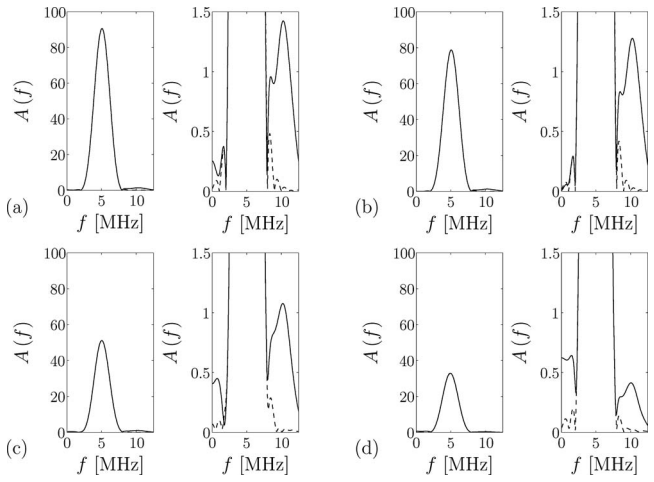


FIG. 6. FFT of the particle velocity $\dot{u}_r$ in nonlinear (solid) and linear (dashed) media at $r=5.2$ mm and different angles θ with zoom: (a) $\theta=0°$, (b) $\theta=30°$, (c) $\theta=60°$, and (d) $\theta=89.5°$.

in performing the FFT. In Ref. 22 it is shown that a fine spatial discretization is required to perform a frequency analysis of the numerical results and that the spatial discretization determines up to which frequency a convergence of the solution in the frequency domain is guaranteed. As mentioned above, a convergence analysis of the applied numerical scheme in the time as well as in the frequency domain can be found in Ref. 26. There it is shown that the numerical scheme used in this research requires a spatial discretization of approximately 60 grid cells per wavelength to guarantee convergence in the frequency domain. For the parameters used here, this condition is met for the longitudinal wavelength up to about 12.5 MHz. Hence, only frequencies below 12.5 MHz are considered. Note that the input signal according to Eq. (12) together with the parameters of Table I contains a single frequency at $f_f=5$ MHz, since the cosine part of the input signal is only applied over one period and hence has no effect on the fundamental frequency of the input signal. Thus, the frequency range up to 12.5 MHz includes both the fundamental frequency (5 MHz, the frequency of the applied load) and its second-order harmonic (10 MHz).

Figure 6 shows the magnitudes of the FFT spectra of $\dot{u}_r$, which represents the particle velocity of the longitudinal wave for both the linear (dashed lines) and nonlinear (solid lines) Lamb problems. The different sub-figures are the results from time domain signals observed at $r=5.2$ mm and $\theta=0°$, 30°, 60°, and 89.5°, respectively. There are two plots in each sub-figure with different vertical scales to reveal more details near the second-order harmonic (10 MHz).

It is seen that the amplitude $A(f)$ of the spectra at the fundamental frequency denoted by $A_1=A(f)|_{f=5\ \mathrm{MHz}}$ is virtually the same for both the linear and nonlinear Lamb problems. The nonlinearity plays a role only at the second-order harmonic amplitude $A_2=A(f)|_{f=10\ \mathrm{MHz}}$. Such phenomenon has been observed and used for characterizing material nonlinearity.[5–10] For the nonlinear problem (solid lines), some side lobes are observed in the second harmonic. This is caused by the use of the Hann window in performing the FFT. It is speculated that the zero frequency component shown in Fig. 6 could be an effect of material nonlinearity.

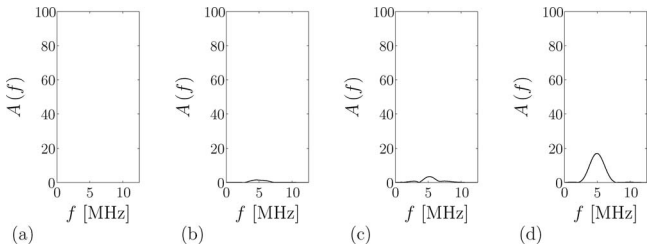


FIG. 7. FFT of the particle velocity $\dot{u}_\theta$ in nonlinear (solid) and linear (dashed) media at $r=5.2$ mm and different angles θ: (a) $\theta=0°$, (b) $\theta=30°$, (c) $\theta=60°$, and (d) $\theta=89.5°$.

It is noticed that the amplitude of the second-order harmonic decreases with increasing θ. This change is due to the interaction among the different waves. As discussed earlier, the wave field is no longer pure longitudinal at any point in the region $\theta>\theta_c=30.47°$. Such interaction is also demonstrated in Fig. 7 where the FFT of $\dot{u}_\theta$ is shown, representing the shear component of particle motion, at $r=5.2$ mm and $\theta=0°$, 30°, 60°, and 89.5°, respectively. This shear component of particle motion in the longitudinal wave part is due to reflections from the free surface. It is seen that the shear component of the particle motion is virtually absent at $\theta=0°$, and gradually increases with increasing θ. At $\theta=89.5°$, the magnitude of $\dot{u}_\theta$ is almost one-half of that of $\dot{u}_r$, indicating a strong presence of the shear motion.

In Ref. 4, it is shown that the value of $A_2/A_1^2k_l^2$, where k_l is the wavenumber of the longitudinal wave, increases linearly with the propagation distance in the case of cylindrical acoustical waves, where only a longitudinal wave occurs. It is thus expected that for the nonlinear Lamb problem considered here, the ratio $A_2/A_1^2k_l^2$ should also change linearly with propagation distance in the region where only the longitudinal wave is present. Figure 8 shows $A_2/A_1^2k_l^2$ for $\dot{u}_r$ versus the propagation distance r for different values of θ. It is seen that for each θ the ratio of $A_2/A_1^2k_l^2$ becomes almost proportional to r after a certain distance beyond which the longitudinal wave becomes dominant (or is separated from the other waves). To quantify such proportionality, define

$$\kappa_r=\frac{A_2}{rA_1^2k_l^2}. \tag{14}$$

Analogous to planar bulk waves, the slope κ_r is referred to as the acoustic nonlinearity parameter[3,4] for the longitudinal

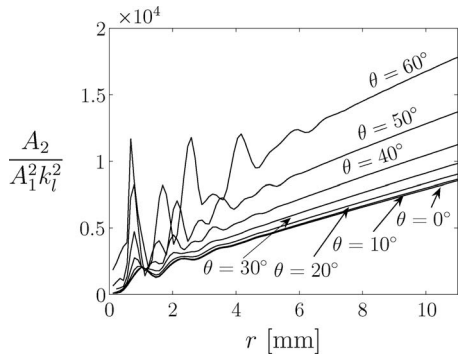


FIG. 8. $A_2/A_1^2k_l^2$ of the FFT of $\dot{u}_r$ over the propagation distance r for different angles θ.

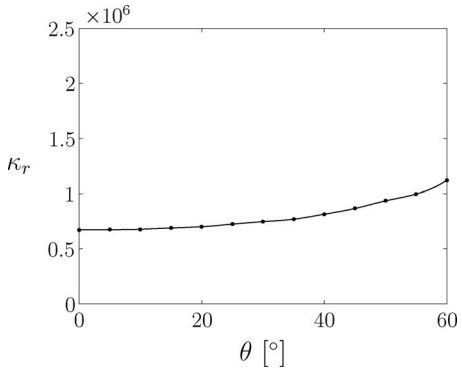


FIG. 9. Nonlinearity parameter κ_r over the angle θ.

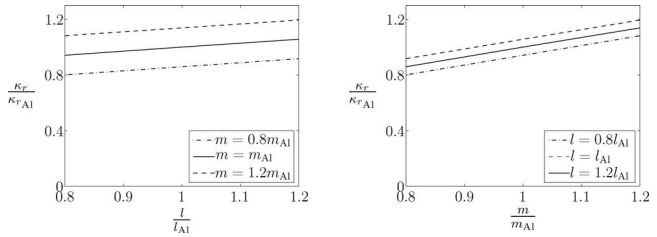


FIG. 10. κ_r normalized by its value for aluminum at $\theta=0°$ for different fixed m ($0.8m_{\mathrm{Al}}$, m_{Al}, and $1.2m_{\mathrm{Al}}$) in dependence of l (left) and for different fixed l ($0.8l_{\mathrm{Al}}$, l_{Al}, and $1.2l_{\mathrm{Al}}$) in dependence of m (right).

wave in the Lamb problem. The variation of κ_r with respect to θ is shown in Fig. 9. Clearly, κ_r is almost constant for $\theta < 30°$ where the longitudinal wave overwhelmingly dominates. For larger θ, more shear motion appears and κ_r gradually increases.

Such dependence of κ_r on θ may be explained by the fact that when both longitudinal and transverse waves are present, the second-order harmonic may be written as in Ref. 13

$$u_x = B_1 \cos(k_l x - \omega t) + \underbrace{B_1^2 F_1 x \cos[2(k_l x - \omega t)] + B_2^2 F_2 \sin[(k_t - k_l)x]\cos[(k_t + k_l)x - 2\omega t]}_{\text{second-order harmonic}},$$

$$u_y = B_2 \cos(k_t x - \omega t) + \underbrace{B_1 B_2 F_3 \{\sin[2(k_t x - \omega t)] - \sin[(k_t + k_l)x - 2\omega t]\}}_{\text{second-order harmonic}}, \tag{15}$$

where for the displacement components u_x and u_y, B_1 and B_2 are, respectively, the amplitudes of longitudinal and transverse components of a time-harmonic plane wave propagating in the x-direction with a fundamental frequency ω, x is the propagation distance, k_l and k_t are the wavenumbers of longitudinal and transverse waves, respectively, and the constants F_i are related to the third-order elastic constants in the following fashion:

$$F_1 \propto 2m + l, \quad F_2 \propto m, \quad F_3 \propto m. \tag{16}$$

As shown in Sec. IV A, for the nonlinear Lamb problem considered, a pure longitudinal wave field can only exist in the region $\theta < \theta_c \approx 30°$. Even in this region, the propagation distance has to be long enough so the longitudinal wave can be fully separated from the transverse wave. In such a case, the amplitude of the transverse wave B_2 in Eq. (15) is zero. Accordingly, the nonlinearity parameter becomes simply $\kappa_r = F_1$. For $\theta > 30°$ the amplitude of the transverse component changes with increasing θ. In other words, B_2 in Eq. (15) would depend on θ so that κ_r would also depend on θ.

The solution given by Eq. (15) also explains the oscillations shown in Fig. 8. For example, in the case of $\theta=60°$, the ratio $A_2/A_1^2 k_l^2$ oscillates until after $r > 7$ mm. Such oscillation is caused by the beating term, $\sin[(k_t - k_l)x]$, appearing in one of the second-order harmonics that is due to the presence of shear motion of the particle velocity.

To investigate how the nonlinearity parameter κ_r changes with the third-order elastic constants, numerical simulations are conducted by varying l and m by $\pm 20\%$ from their respective values for aluminum. For $\theta=0°$, as discussed earlier, it is seen that $\kappa_r = F_1 \propto 2m + l$. This is confirmed by the numerical results, such as in Fig. 10. This relationship holds until about $\theta=30°$. After that, the transverse wave gets involved and the relationship between κ_r and the third-order elastic constants becomes more complex (see Fig. 11) where κ_r normalized by its value for aluminum (see Fig. 9) is presented. Such complex dependency of κ_r on

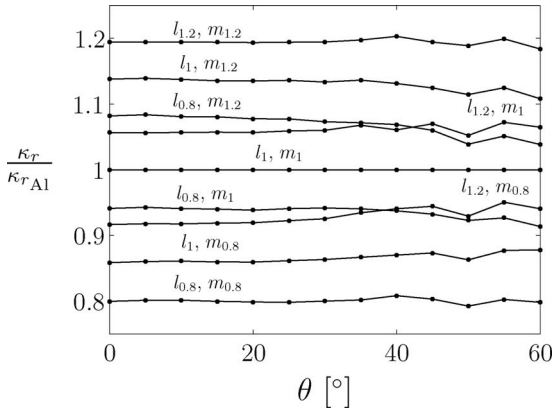


FIG. 11. Normalized nonlinearity parameter $\kappa_r/\kappa_{r_{\mathrm{Al}}}$ over the angle θ for different l and m. The parameters l_a and m_b denote al_{Al} and bm_{Al}, respectively.

the third-order elastic constants poses a challenge on determining l and m individually by measuring the amplitude of the second-order harmonic.

V. CONCLUSIONS

This research studies the application of a high-resolution numerical scheme to solve the two-dimensional Lamb problem in an elastic half-space with quadratic nonlinearity. Furthermore, the acoustic nonlinearity caused by a complex wave front is investigated. For this, the numerical library CENTPACK (Ref. 2) is used, where the problem considered requires the incorporation of several modifications in order to deal with the semi-infinite spatial domain. The convergence and accuracy of the numerical scheme are validated by considering the respective linear Lamb problem and comparing the numerical solution to an available analytical solution. In addition, a convergence check was performed by varying the spatial discretization in both directions, i.e., increasing the number of grid points in the computational domain. Both scenarios confirm the high accuracy of the numerical solution approach considered.

The numerical results obtained show that the effects due to the quadratic nonlinearity are best observed in the frequency domain by considering the spectra of the time domain signals. In particular, the evolution of higher-order harmonics can be observed. Thereby it is shown that the amplitude of the second-order harmonic depends linearly on the third-order elastic constants, namely, $\kappa_r \propto \alpha l + \gamma m$, where α and γ depend on the relative amplitude of the longitudinal and shear waves in a general wave field. Consequently, the relationship between the observed κ_r and the third-order elastic constants depend on the point of observation.

ACKNOWLEDGMENT

This work was partially supported by the German Academic Exchange Service (DAAD) and the National Science Foundation under Grant No. CMMI-0653883.

APPENDIX: NUMERICAL SCHEME

To approximate the solution for wave propagation in an elastic half-space with quadratic nonlinearity, this research uses a semi-discrete central scheme proposed by Kurganov and Tadmor.[1] The discretization of the half-space into a Cartesian grid used in this paper is given by Eq. (9). The flux vectors $\mathbf{f}(x_1,x_2,t)$ and $\mathbf{g}(x_1,x_2,t)$ according to Eq. (8) are approximated by numerical fluxes, which express the flux from one grid cell into a bordering grid cell at time t. With these numerical fluxes the semi-discrete central scheme[1] can be written as

$$\frac{d}{dt}\mathbf{q}_{j,k}(t) = -\frac{\mathbf{F}_{j+1/2,k}(t) - \mathbf{F}_{j-1/2,k}(t)}{\Delta x_1} - \frac{\mathbf{G}_{j,k+1/2}(t) - \mathbf{G}_{j,k-1/2}(t)}{\Delta x_2}, \tag{A1}$$

where $\mathbf{q}_{j,k}(t)$ is the state vector at $x_1 = x_{1_j}$ and $x_2 = x_{2_k}$ and $\mathbf{F}_{j\pm1/2,k}(t)$ and $\mathbf{G}_{j,k\pm1/2}(t)$ are the numerical fluxes approximating the flux in the x_1- and x_2-directions from one grid cell into the next grid cell (see Fig. 12).

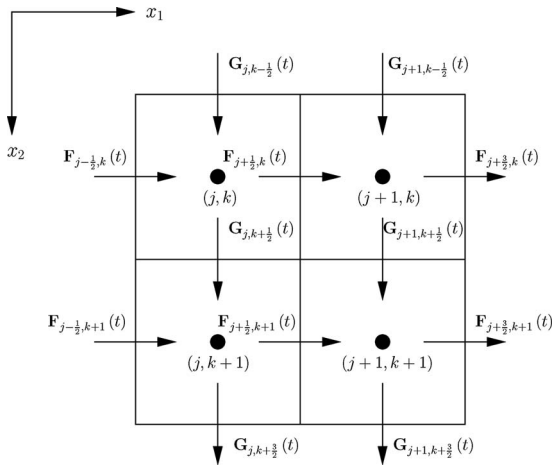


FIG. 12. Discretized Cartesian grid with the corresponding numerical fluxes.

The numerical fluxes are given by

$$\mathbf{F}_{j+1/2,k}(t) = \frac{\mathbf{f}(\mathbf{q}^+_{j+1/2,k}(t)) + \mathbf{f}(\mathbf{q}^-_{j+1/2,k}(t))}{2} - \frac{a_{j+1/2,k}(t)}{2}[\mathbf{q}^+_{j+1/2,k}(t) - \mathbf{q}^-_{j+1/2,k}(t)],$$

$$\mathbf{G}_{j,k+1/2}(t) = \frac{\mathbf{g}(\mathbf{q}^+_{j,k+1/2}(t)) + \mathbf{g}(\mathbf{q}^-_{j,k+1/2}(t))}{2} - \frac{b_{j,k+1/2}}{2}[\mathbf{q}^+_{j,k+1/2}(t) - \mathbf{q}^-_{j,k+1/2}(t)], \tag{A2}$$

which are expressed in terms of the intermediate values

$$\mathbf{q}^{\pm}_{j+1/2,k}(t) = \mathbf{q}_{j+1,k}(t) \mp \frac{\Delta x_1}{2}\left(\frac{\partial \mathbf{q}}{\partial x_1}\right)_{j+1/2\pm1/2,k}(t),$$

$$\mathbf{q}^{\pm}_{j,k+1/2}(t) = \mathbf{q}_{j,k+1}(t) \mp \frac{\Delta x_2}{2}\left(\frac{\partial \mathbf{q}}{\partial x_2}\right)_{j,k+1/2\pm1/2}(t), \tag{A3}$$

and the local speeds of propagation

$$a_{j+1/2,k}(t) = \max\left\{\rho\left(\frac{\partial \mathbf{f}}{\partial \mathbf{q}}(\mathbf{q}^+_{j+1/2,k}(t))\right), \rho\left(\frac{\partial \mathbf{f}}{\partial \mathbf{q}}(\mathbf{q}^-_{j+1/2,k}(t))\right)\right\},$$

$$b_{j,k+1/2}(t) = \max\left\{\rho\left(\frac{\partial \mathbf{g}}{\partial \mathbf{q}}(\mathbf{q}^+_{j,k+1/2}(t))\right), \rho\left(\frac{\partial \mathbf{g}}{\partial \mathbf{q}}(\mathbf{q}^-_{j,k+1/2}(t))\right)\right\}, \tag{A4}$$

where ρ denotes the spectral radius of the Jacobian matrix at the specified values. The derivatives $\partial\mathbf{q}/\partial x_1$ and $\partial\mathbf{q}/\partial x_2$ are determined (componentwise) by a slope limiting method, i.e., at the Θ-dependent family of min mod-like limiters[1]

$$\left(\frac{\partial \mathbf{q}}{\partial x_1}\right)_{j,k} = \text{minmod}\left(\Theta\frac{\mathbf{q}_{j,k}(t) - \mathbf{q}_{j-1,k}(t)}{\Delta x_1}, \frac{\mathbf{q}_{j+1,k}(t) - \mathbf{q}_{j-1,k}(t)}{2\Delta x_1}, \Theta\frac{\mathbf{q}_{j+1,k}(t) - \mathbf{q}_{j,k}(t)}{\Delta x_1}\right),$$

$$\left(\frac{\partial \mathbf{q}}{\partial x_2}\right)_{j,k} = \text{minmod}\left(\Theta \frac{\mathbf{q}_{j,k}(t) - \mathbf{q}_{j,k-1}(t)}{\Delta x_2}, \frac{\mathbf{q}_{j,k-1}(t) - \mathbf{q}_{j,k+1}(t)}{2\Delta x_2}, \Theta \frac{\mathbf{q}_{j,k+1}(t) - \mathbf{q}_{j,k}(t)}{\Delta x_2}\right), \tag{A5}$$

where $1 \leqslant \Theta \leqslant 2$ together with the multivariable minmod function

$$\text{minmod}(s_1, s_2, \ldots) = \begin{cases} \min\{s_j\} & \text{if } s_j > 0, \quad \forall j \\ \max\{s_j\} & \text{if } s_j < 0, \quad \forall j \\ 0 & \text{otherwise.} \end{cases} \tag{A6}$$

The introduction of the parameter Θ allows the adjustment of the numerical dissipation, where $1.1 \leqslant \Theta \leqslant 1.5$ is recommended for systems.[1] This research uses $\Theta = 1.4$.

In order to obtain convergence of the scheme, the following algorithm-dependent CFL-condition[1] must be satisfied:

$$\max\left(\frac{\Delta t}{\Delta x_1} a_{j+1/2,k}(t), \frac{\Delta t}{\Delta x_2} b_{j,k+1/2}(t)\right) \leq \frac{1}{2}, \quad \forall j,k, \tag{A7}$$

where $a_{j+1/2,k}(t)$ and $b_{j,k+1/2}(t)$ are evaluated by Eq. (A4).

The resulting system of ODEs according to Eq. (A1) is solved via an explicit second-order Runge–Kutta method (modified Euler method).[20]

[1] A. Kurganov and E. Tadmor, "New high-resolution central schemes for nonlinear conservation laws and convection-diffusion equations," J. Comput. Phys. **160**, 241–282 (2000).
[2] J. Balbás and E. Tadmor, CENTPACK, available at http://www.cscamm.umd.edu/centpack/software (Last viewed March, 2008).
[3] J. H. Cantrell, "Acoustic-radiation stress in solids. I. Theory," Phys. Rev. B **30**, 3214–3220 (1984).
[4] M. F. Hamilton and D. T. Blackstock, *Nonlinear Acoustics* (Academic, San Diego, CA, 1998).
[5] J. Herrmann, J. Y. Kim, L. J. Jacobs, J. Qu, J. W. Littles, and M. F. Savage, "Assessment of material damage in a nickel-base superalloy using nonlinear Rayleigh surface waves," J. Appl. Phys. **99**, 124913 (2006).
[6] J. Y. Kim, L. J. Jacobs, J. Qu, and J. W. Littles, "Experimental characterization of fatigue damage in a nickel-base superalloy using nonlinear ultrasonic waves," J. Acoust. Soc. Am. **120**, 1266–1273 (2006).
[7] J. H. Cantrell and W. T. Yost, "Nonlinear ultrasonic characterization of fatigue microstructures," Int. J. Fatigue **23**, 487–490 (2001).
[8] P. B. Nagy, "Fatigue damage assessment by nonlinear ultrasonic material characterization," Ultrasonics **36**, 375–381 (1998).
[9] J. Frouin, S. Sathish, T. E. Matikas, and J. K. Na, "Ultrasonic linear and nonlinear behavior of fatigued Ti-6Al-4V," J. Mater. Res. **14**, 1295–1298 (1999).
[10] C. Pruell, J. Y. Kim, J. Qu, and L. J. Jacobs, "Evaluation of plasticity driven material damage using Lamb waves," Appl. Phys. Lett. **91**, 231911 (2007).
[11] L. D. Landau and E. Lifshitz, *Theory of Elasticity* (Pergamon, London, 1959).
[12] F. D. Murnaghan, *Finite Deformation of an Elastic Solid* (Wiley, New York, 1951).
[13] A. L. Polyakova, "Nonlinear effects in a solid," Sov. Phys. Solid State **6**, 50–54 (1964).
[14] R. E. Green, *Ultrasonic Investigation of Mechanical Properties*, Treatise on Materials Science and Technology Vol. **3** (Academic, New York, 1973).
[15] T. Meurer, J. Qu, and L. J. Jacobs, "Wave propagation in nonlinear and hysteretic media—a numerical study," Int. J. Solids Struct. **39**, 5585–5614 (2002).
[16] E. Godlewski and P. A. Raviart, *Numerical Approximation of Hyperbolic Systems of Conservation Laws*, Applied Mathematical Sciences Vol. **118** (Springer, New York, 1996).
[17] M. E. Taylor, *Partial Differential Equations III—Nonlinear Equations*, Applied Mathematical Sciences Vol. **117** (Springer, New York, 1996).
[18] G. B. Whitham, *Linear and Nonlinear Waves*, Pure and Applied Mathematics (Wiley, New York, 1999).
[19] R. J. LeVeque, "Wave propagation algorithms for multidimensional hyperbolic systems," J. Comput. Phys. **131**, 327–353 (1997).
[20] C. F. Gerald and P. O. Wheatley, *Applied Numerical Analysis* (Addison-Wesley, Reading, MA, 1994).
[21] R. J. LeVeque, *Finite Volume Methods for Hyperbolic Problems*, Cambridge Texts in Applied Mathematics (Cambridge University Press, Cambridge, 2002).
[22] S. Vanaverbeke and K. V. D. Abeele, "Two-dimensional modeling of wave propagation in materials with hysteretic nonlinearity," J. Acoust. Soc. Am. **122**, 58–72 (2007).
[23] J. D. Achenbach, *Wave Propagation in Elastic Solids*, Applied Mathematics and Mechanics Vol. **16** (Elsevier Science, New York, 1975).
[24] M. J. Forrestal, L. E. Fugelso, G. L. Neidhardt, and R. A. Felder, "Response of a half-space to transient loads," Proceedings of the Engineering Mechanics Division Specialty Conference, ASCE, New York, NY (1966), pp. 719–751.
[25] J. Miklowitz, *The Theory of Elastic Waves and Waveguides*, Applied Mathematics and Mechanics Vol. **22** (Elsevier Science, New York, 1978).
[26] S. Küchler, "Wave propagation in an elastic half-space with quadratic nonlinearity," MS thesis, Georgia Institute of Technology, Atlanta, GA, 2007.
[27] S. Küchler, T. Meurer, J. Qu, and L. J. Jacobs, "On the solution to the Lamb problem in an elastic half-space with quadratic nonlinearity," *Review of Progress in Quantitative Nondestructive Evaluation*, AIP Conf. Proc. No. 975 (AIP, New York, 2007), pp. 70–77.
[28] R. T. Smith, R. Stern, and R. W. B. Stephens, "Third-order elastic moduli of polycrystalline metals from ultrasonic velocity measurements," J. Acoust. Soc. Am. **40**, 1002–1008 (1966).

Nonlinear nonclassical elasticity applied to the analysis of low frequency flexural vibrations: Theory and experiments

Ignacio Tinao Perez-Miravete and Cleofé Campos-Pozuelo
Instituto de Acústica, CSIC, Serrano 144, 28006 Madrid, Spain

Angel Perea
Instituto de Optica, CSIC, Serrano 121, 28006 Madrid, Spain

(Received 28 September 2007; revised 16 December 2008; accepted 28 December 2008)

Motivated by the increasing interest on nonlinear nondestructive damage detection, a comparison of the nonlinear elastic behavior of damaged samples and their intact state is presented. Flexural vibration is induced in thin glass plates with laser thermal shock induced micro-cracks by means of two thin piezo-ceramic patches glued in a bimorph configuration. The cases of direct excitation of a resonance and excitation of an internal resonance cases are considered. The resulting nonlinear vibrations exhibit the main features of quadratic hysteresis: linear variation of the resonance frequency and quadratic production of the third harmonic. A theoretical model for nonlinear resonant flexural vibrations based on the Preisach–Mayergoyz constitutive relations is proposed for damage quantification. Experimental comparison between the intact and damaged sample indicates an increase in the relevant nonlinearity parameter, indicating a widening of the hysteresis loop due to damage. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3075595]

PACS number(s): 43.25.Dc, 43.25.Gf, 43.25.Ts [OAS] Pages: 1302–1309

I. INTRODUCTION

Ultrasonic spectroscopy is one of the most common methods for characterizing elastic materials. The underlying idea behind this method is the fact that the elastic properties of a material can be studied by measuring the response to harmonic excitation. In particular, linear resonance analysis has been employed for some time to characterize damaged materials.[1]

Modern studies show nonlinear properties to be much more sensitive to increased damage[2] than their linear counterparts, making nonlinear resonance analysis a powerful tool for NDT. Several methods have been proposed to quantify the changes in the response of the system with increasing amplitude, such as NEWS methods SIMONRAS and NWMS.[3–16,2]

A remarkable property of materials with incipient microdamage is quadratic hysteresis, which is caused by the existence of a multi-valued quadratic stress-strain relation. The most prominent signatures under single frequency excitation are a linear shift of resonance frequency with increasing excitation level and the generation of a third harmonic whose amplitude is quadratically related to the amplitude of the fundamental.[6]

The birth of nonclassical nonlinear elasticity theory was motivated by the observation of nonlinear effects in some materials that could not be attributed to lattice anharmonicity. Nonlinearity was significantly increased when internal boundaries (defects) were formed. As reviewed in Ref. 7, this led first to the study of nonlinear behavior of nonbonded interfaces, where harmonic generation and frequency modulation were found and attributed to "clapping" between the faces of the contact. Then, the physics of a bonded interface was studied and described as a nonlinear relaxation system, with two possible equilibrium positions, "open" and "closed," depending on external conditions.[8]

This kind of contact has been demonstrated to be a good model for a defect and, under harmonic excitation, it reproduces dynamic hysteresis and several strongly nonlinear effects observed in acoustic wave-crack interaction.[7–9] This behavior was related to the observed nonclassical nonlinearity of rocks, with abnormally high cubic anharmonicity and end-point memory.

The first mathematical model describing the macroscopic response of a system containing bi-stable mesoscopic units is due to Preisach, in the context of magnetism. This was later adapted to nonlinear elasticity, leading to what is known as the phenomenological Preisach–Mayergoyz (PM) formalism.[10,11] The PM model considers defects as ideal relays with opened and closed equilibrium positions, depending on the loading. The contribution of these hysteretic units gives rise to nonclassical nonlinear effects, such as end-point memory,[10,11] increased nonlinear wave modulation, linear shift of resonance frequency, and the generation of the quadratic third harmonic in stationary vibrations, not only in rocks but in other materials subjected to crack damage, such as Plexiglas,[3] slate,[4] concrete,[5] and more recently, fatigued metal alloys.[12]

Modern approaches to hysteresis start with modeling the local micro-potential for contact defects. Then, the macroscopic response is coupled to the mesoscopic dynamics to obtain the constitutive relations. This kind of approach[2] leads to some qualitative deviations (for example, a second harmonic may be generated) from the simple PM formalism, which reduces the mesodynamics to a switching between two states.

 0001-4966/2009/125(3)/1302/8/$25.00

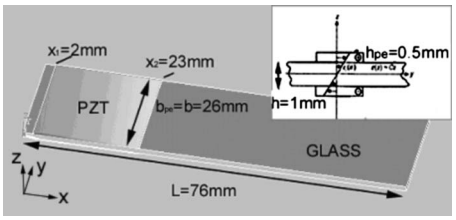


FIG. 1. System geometry.

Several theoretical studies (see, for example, Chap. 23 in Ref. 2) have presented models for resonant longitudinal vibrations in hysteretic media based on the PM model. In this paper, the nonlinear behavior of low frequency resonant flexural vibrations in thin plates is theoretically and experimentally studied using the simplest equation of state that contains the effects of quadratic hysteresis under single frequency excitation.

In Sec. II, a new theoretical model is proposed to describe flexural vibrations in mesoscopic materials. The model is based in the Euler theory of beams and the PM material model.

In Sec. III, the experimental procedure is described and the degree of nonlinearity is quantified for a glass plate with distributed thermal shock damage and compared with the intact state. This is followed by a discussion of the difficulties presented and the possible future trends on this area.

II. THEORY

In this section, a new theoretical model describing nonlinear vibrations of thin plates that attempts to account for the effects of micro-cracks on the macroscopic elastic characteristics of the material is presented. The objective is to relate the measured nonlinear features of low frequency standing waves with the material state, described by means of the PM model. Pure bending waves of finite but moderate amplitude are considered. Moreover, uniaxial "one-dimensional (1D) modes" are considered. This means that the vertical displacement w is considered to depend only on the x coordinate along the axis.

A. Hysteretic Euler beam model

The sample considered is a thin glass plate, to which two piezoelectric [lead zirconate titanate (PZT)] patches have been bonded in a bimorph configuration to excite flexural vibrations with free boundary conditions (see Fig. 1).

The dynamics of a thin plate can be modeled as a 1D Euler beam. Such an approximation is valid only for excitations near those resonances of the system that can be considered "uniaxial," that is, where the vertical displacement can be considered to depend only on the coordinate along the longest axis. Figure 2(b) shows the three-dimensional (3D) shape of such a mode, while Fig. 2(a) shows a comparison between the vertical displacement of the 3D (along the centerline of the top face) and 1D models obtained by finite element model for the fourth flexural mode with free boundary conditions.

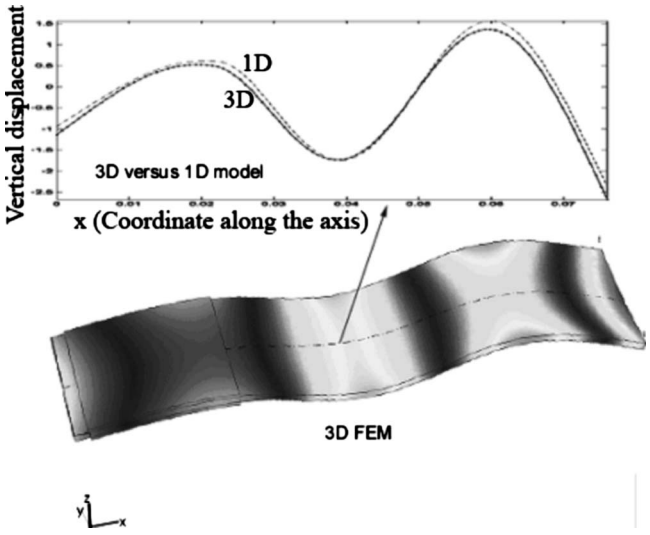


FIG. 2. Comparison between the 3D and 1D linear models for a uniaxial mode.

According to the Euler beam assumption, the vertical displacement $w(x)$ will be assumed to be solution of the following differential equation (see, for example, Refs. 13–15):

$$\rho(x)\frac{\partial^2 w}{\partial t^2} + \gamma\frac{\partial w}{\partial t} + \frac{\partial^2 m}{\partial x^2} = -\frac{\partial^2 m_{\mathrm{pe}}}{\partial x^2}, \tag{1}$$

where

$$\rho(x) = \rho hb + 2\rho_{\mathrm{pe}} h_{\mathrm{pe}} b_{\mathrm{pe}} \chi_{\mathrm{pe}}(x) \tag{2}$$

accounts for the mass distribution of the beam, including the piezoelectric contribution. Here, ρ, h, b and $\rho_{\mathrm{pe}}, h_{\mathrm{pe}}, b_{\mathrm{pe}}$ are the density, thickness, and width of the glass layer and PZT patches, respectively, and $\chi_{\mathrm{pe}}(x)$ is the characteristic function, which is equal to 1 in the region covered by PZT and 0 elsewhere. The γ term represents viscous air damping, while m and m_{pe} represent the moments induced by internal forces and piezoelectric actuator, respectively. The latter is given by[14,16]

$$m_{\mathrm{pe}} = -Y_{\mathrm{pe}} d_{31} b_{\mathrm{pe}} (h + h_{\mathrm{pe}}) \chi_{\mathrm{pe}}(x) V(t) \equiv -\kappa \chi_{\mathrm{pe}}(x) V(t), \tag{3}$$

where Y_{pe} is the Young modulus of the piezoelectric, d_{31} is the piezoelectric coefficient coupling the electric field to the deformation of the PZT layer, and $V(t)$ is the applied voltage. As for the internal moment, we distinguish between the linear ($m^{(\mathrm{lin})}$) and nonlinear ($m^{(\mathrm{nl})}$) part, the former being given by[14]

$$m^{(\mathrm{lin})} = YI(x)\frac{\partial^2 w}{\partial x^2} + cI(x)\frac{\partial^3 w}{\partial x^2 \partial t}, \tag{4}$$

where

$$YI(x) = \tfrac{1}{12} Y_0 h^3 b + \tfrac{2}{3} Y_{\mathrm{pe}}[(h/2 + h_{\mathrm{pe}})^3 - (h/2)^3] b_{\mathrm{pe}} \chi_{\mathrm{pe}}(x) \tag{5}$$

accounts for the total rigidity of the cross section, and

$$cI(x) = \tfrac{1}{2}bh^3c + \tfrac{2}{3}b_{\mathrm{pe}}[(h/2+h_{\mathrm{pe}})^3 - (h/2)^3]c_{\mathrm{pe}}\chi_{\mathrm{pe}}(x), \quad (6)$$

for the structural attenuation. Here, c and c_{pe} are Kelvin–Voight coefficients representing the anelastic material damping for both materials. Combining Eqs. (1)–(6), we obtain the following fifth order partial differential equation:

$$\rho(x)\frac{\partial^2 w}{\partial t^2} + \gamma\frac{\partial w}{\partial t} + cI(x)\frac{\partial^5 w}{\partial t\,\partial x^4} + YI(x)\frac{\partial^4 w}{\partial x^2} = \kappa\chi''_{\mathrm{pe}}(x)V(t) - \frac{\partial^2 m^{(\mathrm{nl})}}{\partial x^2}. \quad (7)$$

Note the appearance of the singular distribution $\chi''_{\mathrm{pe}} = \delta'(x-x_1) - \delta'(x-x_2)$ which can be defined in terms of the integral relation

$$\int f(x)\delta'(x-x_0)dx = \int -f'(x)\delta(x-x_0)$$

$$dx = -\left.\frac{df(x)}{dx}\right|_{x=x_0},$$

if the point x_0 belongs to the integration range. We shall now concentrate on the moment due to the nonlinear part of the internal stresses. This is obtained by integrating the moment of the axial stress across a section

$$m^{(\mathrm{nl})} = \int_S \sigma^{(\mathrm{nl})}(\varepsilon_x(x,z))z\,dy\,dz. \quad (8)$$

For small curvatures, the strain is reasonably approximated by $\varepsilon_x \approx w''(x)z$. The model is completed by specifying the constitutive equation $\sigma = \sigma(\varepsilon)$.

1. The hysteretic part of the stress-strain relation

In this work, the constitutive equation is modified to include the effect of distributed damage. According to the literature (see Ref. 10, for example), materials with mesoscopic microdamage exhibit nonclassical nonlinear behavior. In particular, the PM formalism has been shown to be a good description for the end-point memory effects attributed to micro-cracks. Following Chap. 23 in Ref. 2, the simplest constitutive equation describing quadratic hysteresis will be employed. The system is driven in the so-called simplex regime, where the constitutive equation represents only one loop (Fig. 3). This means that minor loops (which would arise in the presence of other harmonic components in the excitation or strong internal resonances) are neglected. Under these conditions, the additional contribution to the stress-strain relation takes the form

$$\sigma^{(\mathrm{nl})} = -Y_0\Gamma\left[\varepsilon\varepsilon_{\max} + \tfrac{1}{2}\mathrm{sign}(\dot{\varepsilon})(\varepsilon^2 - \varepsilon_{\max}^2)\right]. \quad (9)$$

In Ref. 10, the parameter Γ is related to the density of micro-cracks, and its increase implies a widening of the hysteresis loop. The goal of this work is to relate the increase in this parameter to the degree of damage.

Performing the integration specified in Eq. (8) the induced moment on the section is obtained:

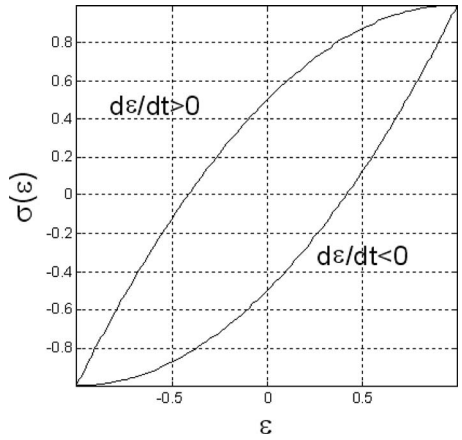


FIG. 3. Illustration of the constitutive equation $\sigma = \sigma(\varepsilon)$ for hysteretic media.

$$m^{(\mathrm{nl})} = \int_S \sigma^{(\mathrm{nl})} z\,dS = -Y_0\left[w''w''_{\max} + \frac{1}{2}\mathrm{sign}(\dot{w''})(w''^2 - w_{\max}{}''^2)\right]\int_S\int \Gamma z^2|z|\,dy\,dz, \quad (10)$$

where S stands for the glass layer section. We further assume Γ to be constant, i.e., the density of cracks across the section and in the damaged area is uniform, so that

$$m_H = \int_S \sigma_H z\,dS = -Y_0\Gamma S\frac{h^3}{32}\left[w''w''_{\max} + \frac{1}{2}(w''^2 - w_{\max}{}''^2)\right]. \quad (11)$$

By inserting Eq. (11) into Eq. (7), a new nonclassical nonlinear equation for flexural vibration including hysteresis is obtained:

$$\rho(x)\frac{\partial^2 w}{\partial t^2} + \gamma\frac{\partial w}{\partial t} + cI(x)\frac{\partial^5 w}{\partial t\,\partial x^4} + YI(x)\frac{\partial^4 w}{\partial x^2} = \kappa\chi''_{\mathrm{pe}}(x)V(t) + \frac{Y_0h^3}{12}\frac{3}{8}\Gamma h\frac{\partial^2}{\partial x^2}\left\{w''w''_{\max} + \frac{1}{2}\mathrm{sign}(\dot{w''})(w''^2 - w_{\max}{}''^2)\right\}. \quad (12)$$

We then obtain a dimensionless version of this equation by writing it in terms of the following variables:

$$\hat{x} \equiv x/L, \quad \hat{t} \equiv \sqrt{\frac{Y_0h^2}{12\rho L^4}}t, \quad \hat{w} \equiv w/h. \quad (13)$$

After omitting the hats, we then rewrite Eq. (12) as

$$(1+g_1\chi_{\mathrm{pe}}(x))\frac{\partial^2 w}{\partial t^2}+\left[\gamma^*\frac{\partial w}{\partial t}+c^*(x)\frac{\partial^5 w}{\partial t\,\partial x^4}\right]+(1+g_2\chi_{\mathrm{pe}}(x))\frac{\partial^4 w}{\partial x^2}=\kappa^*\chi''_{\mathrm{pe}}(x)V(t)+\frac{3}{8}\Gamma\left(\frac{h}{L}\right)^2\frac{\partial^2}{\partial x^2}\left\{w''w''_m+\frac{1}{2}\mathrm{sign}(\dot{w}'')(w''^2-w_m''^2)\right\},\tag{14}$$

where

$$g_1=\frac{2h_{\mathrm{pe}}\rho_{\mathrm{pe}}b_{\mathrm{pe}}}{h\rho b}=2.85,\quad g_2=\frac{2}{3}\frac{Y_{\mathrm{pe}}}{Y_0}[(1+2h_{\mathrm{pe}}/h)^3-1]\frac{b_{\mathrm{pe}}}{b}=4.48,$$

$$\kappa^*=\frac{Y_{\mathrm{pe}}}{Y_0}d_{31}\frac{12}{h_{\mathrm{pe}}}\left(\frac{L}{h}\right)^2\frac{h_{\mathrm{pe}}}{h}\left(\frac{h_{\mathrm{pe}}}{h}+1\right)=1.1\times10^{-2}.\tag{15}$$

The form of γ^* and c^* are unimportant, since the effect of attenuation will have to be experimentally measured.

2. Direct excitation of a resonance

We now consider a sinusoidal excitation voltage, $V(t)=V_0\cos(\Omega t)$, with a frequency near a resonance of the linear system, so that $\Omega\approx\omega_{n_1}$. We then make the ansatz that the response of the system is of the form

$$w=\tfrac{1}{2}(A_1e^{i\Omega t}+\mathrm{c.c.})\phi_{n_1}(x)=|A_1|\cos(\Omega t+\varphi_1)\phi_{n_1}(x),\tag{16}$$

where $A_1=|A_1|e^{i\varphi_1}$ is a complex parameter to be determined (c.c. stands for complex conjugate), and $\phi_{n_1}(x)$ is an eigenfunction for the unperturbed, nondissipative system:

$$\omega_{n_1}^2(1+g_1\chi_{\mathrm{pe}}(x))\phi_{n_1}+(1+g_2\chi_{\mathrm{pe}}(x))\frac{\partial^4\phi_{n_1}}{\partial x^2}=0.\tag{17}$$

These eigenfunctions are normalized so that

$$\int_0^L\phi_n(1+g_1\chi_{\mathrm{pe}}(x))\phi_m dx=\delta_{nm},$$

$$\int_0^L\phi_n(1+g_2\chi_{\mathrm{pe}}(x))\phi_m^{iv}dx=\omega_n^2\delta_{nm}.\tag{18}$$

We then introduce Eq. (16) into Eq. (14), multiply by $\phi_{n_1}(x)$, and integrate over x, taking Eq. (18) into account:

$$[(\omega_{n_1}^2-\Omega^2+2\xi_{n_1}\omega_{n_1}i\Omega)]A_1e^{i\Omega t}+\mathrm{c.c.}+\omega_{n_1}^2\hat{\Gamma}_{n_1n_1}\times[-A_1|A_1|e^{i\Omega t}+\mathrm{c.c.}+\mathrm{sign}(\sin(\Omega t+\varphi_1))[A_1^2e^{i2\Omega t}-2|A_1|^2+\bar{A}_1^2e^{-i2\Omega t}]]=\kappa^*V_0e^{i\Omega t}[\phi'_{n_1}(x_2)-\phi'_{n_1}(x_1)],\tag{19}$$

where $\xi_{n_1}\omega_{n_1}$ represents the total modal damping, and is given by

$$\xi_{n_1}\omega_{n_1}=\gamma^*\int_0^L\phi_{n_1}^2(x)dx+\int_0^L\phi_{n_1}(x)c^*(x)\phi_{n_1}(x)dx\tag{20}$$

and

$$\hat{\Gamma}_{n_1n_1}=\Gamma\int_{\text{damaged zone}}\frac{3}{8}\left(\frac{h}{L}\right)^2\frac{1}{\omega_{n_1}^2}|\phi''^3_{n_1}|dx\equiv\Gamma\int_{\text{damaged zone}}I_{n_1}(x)dx\tag{21}$$

represents the effect of the damaged zone localization influence for a given mode. Integration by parts has been performed, using the fact that boundary terms vanish because of the contour conditions (free plate).

Only contributions at frequency Ω are consistent with the approximation. Noting that

$$\mathrm{sign}(\sin(\Omega t+\varphi))=-\frac{2i}{\pi}\left(e^{i(\varphi+\Omega t)}+\frac{1}{3}e^{i3(\varphi+\Omega t)}+\frac{1}{5}e^{i5(\varphi+\Omega t)}+\cdots\right)+\mathrm{c.c.},\tag{22}$$

we finally obtain

$$e^{i\Omega t}:A_1\left[(\omega_{n_1}^2(1-\hat{\Gamma}_{n_1n_1}|A_1|)-\Omega^2)+i\left(2\xi_{n_1}\omega_{n_1}\Omega+\omega_{n_1}^2\frac{4\hat{\Gamma}_{n_1n_1}}{3\pi}|A_1|\right)\right]=\kappa^*V_0[\phi'_{n_1}(x_2)-\phi'_{n_1}(x_1)].\tag{23}$$

Equation (23) relates the experimentally obtained measures to the linear and nonlinear parameters of the model. In particular, it shows how the presence of microdamage (quantized by the parameter Γ) induces a linear shift of resonance frequency ($\Delta\omega_n/\omega_n\sim\Gamma_{n_1n_1}A/2$) and an increase in attenuation with amplitude.

3. Internal resonance excitation

Harmonic generation is another important feature of nonlinear behavior. The main difference with the nondispersive case is that the harmonics are in general not resonant, a result of the quadratic dependence of the resonance frequency on the wave vector k. This means that, when exciting a primary resonance, as in the above paragraph, we should not expect the harmonic generation to be large enough to be measurable. In the dispersive case, however, we may excite in a region where the fundamental response is nonresonant (corresponding to a slow variation in the frequency response), while 3Ω, for instance, corresponds to an eigenmode of the system. The possibility of exciting this internal resonance allows us to quantify the process of harmonic generation separately from frequency shift.

The case where $3\Omega\approx\omega_n$ is considered to obtain a measurable response on the third harmonic. The displacement is assumed to have the form

$$w = \tfrac{1}{2}A_1 e^{i\Omega t}\phi_{n_1}(x) + \mu\tfrac{1}{2}A_3 e^{i3\Omega t}\phi_{n_3}(x) + \text{c.c.}, \tag{24}$$

where μ is a small parameter, which will be considered of order $\mu \approx O(\Gamma)$. ϕ_n is assumed to fulfill Eq. (17). To order unity, we obtain by the same procedure followed in Sec. II A 2:

$$[\omega_{n_1}^2 - \Omega^2 + 2i\Omega\xi_{n_1}\omega_{n_1}]A_1 e^{i\Omega t} + \text{c.c.} = \kappa^* V_0 e^{i\Omega t}[\phi'_{n_1}(x_2) - \phi'_{n_1}(x_1)] + \text{c.c.}, \tag{25}$$

which is the linear fundamental response. Nonlinearities are neglected, as we assume that this contribution is not resonant. At order $O(\mu \approx \Gamma)$, we obtain by multiplying by $\phi_{n_3}(x)$ and integrating in x

$$[-(3\Omega)^2 + 2i(3\Omega)\xi_{n_3}\omega_{n_3} + \omega_{n_3}^2]A_3 e^{i3\Omega t} + \text{c.c.} = -\omega_{n_3}^2 \hat{\Gamma}_{n_3 n_1}\,\text{sign}(\sin(\Omega t + \varphi_1)) \times \left[\frac{A_1^2 e^{i2\Omega t} - 2|A_1|^2 + \bar{A}_1^2 e^{-i2\Omega t}}{4}\right], \tag{26}$$

where $\xi_{n_3}\omega_{n_3}$ is, again, the total damping for the mode $\phi_{n_3}(x)$ [given by Eq. (20)] and

$$\hat{\Gamma}_{n_3 n_1} = \frac{3}{8}\Gamma\left(\frac{h}{L}\right)^2 \frac{1}{\omega_{n_3}^2}\int_{\substack{\text{damaged}\\ \text{zone}}} \phi_{n_3}\frac{\partial^2}{\partial x^2}(\phi''_{n_1}|\phi''_{n_1}|)dx \tag{27}$$

represents the contribution of the crack location.

Using again Eq. (22), we can extract the 3Ω component of Eq. (26), yielding

$$A_3 = \omega_{n_3}^2 \hat{\Gamma}_{n_3 n_1}\frac{4}{15\pi}\frac{i|A_1|^2 e^{i3\varphi_1}}{[\omega_{n_3}^2 - (3\Omega)^2 + 2i(3\Omega)\xi_{n_3}\omega_{n_3}]}. \tag{28}$$

Fitting the experimental data on third harmonic generation to expression (28) provides another method to compare the increase of nonlinearity due to the presence of microcracks.

III. SAMPLES AND EXPERIMENTAL SETUP

The sample studied is a thin $75\times25\times2$ mm^3 glass plate. For sample damaging, an ArF ($\lambda=193$ nm) excimer laser beam having a pulse duration of 20 ns and energy per pulse of 290 mJ was focused on the glass plate surface at normal incidence to "write" damage lines. The spot at the glass plate surface was 1.1×0.4 mm^2 ($x\times y$) leading to an average energy density close to 16 J/cm^2. The glass plate was mounted on a motorized translational stage and moved uniformly along the short dimension of the glass plate (y-axis) at a speed of 0.1 mm/s using a computer controlled step motor while the laser pulse repetition rate was kept at 20 Hz. 11 lines parallel to the y-axis with a length of 18 mm and separated 2 mm were written. Taking into account the speed of the motor and the repetition rate, we compute that each point of the 14 mm central part of the line received approximately 200 laser shots.

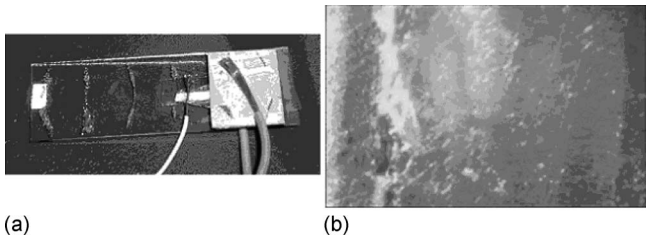


FIG. 4. (a) Glass sample and (b) zoom of a region of the thermally damaged zone.

Figure 4(a) shows the full system after the damaging stage. Figure 4(b) shows an optical image of the extended thermal damage induced by laser irradiation.

A. Experimental setup

The setup is schematically shown in Fig. 5. A Yokogawa FG300 signal generator is used to excite, through an amplifier, the two PZT patches. The vibration of the center of the beam tip is measured with a Polytec vibrometer, transferred via GPIB, and processed in the computer.

B. Direct resonance excitation

For the nonlinear characterization, the first four flexural modes of the plate were tested. For each of them, a set of resonance curves was obtained by sweeping the excitation over a range containing the corresponding resonance frequency for different excitation amplitudes.

When characterizing the different curves, the data for the intact sample should be fitted to the expression

$$\frac{|A_1|}{|A_{1,\max}|} = \frac{2\xi_{n_1}\omega_{n_1}^2}{\sqrt{[-\Omega^2 + \omega_{n_1}^2]^2 + [2\xi_{n_1}\omega_{n_1}\Omega]^2}}. \tag{29}$$

When trying to do so, however, discrepancies are shown even by the intact sample, and a linear shift in the resonance frequency ω_{n_1} with increasing driving voltage is observed [see Fig. 7(a), intact sample]. This attributed to the fact that domain switching processes induced by the time varying electrical and stress field in the PZT layers (see Refs. 17 and 18) act as an additional source for hysteretic behavior in the undamaged sample and force the characterization to be comparative. The experimental objective is thus to compare the degree of hysteretic behavior of both samples.

The modulus of expression (23) defines the complex amplitude A implicitly. To characterize the nonlinear reso-

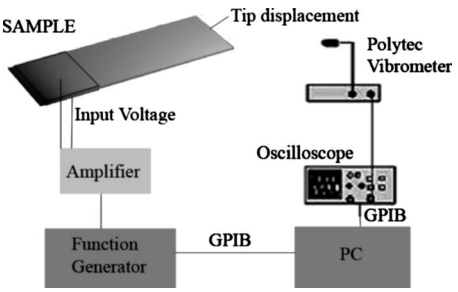


FIG. 5. Experimental setup

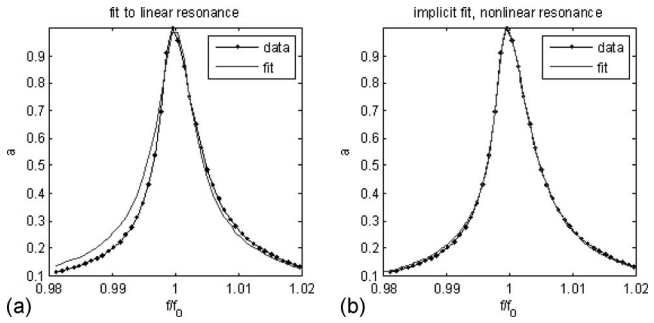


FIG. 6. Fitting of experimental data to (a) expression (29) and (b) expression (30).

nances, first, the experimental data are fitted to the function in Eq. (29), obtaining an estimation for ω_{n_1} and ξ_{n_1}. Both data for the intact and damaged sample are then fitted to

$$a = \frac{2\xi_{n_1}}{\sqrt{[-f^2 + (1 - \beta_1 - \beta_2 a)]^2 + [2\xi_{n_1} f]^2}}, \tag{30}$$

where $f = \Omega/\omega_{n_1}$, $a = |A_1|/|A_1|_{\max}$, and $\beta_2 = \hat{\Gamma}_{n_1 n_1}|A_{1,\max}|$. Notice that although we have neglected the effect of nonlinear attenuation, ξ_{n_1} will still show some shift with increasing amplitude. The introduction of the additional parameter β_1 allows for correcting the estimated frequency ω_{n_1}.

Expression (30) is an implicit relation between experimental variables a and f, depending on parameters $\boldsymbol{\alpha} \equiv (\beta_1, \beta_2, \xi_{n_1})$. To obtain the values of these parameters, a nonlinear least-squares fit is performed, minimizing the expression

$$F(\{a,f\};\alpha) = \sum_i [a_i^2[[-f_i^2 + (1 - \beta_1 - \beta_2 a_i)]^2 + [2\xi_{n_1} f_i]^2] - (2\xi_{n_1})^2]^2 \tag{31}$$

with respect to $(\beta_1, \beta_2, \xi_{n_1})$. The equations $\partial F/\partial \alpha_j = 0$ are solved iteratively with a standard Newton–Raphson scheme. Note that identifying $f_n(a)^2 = (1 - \beta_1 - \beta_2 a)$ means that $(\beta_1 + \beta_2)/2 \sim \Delta f_n/f_n \sim \Gamma_{n_1 n_1} A_{\max}/2$ approximately corresponds then to the frequency shift at the maximum amplitude. The difference between fitting data to expressions (29) and (30) is illustrated in Fig. 6. It should be noted too that the variable measured is the tip displacement, $w(L) = A_1 \phi_{n_1}(L)$. This scaling factor depends on the mode considered, but because we are only interested in a mode by mode comparison between the intact and damaged samples, we will omit this difference.

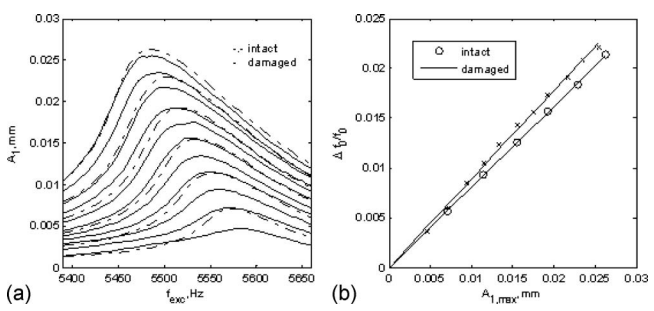


FIG. 7. Resonance curves for measured amplitude (a), $\Gamma_{n_1 n_1}$ parameter, as obtained from the nonlinear least-squares fitting.

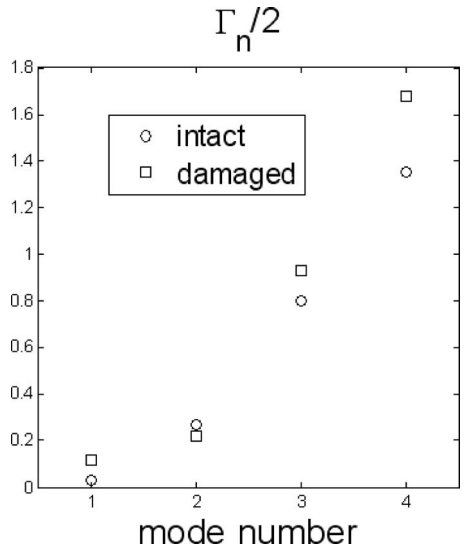


FIG. 8. Nonlinear characterization of the first four flexural modes.

Figure 7(a) shows the amplitude response at the excitation frequency for the third flexural mode. Results in Fig. 7(b) refer to experimental data from intact and damaged glass thin plates. The results for the characterization of the first four modes by this procedure are presented in the following table, and displayed in Fig. 8. Except for the second mode, an increase in the nonlinear hysteretic behavior is observed, indicating a widening of the hysteresis loop with extended damage.

Intact		Damaged	
f_0	$\Gamma_{n_1 n_1}/2$	f_0	$\Gamma_{n_1 n_1}/2$
928 Hz	0.039	933	0.116
2813 Hz	0.322	2751	0.221
5599 Hz	0.809	5610	0.887
8788 Hz	1.350	8700	1.685

C. Internal resonance excitation

To experimentally quantify the third harmonic, we choose a situation in which the fundamental response of the system is relatively flat, but where the third harmonic is resonant [Fig. 9(a)]. For this part, we chose to excite an internal resonance with the third mode. For this, the plate was excited

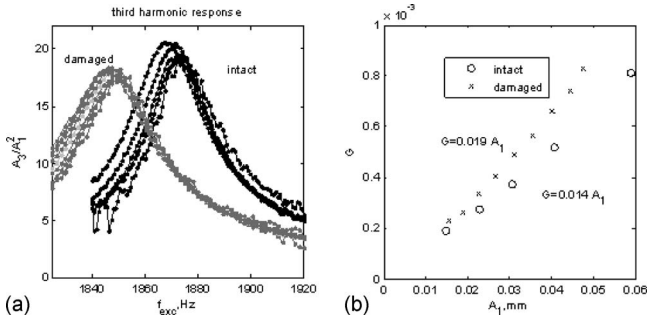


FIG. 9. Comparative analysis of third harmonic generation. (a)Third harmonic response. (b) $G \propto \Gamma_{n_3 n_1}|u_1|$ vs $|u_1|$.

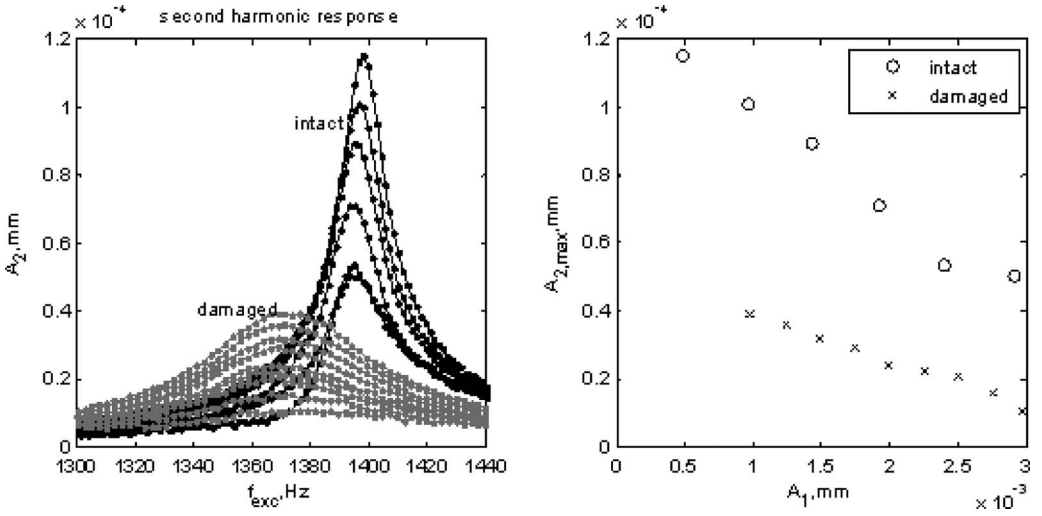


FIG. 10. Second harmonic generation.

at a frequency range near 1880 Hz (=5640 Hz/3). The coefficient $\hat{\Gamma}_{n_3 n_1}$ can be obtained fitting experimental data to the following expression:

$$\frac{|A_3|}{|A_1|} = G_{n_3 n_1} \frac{\omega_{n_3}^2}{\sqrt{[(3\Omega)^2 - \omega_{n_3}^2]^2 + (3\Omega)^2 [2\xi_{n_1}\omega_{n_3}]^2}}, \tag{32}$$

where the parameter $G_{n_3 n_1} = (4/15\pi)\hat{\Gamma}_{n_3 n_1}|A_1|$. This parameter is also proportional to Γ. Measured data, shown in Fig. 9(a) is then fitted to expression (32) using nonlinear least-squares. The set of obtained $\{u_1, G_{n_3 n_1}\}$ data points is fitted to the fundamental amplitude u_1, giving an estimate for the coefficient $\hat{\Gamma}_{n_3 n_1}$ and thus for the coefficient Γ.

In Fig. 9(b), we can see the result of this fitting, again showing increased nonlinear behavior in the damaged sample.

Finally, we would like to remark that a difference between classical nonlinearity and hysteresis is the preference of the later for odd harmonics. To show this, an internal resonance with the second mode was excited by driving the plate at a frequency near 1400 Hz (=2800 Hz/2). The second harmonic response is shown in Fig. 10, where we observe a decrease in the maximum amplitude with increasing driving amplitude.

We have proposed two different methods based on the simplest differential equation for low frequency flexural vibrations accounting for quadratic hysteresis. These have been shown to be equivalent methods for performing a comparative analysis between a sample with different degrees of damage.

IV. CONCLUSIONS

An experimental system for nonlinear characterization has been presented. Characterization is carried out by means of the analysis of nonlinear flexural vibrations of thin plates. A new theoretical model for nonlinear hysteretic flexural vibrations, based on the PM model for hysteresis, is proposed. Results for thin glass plates, intact and with laser induced damage clearly demonstrate the opening of the hysteresis loop. The experimental data obtained have been fitted in order to relate the degree of damage to the nonlinear parameter of the model. The possibility of exciting and analyzing this type of vibration for nonlinear detection of damage has been demonstrated in this way.

It is appropriate at this point to note that the intact sample proved to suffer from a high degree of nonlinearity, of the same order as that in the damaged sample. This is a consequence of the fact that thin patches of piezoelectric materials are known to show electromechanical hysteresis, due to domain reorientation/movement. This is often neglected in the context of nondestructive testing, but it has proved to be an important issue for the present analysis. This topic will be explored in a future work.

ACKNOWLEDGMENTS

This work is supported by the European Union Project No. STREP FP6-502927 (AERONEWS). Professor C. N. Afonso, from Instituto de Optica, CSIC Madrid, is thanked for the helpful discussions on laser irradiation of glasses. We would like to thank as well David K. Anthony and Jeff Porter, for his suggestions redacting this manuscript.

[1] A. D. Dimarogonas, "Vibration of cracked structures: A state of the art review," Eng. Fract. Mech. **55**, 831–857 (1996).
[2] *Universality of Nonclassical Nonlinearity*, edited by P. P. Delsanto (Springer, New York, 2006).
[3] K. Van Den Abeele, P. A. Johnson, and A. Sutin, "Nonlinear elastic waves spectroscopy techniques to discern material damage. Part I: Nonlinear wave modulation spectroscopy," Res. Nondestruct. Eval. **12**, 17–30 (2000).
[4] K. Van Den Abeele, J. Carmeliet, J. A. TenCate, and P. A. Johnson, "Nonlinear elastic waves spectroscopy techniques to discern material damage. Part II: Single mode nonlinear resonance acoustic spectroscopy," Res. Nondestruct. Eval. **12**, 31–42 (2000).
[5] K. Van Den Abeele and J. Visscher, "Damage assessment in reinforced concrete using spectral and temporal nonlinear vibration techniques," Cement Concr. Res. **30**, 1453–1464 (2000).
[6] K. Van Den Abeele, A. Sutin, J. Cameliet, and P. A. Johnson, "Microdamage diagnostics using nonlinear elastic wave spectroscopy (NEWS)," NDT Int. **34**, 239–248 (2001).
[7] I. Yu. Solodov, "Ultrasonics of nonlinear interfaces in solids: New physical aspects and NDE applications," WCU 2003, Paris, 7–10 September 2003.

[8]V. Gusev, B. Castagnede, and A. Moussatov, "Hysteresis of nonlinear bistable interface to continuously varying acoustic loading," Ultrasonics **41**, 643–654 (2003).
[9]I. Yu. Solodov and B. A. Korshak, "Instability, chaos, and memory in acoustic wave-crack interaction," Phys. Rev. Lett. **88**, 014303 (2002).
[10]K. R. McCall and R. A. Guyer, "Equation of state and wave propagation in hysteretic nonlinear elastic materials," J. Geophys. Res. **99**, 23887 (1994).
[11]R. A. Guyer and P. A. Johnson, "Nonlinear mesoscopic elasticity: Evidence for a new class of materials," Phys. Today **52**, 30–35 (1999).
[12]C. Campos-Pozuelo, C. Vanhille, and J. A. Gallego-Juárez, "Comparative study of the nonlinear behavior of fatigued and intact samples of metallic alloys," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **53**, 175–184 (2006).
[13]P. M. Morse and K. Uno Ingard, *Theoretical Acoustics* (McGraw-Hill, New York, 1968).
[14]J. M. Renno and D. J. Inman, "An experimentally verified model of a membrane mirror strip actuated using a piezoelectric bimorph," J. Vibr. Acoust. **129**, 631–640 (2007).
[15]H. T. Banks and D. J. Inman, "On damping mechanisms in beams," ASME J. Appl. Mech. **58**, 716–723 (1991).
[16]E. F. Crawley and E. H. Anderson, "Detailed models of piezoceramic actuation of beams," J. Intell. Mater. Syst. Struct. **1**, 4–25 (1990).
[17]B. L. Ball and R. C. Smith, "A stress dependent hysteresis model for PZT-based transducers," Proc. SPIE **5383**, 23–30 (2004).
[18]R. C. Smith, S. Seelecke, Z. Ounaies, and J. Smith, "A free energy model for hysteresis in ferroelectric materials," J. Intell. Mater. Syst. Struct. **14**, 719–739 (2003).

Nonlinear frequency shifts in acoustical resonators with varying cross sections

Mark F. Hamilton, Yurii A. Ilinskii, and Evgenia A. Zabolotskaya
Applied Research Laboratories, The University of Texas at Austin, Austin, Texas 78713-8029

(Received 8 October 2008; revised 16 December 2008; accepted 20 December 2008)

The frequency response and nonlinear resonance frequency shift of an acoustical resonator with losses and having a varying cross section were investigated previously using Lagrangian mechanics and perturbation for resonator shapes that are close to cylindrical [M. F. Hamilton, *et al.*, J. Acoust. Soc. Am. **110**, 109–119 (2001)]. The same approach is extended here to include resonators having any shape for which the Webster horn equation is a valid model in the linear approximation. Admissible shapes include cones and bulbs proposed for acoustical compressors. The approach is appropriate for approximate but rapid parameter estimations for resonators with complicated shapes, requiring far less computation time than for direct numerical solution of the one-dimensional model equation frequently used for such resonators [Ilinskii *et al.*, J. Acoust. Soc. Am. **104**, 2664–2674 (1998)]. Results for cone and bulb shaped resonators with losses are compared with results from the direct numerical solution. The direction of the resonance frequency shift is determined by the efficiency of second-harmonic generation in modes having natural frequencies below versus above the frequency of the second harmonic, and how the net effect of this coupling compares with the frequency shifts due to cubic nonlinearity and static deformation.
 [DOI: 10.1121/1.3075585]

PACS number(s): 43.25.Gf [OAS] Pages: 1310–1318

I. INTRODUCTION

During the 1990s, MacroSonix Corporation worked on development of resonant acoustical compressors for a variety of applications. Resonators of different shapes were fabricated and tested, and they were described using mathematical models developed in the company. A portion of this work was reported in Refs. 1 and 2. An important part of the research and development of these acoustical compressors was modeling the nonlinear behavior of the resonator and estimation of losses. The main tool for this purpose was a code for numerical integration of gasdynamic equations for the volume of the resonator and simultaneously within the boundary layer. The gasdynamic equations were written and integrated in curvilinear coordinates corresponding to the resonator shape. Different curvilinear coordinates were required for each different resonator shape. These coordinates are not orthogonal, and the equations were expressed in terms of covariant derivatives. Stable finite difference schemes were developed to solve these equations. The principal application of this code was calculation of losses.

For rough estimation of losses, a one-dimensional (1D) model was developed which in the linear approximation is equivalent to the Webster horn equation.[2] An efficient algorithm was developed to solve this equation numerically, and it was used for preliminary optimization of the resonator shape. Losses were minimized using Powell's algorithm,[3] and approximately 20 different parameters were adjusted to perform the optimization. The algorithm based on the Webster approximation was shown in Ref. 2, by comparison with experiment, to accurately predict the waveform in a cone resonator at a positive peak acoustic pressure amplitude of 2 atm. Predicted amplitudes of the lowest three modes were also shown to be in good agreement with experiment.

The model equation derived in Ref. 2 was also solved by Erickson and Zinn,[4] who used a Galerkin method. Selection of an appropriate trial function is important for this method, and when the choice is natural for a given resonator shape this approach is very efficient. The authors considered the specific case of an exponential horn. Chun and Kim,[5] numerically integrating the 1D conservation equations directly, simulated resonators having several different shapes.

For resonators with shapes close to cylindrical, as are used in thermoacoustic devices, an asymptotic theory in Lagrangian coordinates was developed using an expansion in powers of the amplitudes of the normal modes.[6] This perturbation method permitted simple explicit expressions to be derived for the frequency response and the shift in resonance frequency. Further discussion of this method may be found in Ref. 7.

Additional literature on the nonlinear theory of resonators based on geometrical assumptions consistent with the Webster horn equation can be found in the work already cited,[1,2,6] and especially in a more recent paper by Mortell and Seymour.[8]

In the present paper, the restriction in Ref. 6 to resonators that are close to cylindrical in shape is removed. The accuracy of the present model is demonstrated by comparison with numerical solutions of the fully nonlinear model equation presented in Ref. 2. The present theory is valid for any resonator shape for which the Webster horn equation is an accurate model in the linear approximation. For acoustic Mach numbers up to about 0.2 or 0.3, it is shown to provide reasonably accurate predictions of the frequency response while requiring substantially less computation time, on the

order of 1%, than needed for the fully numerical solution in Ref. 2. The normal mode approach also lends itself better to physical interpretation than purely numerical methods.

Finally, the present model permits quantitative assessment of the two principal assumptions made in Ref. 6 on the basis of the resonator being close to cylindrical in shape. One assumption was that it is sufficient to include only the fundamental and second mode, and the other is that the eigenfunctions for a cylindrical resonator can be used to calculate the nonlinearity coefficients. The relative effect of these two approximations is discussed in the present paper.

II. LAGRANGE'S EQUATIONS

In a previous paper,[6] the nonlinear equations for the dynamics of an ideal gas in a one-dimensional resonator with rigid walls were presented in terms of Lagrange's equations. The kinetic energy K and potential energy U of the gas, given by Eqs. (60) and (70) in that paper, respectively, are

$$K = \frac{\rho_0}{2}\int_0^l (v_0 + \dot{\xi})^2 S(x)dx, \tag{1}$$

$$U = \frac{P_0}{\gamma - 1}\int_0^l \frac{S^\gamma(x)dx}{[(1 + \partial\xi/\partial x)S(x+\xi)]^{\gamma-1}}, \tag{2}$$

where x is the Lagrangian coordinate along the axis of the resonator, $S(x)$ is the cross-sectional area, l is the length of the resonator, $v_0(t)$ is the velocity at which the resonator is being shaken along the x axis, ρ_0 and P_0 are the ambient density and pressure of the gas inside, respectively, γ is the gas constant, $\xi(x,t)$ is the particle displacement of the gas, and overdots indicate time derivatives. Lagrange's equation in terms of particle displacement is

$$\frac{\partial}{\partial t}\left(\frac{\delta L}{\delta \dot{\xi}}\right) - \frac{\delta L}{\delta \xi} = 0, \tag{3}$$

where $L=K-U$ is the Lagrangian, with $\delta L/\delta\xi$ and $\delta L/\delta\dot{\xi}$ its functional derivatives. In terms of the Lagrangian density $\mathcal{L}$, where $L=\int_0^l \mathcal{L}dx$, Lagrange's equation becomes

$$\frac{\partial}{\partial t}\left(\frac{\partial \mathcal{L}}{\partial \dot{\xi}}\right) + \frac{\partial}{\partial x}\left(\frac{\partial \mathcal{L}}{\partial(\partial\xi/\partial x)}\right) - \frac{\partial \mathcal{L}}{\partial \xi} = 0. \tag{4}$$

Making use of Eqs. (1) and (2) one obtains, in explicit form,

$$\frac{\partial^2 \xi}{\partial t^2} = -\frac{P_0}{\rho_0}\frac{S(x+\xi)}{S(x)}\frac{\partial}{\partial x}\left(\frac{S(x)}{(1+\partial\xi/\partial x)S(x+\xi)}\right)^\gamma - \dot{v}_0, \tag{5}$$

where $-\dot{v}_0$ may be interpreted as an effective body force per unit mass. This dynamical equation in Lagrangian coordinates is fully equivalent to the lossless form of Eq. (28) in Ref. 2, which is expressed in terms of velocity potential and Eulerian coordinates. For a cylindrical resonator (S=const), Eq. (5) reduces to the classical exact result for plane waves:[9]

$$\frac{\partial^2 \xi}{\partial t^2} = \frac{c_0^2}{(1+\partial\xi/\partial x)^{\gamma+1}}\frac{\partial^2 \xi}{\partial x^2} - \dot{v}_0, \tag{6}$$

where $c_0^2=\gamma P_0/\rho_0$. In the linear approximation, Eq. (5) reduces to

$$\frac{\partial}{\partial x}\left(\frac{1}{S}\frac{\partial(S\xi)}{\partial x}\right) - \frac{1}{c_0^2}\frac{\partial^2 \xi}{\partial t^2} = \frac{\dot{v}_0}{c_0^2}, \tag{7}$$

which is the Webster horn equation expressed in terms of particle displacement.

At this point one may expand Eq. (5) in powers of ξ and its derivatives up to the desired order of approximation, and then seek solutions of the resulting equation. Instead, the procedure used previously[6] is followed here to obtain a set of coupled dynamical equations for the amplitudes $q_n(t)$ of the normal modes of the resonator, where n is the mode number. These amplitudes serve as the generalized coordinates of the system, in terms of which Lagrange's equation is

$$\frac{\partial}{\partial t}\left(\frac{\partial L}{\partial \dot{q}_n}\right) - \frac{\partial L}{\partial q_n} = 0. \tag{8}$$

III. NORMAL MODES

The sound field is expressed in terms of the normal mode expansion

$$\xi(x,t) = \sum_{n=1}^{\infty} q_n(t)\xi_n(x), \tag{9}$$

where the dimensionless quantities ξ_n are eigenfunctions of Eq. (7),

$$\frac{d}{dx}\left(\frac{1}{S}\frac{d(S\xi_n)}{dx}\right) = -\frac{\omega_n^2}{c_0^2}\xi_n, \tag{10}$$

ω_n are the natural angular frequencies, and the modal amplitudes q_n have units of displacement. Solutions for arbitrary functions $S(x)$ may be obtained numerically as follows. The dimensionless variables $w_n=\tilde{S}\xi_n$ and $p_n=\tilde{S}^{-1}d(\tilde{S}\xi_n)/d\tilde{x}$ are introduced, where $\tilde{x}=x/l$ is the dimensionless coordinate and $\tilde{S}(x)=S(x)/S_0$ is a dimensionless cross-sectional area, the normalization for which is chosen to be $S_0=\pi l^2$. Equation (10) then separates into the two coupled first-order equations

$$\frac{dw_n}{d\tilde{x}} = \tilde{S}p_n, \tag{11}$$

$$\frac{dp_n}{d\tilde{x}} = -\pi^2\tilde{\omega}_n^2\frac{w_n}{\tilde{S}}, \tag{12}$$

where $\tilde{\omega}_n=\omega_n/\omega_{01}$ is the dimensionless natural frequency and $\omega_{01}=\pi c_0/l$ is the fundamental natural frequency of a cylindrical resonator of length l. The physical sense of p_n is that it is proportional to the acoustic pressure of the nth mode. The boundary conditions on Eq. (11) are

$$w_n(0) = w_n(1) = 0. \tag{13}$$

Any nonzero value can be used for $p_n(0)$, as this condition affects only the amplitude of the solution. The amplitude of ξ_n is ultimately adjusted so that it satisfies the orthogonality relation

$$\int_0^1 \xi_m(\tilde{x})\xi_n(\tilde{x})\tilde{S}(\tilde{x})d\tilde{x} = \frac{1}{2}\delta_{mn}, \tag{14}$$

where δ_{mn} is the Kronecker delta. A standard shooting method was used to obtain solutions for $\tilde{\omega}_n$ and $\xi_n(\tilde{x})$. For reference, the solutions for a cylindrical resonator are $\tilde{\omega}_n = n$ and $\xi_n = \sin(n\pi\tilde{x})$.

Concerning the choice of $S_0 = \pi l^2$, the motivation is simply that for axisymmetric resonators the shape is most often defined in terms of the radius $r(x)$, in which case the dimensionless cross-sectional area is $\tilde{S}(\tilde{x}) = r^2(\tilde{x})/l^2$. In Secs. IV–VIII, however, the reference area S_0 remains arbitrary, the choice of which does not affect the results.

IV. DYNAMICAL EQUATIONS

The next step is to expand the kinetic and potential energies in terms of the modal amplitudes, and then use Eq. (8) to obtain the dynamical equation.

Expressed in terms of the modal amplitudes in Eq. (9), the kinetic energy given by Eq. (1) becomes[6]

$$K = \frac{m_0}{4}\sum_n (\dot{q}_n^2 + 2v_0 e_n \dot{q}_n) + \frac{m}{2}v_0^2, \tag{15}$$

where $m = \rho_0 \int_0^l S(x)dx$ is the mass of gas in the resonator, $m_0 = \rho_0 S_0 l$ is a reference mass, and

$$e_n = \frac{2}{S_0 l}\int_0^l \xi_n(x)S(x)dx. \tag{16}$$

The potential energy given by Eq. (2) is expressed in the form[6]

$$U = \frac{P_0}{\gamma - 1}\int_0^l (1+\alpha)^{-(\gamma-1)}S(x)dx, \tag{17}$$

where

$$\alpha = \frac{1}{S}\frac{\partial}{\partial x}\sum_{k=1}^{\infty}\frac{\xi^k}{k!}\frac{d^{k-1}S}{dx^{k-1}}. \tag{18}$$

The quantity $(1+\alpha)^{-(\gamma-1)}$ is now expanded up to fourth order in powers of α,

$$\begin{aligned}(1+\alpha)^{-(\gamma-1)} = 1 &- (\gamma-1)\alpha + \tfrac{1}{2}(\gamma-1)\gamma\alpha^2 \\ &- \tfrac{1}{6}(\gamma-1)\gamma(\gamma+1)\alpha^3 \\ &+ \tfrac{1}{24}(\gamma-1)\gamma(\gamma+1)(\gamma+2)\alpha^4,\end{aligned} \tag{19}$$

and the potential energy is expanded up to fourth order in powers of the modal amplitudes,

$$U = U_2 + U_3 + U_4. \tag{20}$$

The constant $U_0 = mc_0^2/[\gamma(\gamma-1)]$ is ignored because it does not contribute to the dynamics, the linear term U_1 is zero, the quadratic term is

$$U_2 = \frac{m_0}{4}\sum_n \omega_n^2 q_n^2, \tag{21}$$

and the terms at higher orders are expressed in the general forms

$$U_3 = \sum_{j,k,l} b_{jkl}q_j q_k q_l, \tag{22}$$

$$U_4 = \sum_{j,k,l,m} c_{jklm}q_j q_k q_l q_m, \tag{23}$$

where the coefficients b_{jkl} and c_{jklm} are symmetric with respect to any permutation of their indices.[6]

The coupled set of dynamical equations for the modal amplitudes is obtained by substituting Eqs. (15) and (20)–(23) in Eq. (8):

$$\ddot{q}_n + \omega_n^2 q_n = -e_n\dot{v}_0 - \frac{6}{m_0}\sum_{k,l} b_{nkl}q_k q_l - \frac{8}{m_0}\sum_{k,l,m} c_{nklm}q_k q_l q_m. \tag{24}$$

Since not all terms in the summations are relevant to a leading-order estimation of the amplitude response at the drive frequency, it is necessary at this point to consider the specific problem under consideration. An harmonic excitation of the form

$$\dot{v}_0 = A\cos\omega t \tag{25}$$

is assumed, where A is the drive amplitude of the acceleration applied to the resonator, and ω is close to the natural frequency ω_1 of the fundamental mode. For small drive amplitudes, the amplitude q_1 of the fundamental mode is small and taken to be of first order, and all other modal amplitudes are classified as second order. The objective is to calculate q_1 up to an accuracy of third order, which is the leading order at which change in q_1 occurs due to nonlinearity. Consequently, the only terms required at second order are the ones in the first summation of Eq. (24) that contain the coefficients b_{n11}. The only terms required at third order are the ones in the first summation that contain the coefficients b_{1k1} and b_{11k} for $k \neq 1$, and the one term in the second summation with coefficient c_{1111}.

With terms containing only the aforementioned coefficients taken into account, and since $b_{k11} = b_{1k1} = b_{11k}$, Eq. (24) reduces to

$$\begin{aligned}\ddot{q}_n + \omega_n^2 q_n = &-e_n\dot{v}_0 - \frac{6}{m_0}b_{n11}q_1^2 - \delta_{n1}\frac{12}{m_0}\sum_{k=2}^{N} b_{11k}q_1 q_k \\ &- \delta_{n1}\frac{8}{m_0}c_{1111}q_1^3,\end{aligned} \tag{26}$$

where δ_{ij} is the Kronecker delta, and the upper limit N on the summation was introduced to acknowledge the finite number of modes required for numerical calculations. This is the desired approximate dynamical equation for calculating the leading-order effect of nonlinearity on the acoustical response at the drive frequency. The drive frequency is assumed to be close to the natural frequency of the fundamental mode of the resonator, which is the case that is most often of interest in practice because this choice of drive frequency maximizes the response. The model equation is valid for resonators of arbitrary shape to the extent that the Webster horn equation is an accurate model in the linear approximation.

Ultimately of interest is an approximate equation that can be solved algebraically to determine the frequency response and thus the resonance frequency shift. The next step is to determine the nonlinearity coefficients in Eq. (26).

V. NONLINEARITY COEFFICIENTS

The coefficients $b_{k11}=b_{1k1}=b_{11k}$ of the quadratic terms in Eq. (26) are calculated from U_3. There is no contribution from the term in the expansion of Eq. (17) that is linear in α because its integral is just the summation in Eq. (18), which vanishes when evaluated at the upper and lower limits because $\xi=0$ at the ends of the resonator. The contribution from α^2 is

$$U_3^{(\alpha^2)} = \frac{\gamma P_0}{2} q_1^3 \int_0^l \frac{1}{S}\frac{d(S\xi_1)}{dx}\frac{d(S'\xi_1^2)}{dx}dx + \frac{\gamma P_0}{2}\sum_{k=2}^{N} q_1^2 q_k \int_0^l \left[\frac{2}{S}\frac{d(S\xi_1)}{dx}\frac{d(S'\xi_1\xi_k)}{dx} + \frac{1}{S}\frac{d(S\xi_k)}{dx}\frac{d(S'\xi_1^2)}{dx}\right]dx, \quad (27)$$

where $S'(x)=dS/dx$. Use of Eq. (10), the boundary conditions $\xi_k(0)=\xi_k(l)=0$, and integration by parts yields

$$U_3^{(\alpha^2)} = \frac{\gamma P_0}{2c_0^2}\left[\omega_1^2 q_1^3 \int_0^l \xi_1^3 S' dx + \sum_{k=2}^{N}(\omega_k^2+2\omega_1^2) q_1^2 q_k \int_0^l \xi_1^2 \xi_k S' dx\right]. \quad (28)$$

Likewise, the contribution from α^3 is

$$U_3^{(\alpha^3)} = -\frac{\gamma(\gamma+1)P_0}{2l^3}\left(\frac{1}{3}q_1^3\int_0^l p_1^3 S dx + \sum_{k=2}^{N} q_1^2 q_k \int_0^l p_1^2 p_k S dx\right). \quad (29)$$

Summing Eqs. (28) and (29) yields

$$U_3 = \frac{\pi}{2}\gamma P_0\left(\frac{1}{3}B_1 q_1^3 + \sum_{k=2}^{N} B_k q_k q_1^2\right), \quad (30)$$

and by comparison with Eq. (22)

$$b_{k11} = b_{1k1} = b_{11k} = \frac{\pi}{6}\gamma P_0 B_k, \quad (31)$$

where the coefficients

$$B_k = \pi^2(\tilde{\omega}_k^2 + 2\tilde{\omega}_1^2)\int_0^1 \xi_1^2\xi_k\tilde{S}'(\tilde{x})d\tilde{x} - (\gamma+1)\int_0^1 p_1^2 p_k \tilde{S}(\tilde{x})d\tilde{x} \quad (32)$$

are expressed in terms of the dimensionless quantities used in Sec. III for numerical calculation of the normal modes and their natural frequencies.

The same procedure is followed to calculate the coefficient c_{1111} of the cubic term in Eq. (26):

$$U_4^{(\alpha^2)} = \frac{\gamma P_0}{8} q_1^4 \int_0^l \left[\frac{1}{S}\frac{d(S\xi_1^2)}{dx}\right]^2 S(x)dx + \frac{\gamma P_0}{6} q_1^4 \int_0^l \left[\frac{1}{S}\frac{d(S\xi_1)}{dx}\right]\left[\frac{1}{S}\frac{d(S''\xi_1^3)}{dx}\right]S(x)dx, \quad (33)$$

$$U_4^{(\alpha^3)} = -\frac{1}{4}\gamma(\gamma+1)P_0 q_1^4 \times \int_0^l \left[\frac{1}{S}\frac{d(S\xi_1)}{dx}\right]^2\left[\frac{1}{S}\frac{d(S'\xi_1^2)}{dx}\right]S(x)dx, \quad (34)$$

$$U_4^{(\alpha^4)} = \frac{1}{24}\gamma(\gamma+1)(\gamma+2)P_0 q_1^4 \int_0^l \left[\frac{1}{S}\frac{d(S\xi_1)}{dx}\right]^4 S(x)dx. \quad (35)$$

Summing Eqs. (33)–(35) yields

$$U_4 = \frac{\pi\gamma P_0}{24l} C q_1^4, \quad (36)$$

and thus from Eq. (23)

$$c_{1111} = \frac{\pi\gamma P_0}{24l} C, \quad (37)$$

where

$$C = 3\int_0^1 \left\{2\xi_1\left[p_1 - \xi_1\frac{\tilde{S}'(\tilde{x})}{\tilde{S}(\tilde{x})}\right]\tilde{S}'(\tilde{x}) + \xi_1^2\tilde{S}''(\tilde{x})\right\}^2\frac{d\tilde{x}}{\tilde{S}(\tilde{x})} + 4\pi^2\tilde{\omega}_1^2\int_0^1 \xi_1^4\tilde{S}''(\tilde{x})d\tilde{x} - 12(\gamma+1)\pi^2\tilde{\omega}_1^2\int_0^1 \xi_1^3 p_1\tilde{S}'(\tilde{x})d\tilde{x} + (\gamma+1)(\gamma+2)\int_0^1 p_1^4\tilde{S}(\tilde{x})d\tilde{x}, \quad (38)$$

and $S''(x)=d^2S/dx^2$.

Evaluation of the integrals in Eqs. (32) and (38) for B_k and C can be accomplished together with numerical integration of Eqs. (11) and (12) for w_k and p_k by including the first-order differential equations for $dB_k/d\tilde{x}$ and $dC/d\tilde{x}$ [i.e., the integrands of Eqs. (32) and (38)] with initial conditions $B_k(0)=C(0)=0$.

VI. FREQUENCY SHIFT

In terms of the coefficients in Eqs. (32) and (38), Eq. (26) becomes

$$\ddot{q}_n + 2\delta_n\dot{q}_n + \omega_n^2 q_n = -e_n\dot{v}_0 - \frac{c_0^2}{l^3}B_n q_1^2 - \delta_{n1}\frac{2c_0^2}{l^3}\sum_{k=2}^{N} B_k q_1 q_k - \delta_{n1}\frac{c_0^2}{3l^4}C q_1^3. \quad (39)$$

The *ad hoc* loss factor δ_n introduced on the left-hand side corresponds to the sound diffusivity coefficients that appear

in Eqs. (28) and (33) of Ref. 2. The applied acceleration and modal amplitudes are now expressed as

$$\dot{v}_0 = \frac{A}{2} e^{i\omega t} + \text{c.c.}, \tag{40}$$

$$q_n = \frac{1}{2} \sum_{m=0}^{\infty} d_{nm} e^{im\omega t} + \text{c.c.}, \tag{41}$$

where c.c. denotes complex conjugate, and d_{nm} is the nominal complex peak amplitude of the particle displacement in mode n and at frequency $m\omega$. Substitution in Eq. (39) yields for d_{11}, the modal amplitude of interest,

$$(\omega_1^2 + i2\delta_1\omega - \omega^2) d_{11} = -e_1 A - \frac{c_0^2}{l^3} \sum_{k=1}^{N} B_k (2 d_{k0} d_{11} + d_{k2} d_{11}^*) - \frac{c_0^2}{4l^4} C |d_{11}|^2 d_{11}, \tag{42}$$

where

$$e_1 = 2 \int_0^1 \xi_1(\tilde{x}) \tilde{S}(\tilde{x}) d\tilde{x} = 2 \int_0^1 w_1(\tilde{x}) d\tilde{x}. \tag{43}$$

Recall that these equations follow from the assumption that the drive frequency is close to the natural frequency of the fundamental mode.

An approximate solution of Eq. (42) is obtained by recognizing that since d_{11} is assumed to be a small quantity of first order, and the resonance frequency shift is proportional to $|d_{11}|^2$, this shift is second order. Then, since in the absence of losses the linear form of Eq. (42) is approximately $2\omega_1(\omega_1 - \omega) d_{11} = -e_1 A$ for drive frequencies in the neighborhood of the resonance frequency, the drive amplitude A is third order. Within this framework, it is straightforward to obtain at second order the following explicit solutions of Eq. (39) for the modal amplitudes d_{k0} and d_{k2} of the dc component and second harmonic in the kth mode, respectively:

$$d_{k0} = -\frac{B_k c_0^2 |d_{11}|^2}{2l^3 \omega_k^2}, \tag{44}$$

$$d_{k2} = -\frac{B_k c_0^2 d_{11}^2}{2l^3 (\omega_k^2 + 4i\omega\delta_k - 4\omega^2)}. \tag{45}$$

Substitution of Eqs. (44) and (45) in Eq. (42) yields an equation for the amplitude d_{11} alone,

$$\Delta_{11}(\omega) d_{11} = -e_1 A + \left\{ \frac{c_0^4}{2l^6} \sum_{k=1}^{N} B_k^2 \left[\frac{2}{\omega_k^2} + \frac{1}{\Delta_{k2}(\omega)} \right] - \frac{c_0^2}{4l^4} C \right\} |d_{11}|^2 d_{11}, \tag{46}$$

where

$$\Delta_{nm}(\omega) = \omega_n^2 - (m\omega)^2 + i2\delta_n m\omega. \tag{47}$$

Finally, Eq. (46) is expressed in terms of the particle velocity $u_1 = i\omega d_{11}$:

$$\Delta_{11}(\omega) u_1 + F(\omega) |u_1|^2 u_1 = -i\omega e_1 A, \tag{48}$$

where

$$F(\omega) = \frac{c_0^2}{4l^4\omega^2} C - \frac{c_0^4}{2l^6\omega^2} \sum_{k=1}^{N} B_k^2 \left[\frac{2}{\omega_k^2} + \frac{1}{\Delta_{k2}(\omega)} \right]. \tag{49}$$

From Eq. (48) it can be seen that for drive frequencies $\omega \simeq \omega_1$, nonlinearity causes the resonance frequency to increase or decrease depending on whether the real part of $F(\omega_1)$ is positive or negative, respectively. The imaginary part of F is due to the damping coefficient in the expression for Δ_{k2}, which can be ignored for the purposes of this discussion. The contribution from the dc component (associated with the first term in the square brackets) to the sign of F is always negative, and the contribution from the cubic nonlinearity depends on the sign of C. All that is noted here about the sign of C is that for resonators having slowly varying shapes that are close to cylindrical, C is determined mainly by the last integral in Eq. (38), which is always positive.

The contribution from the second harmonic is determined through Δ_{k2} by the relation of the frequency of the second harmonic to the natural frequencies of the resonator. If it so happens that the frequency of the second harmonic lies very close to the natural frequency of a particular mode n, meaning that $2\omega_1 \simeq \omega_n$, then $\Delta_{n2}(\omega_1)$ is very small, and the term $k=n$ in the summation provides the dominant contribution to $F(\omega_1)$. In this special case, for $2\omega_1 > \omega_n$ the sign of $F(\omega_1)$ is positive and the resonance frequency is increased (resonance curves bend to the right), and for $2\omega_1 < \omega_n$ the resonance frequency is decreased (resonance curves bend to the left). The same result was obtained previously by perturbation for resonators that are close to cylindrical, in which case $\omega_n \simeq n\omega_1$ and whether the resonance curves bend right or left depends on whether $2\omega_1$ is greater or less than ω_2, respectively.[6] Further comparison with this previous work is presented in Sec. VIII.

As a practical matter, however, when a large response is desired at the drive frequency ω, the resonator is designed so that 2ω is not close to any natural frequency.[7] By suppressing generation of the second harmonic in this way, nonlinear losses are reduced. In this case one may replace $F(\omega)$ by $F(\omega_1)$ in Eq. (48), and, in particular, $\Delta_{k2}(\omega)$ by $\Delta_{k2}(\omega_1)$ in Eq. (49), because ω typically differs from the fundamental natural frequency ω_1 by only a few percent, and by design $2\omega_1$ is far from any natural frequency ω_k.

While in principle it is possible to solve Eq. (48) analytically as a real cubic equation in the quantity $|u_1|^2$, the solution is cumbersome and it is easier to obtain numerical solutions. For this purpose the dimensionless quantities $\tilde{\omega} = \omega/\omega_{01}$, $\tilde{\delta}_n = \delta_n/\omega_{01}$, $\tilde{\Delta}_{nm} = \Delta_{nm}/\omega_{01}^2$, $\tilde{F} = l^2 F$, $\tilde{A} = A/\omega_{01}^2 l$, and $\tilde{u}_1 = u_1/c_0$ are introduced to obtain

$$\tilde{\Delta}_{11}(\tilde{\omega}) \tilde{u}_1 + \tilde{F}(\tilde{\omega}) |\tilde{u}_1|^2 \tilde{u}_1 = -i\tilde{\omega}\pi e_1 \tilde{A}, \tag{50}$$

where

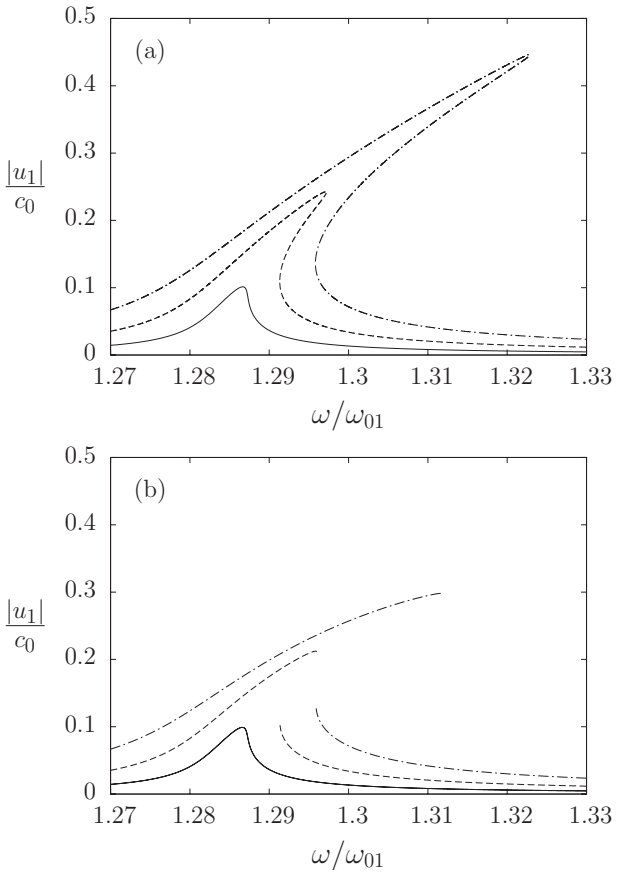


FIG. 1. Frequency-response curves for a cone resonator (a) calculated using Eq. (52) and (b) calculated using the fully nonlinear model in Ref. 2. The cone is defined by $\tilde{S}=(a_0+a_1\tilde{x})^2$, where $a_0=0.032\,94$ and $a_1=0.2680$, with $\gamma=1.2$ and $G=0.78\times10^{-2}$, and for drive amplitudes $\tilde{A}=10^{-4}$ (solid lines), $\tilde{A}=2.5\times10^{-4}$ (dashed lines), and $\tilde{A}=5\times10^{-4}$ (dot-dash lines) at location $\tilde{x}=0.5$.

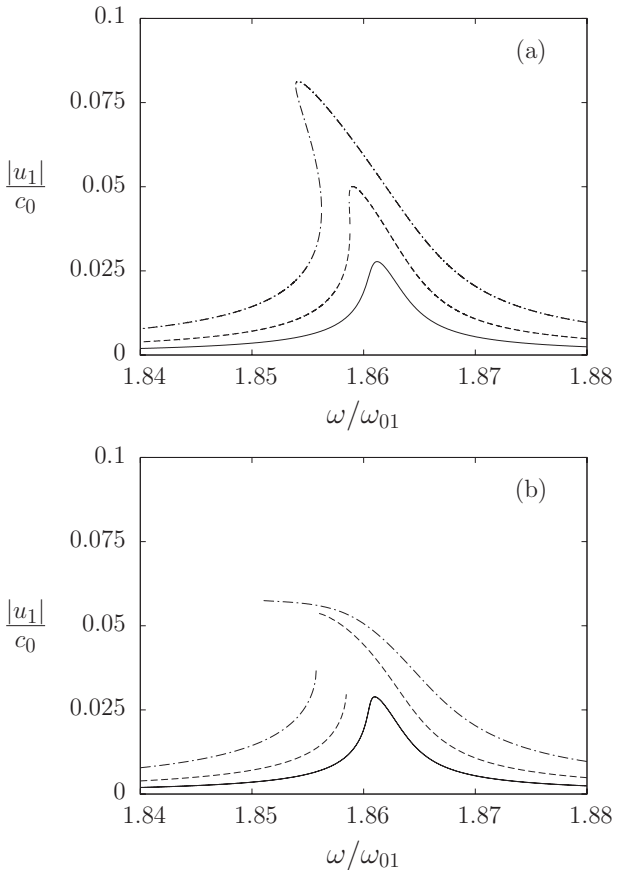


FIG. 2. Frequency-response curves for a bulb resonator (a) calculated using Eq. (52) and (b) calculated using the fully nonlinear model in Ref. 2. The bulb is defined by $\tilde{S}=(a_0+a_1\tilde{x}+a_2\tilde{x}^2+a_3\tilde{x}^3+a_4\tilde{x}^4)^2$, where $a_0=0.025$, $a_1=-0.2$, $a_2=1.15$, $a_3=0$, and $a_4=-0.9$, with $\gamma=1.2$ and $G=0.27\times10^{-2}$, and for drive amplitudes $\tilde{A}=2.5\times10^{-5}$ (solid lines), $\tilde{A}=5\times10^{-5}$ (dashed lines), and $\tilde{A}=10^{-4}$ (dot-dash lines) at location $\tilde{x}=0.5$.

$$\tilde{F}(\tilde{\omega})=\frac{C}{4\pi^4\tilde{\omega}^2}-\sum_{k=1}^{N}\frac{B_k^2}{2\pi^6\tilde{\omega}^2}\left[\frac{2}{\tilde{\omega}_k^2}+\frac{1}{\tilde{\Delta}_{k2}(\tilde{\omega})}\right]. \tag{51}$$

Roots of this cubic algebraic equation determine the desired frequency response $\tilde{u}_1(\tilde{\omega})$. Since it is more convenient numerically to find the root of a real equation, Eq. (50) is multiplied by its complex conjugate to obtain

$$|\tilde{F}(\tilde{\omega})x^2+\tilde{\Delta}_{11}(\tilde{\omega})|^2x^2=\pi^2e_1^2|\tilde{A}|^2\tilde{\omega}^2, \tag{52}$$

which is a real equation that can be solved for the real root $x=|\tilde{u}_1|$.

VII. RESONANCE CURVES

Numerical solutions of Eq. (52) are compared in Figs. 1 and 2 with the fully nonlinear numerical solutions obtained using the model equation and solution procedure described in Ref. 2. Figure 1 shows results for a cone resonator, and Fig. 2 shows results for a bulb resonator. The geometries of these resonators are defined by Eqs. (3) and (5) in the paper by Lawrenson *et al.*[1] The only difference here is that the small flare they introduced at one end of their bulb resonator to accommodate compressor valves was ignored, resulting in our change in the value of the coefficient a_1 that appears in their Eq. (5). All numerical values used for the computations are provided in the captions of Figs. 1 and 2, including the radial contour $r(x)$ for each resonator, where $S(x)=\pi r^2(x)$. The damping coefficient G that appears in the captions is defined as in Ref. 2. It is related to the damping coefficient used in the present work by $\tilde{\delta}_n=\tilde{\omega}_n^2G/2\pi$. The resonance curves in each figure are presented for the midpoint of each resonator, $\tilde{x}=0.5$. The gaps in the curves in Figs. 1(b) and 2(b) are associated with unstable states that cannot be computed with the numerical algorithm in Ref. 2.

The solutions of Eq. (52) shown in Figs. 1(a) and 2(a) are in reasonable quantitative agreement with the fully nonlinear solutions shown in Figs. 1(b) and 2(b) for both the lower and middle drive amplitudes used in each case, and there is still qualitative agreement at the higher drive amplitude. The mode shapes for the particle displacement (and therefore particle velocity) are shown in Fig. 3, where the solid line is for the fundamental mode in each case (possessing no nodes, not counting the end points), the dashed line is the second mode (one node), and the dot-dash line is the third mode (two nodes). The mode shape for the fundamental mode in the cone is relatively symmetric with its maximum located near the midpoint, whereas the maximum of the corresponding mode shape in the bulb is far off center, near $\tilde{x}$

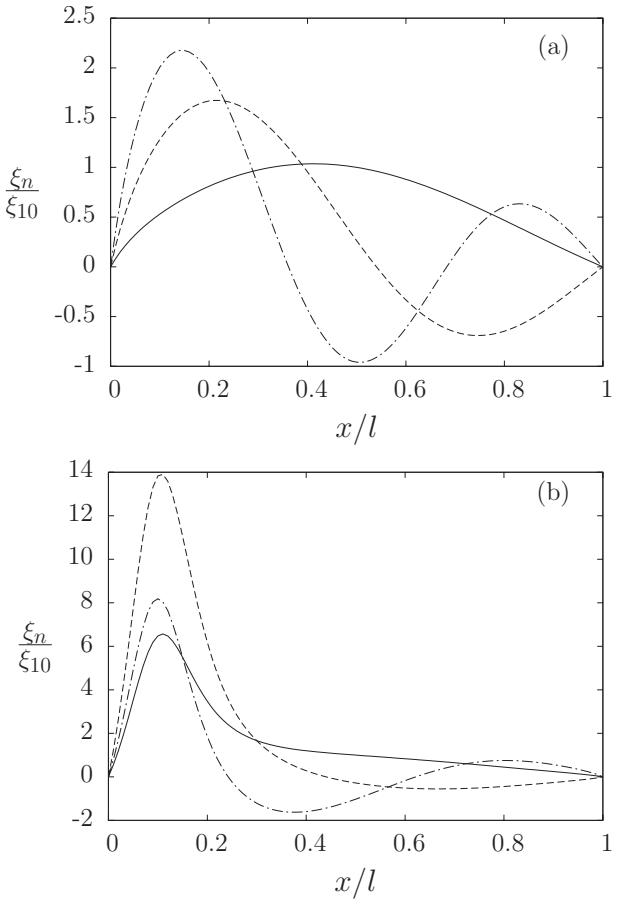


FIG. 3. Mode shapes of the particle displacement (and equivalently the particle velocity) for the fundamental (solid line), second (dashed line), and third (dot-dash line) modes, where the normalization ξ_{10} is the amplitude of the fundamental mode at $\tilde{x}=0.5$, for (a) a cone resonator and (b) a bulb resonator.

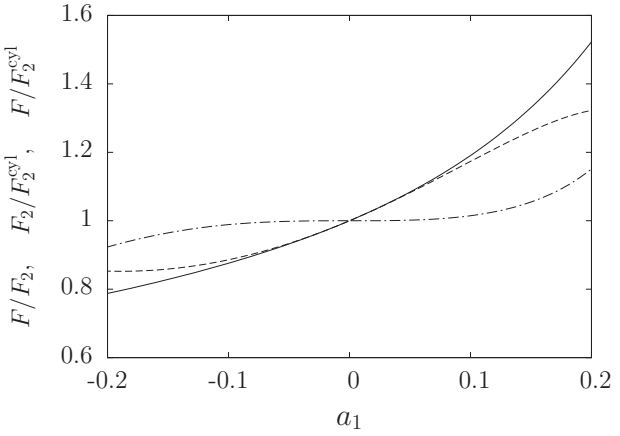


FIG. 4. Relative values of coefficients for the nonlinear frequency shift, $F/F_2^{\rm cyl}$ (solid line), F/F_2 (dot-dash line), and $F_2/F_2^{\rm cyl}$ (dashed line), as functions of the parameter a_1 for a resonator shape defined by Eq. (54): F was calculated using Eq. (49) with N sufficiently large that truncation error is insignificant; F_2 was calculated using Eq. (49) with $N=2$; $F_2^{\rm cyl}$ was calculated using Eq. (49) with $N=2$ and also with the eigenfunctions $\xi_n=\sin(n\pi x/l)$ for a cylindrical resonator used to calculate B_k and C.

$=0.1$, where the amplitude is approximately 6.5 times higher than at the midpoint. From Fig. 3 it is thus observed for the bulb that an acoustic Mach number $|u_1|/c_0=0.05$ in Fig. 2(b), which is calculated at $\tilde{x}=0.5$, corresponds to an acoustic Mach number of approximately 0.3 at $\tilde{x}\simeq 0.1$.

The relations of the first few natural frequencies of the cone resonator to the drive frequency $\omega\simeq\omega_1$ are $\omega_2/\omega_1=1.74$, $\omega_3/\omega_1=2.48$, and $\omega_4/\omega_1=3.24$, and for the bulb resonator they are $\omega_2/\omega_1=1.24$, $\omega_3/\omega_1=1.72$, and $\omega_4/\omega_1=2.28$. Thus for the cone the second harmonic lies midway between the natural frequencies of modes 2 and 3, whereas for the bulb it lies midway between the natural frequencies of modes 3 and 4. Because the second harmonic is not close to a natural frequency in either case, it is not clear in advance without evaluating Eq. (49) whether the resonance frequency is increased or decreased due to nonlinearity.

It was found that $N=40$ modes were sufficient for the calculations based on Eq. (52). Calculations with 100 modes produced results that are essentially the same.

VIII. COMPARISON WITH PERTURBATION APPROACH

The model developed in the present paper can be used to assess the validity of the assumptions made in the previous paper on resonators that are close to cylindrical in shape.[6] In that paper, a perturbation approach was used to obtain explicit expressions for the frequency response $u_1(\omega)$ and resonance frequency $\omega_{\rm res}$.

Two assumptions were made in Ref. 6 to estimate the resonance frequency shift. One is that only two modes need to be taken into account ($N=2$). The other, since the resonator shape was taken to be a small perturbation of a cylinder, is that the eigenfunctions $\xi_n=\sin(n\pi x/l)$ for a cylindrical resonator can be used to calculate the nonlinearity coefficients. With these assumptions one obtains

$$B_1=0,\quad B_2=-\tfrac{1}{2}\pi^3(\gamma+1),\quad C=\tfrac{3}{8}\pi^4(\gamma+1)(\gamma+2), \tag{53}$$

substitution of which into Eqs. (48) and (49) with $e_1=4/\pi$ and $\omega\simeq\omega_{01}$ recovers Eqs. (100) and (101) of Ref. 6.

For the numerical calculations in Ref. 6, the resonator shape was taken to be

$$S=S_0\exp[a_1\cos(2\pi x/l)], \tag{54}$$

where a_1 is the small parameter associated with the shape perturbation. Motivation for choosing this shape is discussed in Ref. 7. The effects of the two assumptions individually are quantified in Fig. 4 for a resonator defined by Eq. (54) by showing $F(\omega_1)$ evaluated as a function of the small perturbation parameter a_1 under different conditions. Losses are ignored ($\delta_n=0$) and therefore F is real. The function F without subscript or superscript in Fig. 4 is Eq. (49) evaluated without approximation, F_2 is Eq. (49) evaluated with only two modes ($N=2$), and $F_2^{\rm cyl}$ is Eq. (49) evaluated not only with just two modes, but with the eigenfunctions $\xi_n=\sin(n\pi x/l)$ for a cylindrical resonator used to calculate B_k and C.

A general observation is that neither the two-mode approximation nor the use of normal modes for a cylindrical resonator to calculate the nonlinearity coefficients changes the sign of F, or equivalently, the direction of the resonance frequency shift. The two approximations affect only the mag-

nitude of the frequency shift in either direction. The deviation of the curve for F/F_2 (dot-dash line in Fig. 4) from unity represents the effect of the two-mode approximation. Similarly, the deviation of the curve F_2/F_2^{cyl} (dashed line) from unity represents the effect of using normal modes for a cylindrical resonator to calculate the nonlinearity coefficients in addition to the two-mode approximation. Comparison of these two curves reveals that the use of cylindrical-resonator modes introduces a significantly greater error than does the two-mode approximation. The curve for F/F_2^{cyl} (solid line) represents the total error due to both approximations used in Ref. 6.

Now consider the explicit expression for the resonance frequency shift derived in Ref. 6:

$$\frac{\omega_{\mathrm{res}}}{\omega_1} \simeq 1 + \frac{(\gamma+1)^2}{64}\frac{|u_1|_{\max}^2}{c_0^2}\left(\frac{\omega_{01}}{\omega_1}\right)^6\left[3\left(\frac{\gamma+2}{\gamma+1}\right)\left(\frac{\omega_1}{\omega_{01}}\right)^2 - 2\left(\frac{2\omega_1}{\omega_2}\right)^2 + \frac{1}{1-(\omega_2/2\omega_1)^2}\right], \tag{55}$$

where $|u_1|_{\max} \simeq 2A/\pi\delta_1$. The right-hand side can also be expressed as $1+F(\omega_1)|u_1|_{\max}^2/2\omega_1^2$, from which it can be seen qualitatively how Eq. (55) follows from Eq. (48). The first term in the square brackets of Eq. (55) is due to cubic nonlinearity and involves only the fundamental component, and the second is due to interaction of the dc component with the fundamental through the quadratic nonlinearity. The effect of cubic nonlinearity, which increases the resonance frequency, is approximately twice that of the quadratic nonlinearity involving the dc component, which decreases the resonance frequency. Therefore the net effect of these two contributions is to increase the resonance frequency.

The third term in the square brackets is due to interaction of the fundamental with the second harmonic through the quadratic nonlinearity. At leading order this term equals $-a_1^{-1}$, the magnitude of which is symmetric about $a_1=0$. Thus, with the net positive frequency shift produced by the second harmonic and cubic nonlinearity taken into account, the increase in the frequency shift for $a_1<0$ is greater than the decrease in the frequency shift for $a_1>0$. This is what is observed in Figs. 4(a) and 5(a) of Ref. 6. The curvature of the solid line for F/F_2^{cyl} in Fig. 4 of the present paper reveals that this asymmetry is reduced when the two-mode and cylindrical-resonator approximations are not employed. The fully nonlinear results in Figs. 4(b) and 5(b) of Ref. 6, which are more symmetric than their counterparts in the corresponding Figs. 4(a) and 5(a), support this conclusion.

We conclude with a few remarks on the physical significance of positive and negative values of a_1. Negative values of a_1 produce a resonator shape resembling an American football. Resonators having this shape were constructed and tested by MacroSonix Corporation, and they were found to have desirable properties for compressors. Positive values of a_1 produce a resonator with a constriction in the middle. The behavior of such resonators tends toward that of two coupled Helmholtz resonators connected neck to neck. In this case a very large particle velocity can be generated near the midpoint of the resonator.

IX. CONCLUSION

Figures 1 and 2 demonstrate that the normal mode theory presented here provides a relatively simple method for making preliminary qualitative estimates of the frequency response of acoustical resonators. The model can be used for any resonator shape for which the Webster horn equation provides an accurate model in the linear approximation. Accurate quantitative calculations for strong nonlinearity require fully numerical solutions in 2D (for axisymmetric resonators), in an appropriate curvilinear coordinate system dependent on resonator geometry, that require huge computational resources. Even development of such codes is a nontrivial matter. Only the 1D code used by MacroSonix Corporation has been described in the literature,[2] although the main calculations for design of their resonators were performed with the 2D code they developed. The 2D code was used primarily to estimate losses, taking into account the automatic control used to track the resonance frequency shift. For calculations of the resonance frequency shift, it was sufficient to use the 1D code.[2] Even so, computation time for the 1D code is of order 100 times longer than for the normal mode theory presented here. Also, the normal mode theory provides physical insight into shape changes required to produce a desired frequency response, information that is not immediately available from purely numerical solutions.

The price paid for any asymptotic method is that its accuracy cannot be known in advance for arbitrary finite values of the small perturbation parameters. In the absence of confirmation by direct numerical simulation or by a theory unconstrained by small parameters, one must settle for qualitative results. For the model presented here, 1D numerical calculations show that significant quantitative differences can occur for large drive amplitudes, especially the nonlinear saturation that is underestimated by the asymptotic solution. Saturation was found to be more pronounced for the bulb resonator (Fig. 2) than for the cone resonator (Fig. 1). One may speculate that the stronger tendency toward saturation in the bulb resonator is associated with the maximum acoustic Mach number being approximately 50% larger than in the cone resonator.

An alternative asymptotic theory in Eulerian coordinates has been developed by Mortell and Seymour.[8] Discussion of saturation is not possible because losses are not included in their model, but their results presented in their Fig. 3(c) for the frequency shift in a bulb resonator may be discussed qualitatively. They compare their results directly with fully nonlinear numerical results from the 1D code presented in Fig. 13 of Ref. 2. However, the bulb geometries in the two cases are substantially different and do not warrant comparison because the ratios of the maximum to the minimum cross-sectional areas (corresponding to the bulb and the neck, respectively) differ by an order of magnitude.[7] As a result, the second harmonic in Mortell and Seymour's[8] bulb lies between the natural frequencies of modes 2 and 3, whereas for the bulbs simulated in Ref. 2 it lies between the natural frequencies of modes 3 and 4 (see Fig. 12 of Ref. 2), as it also does for the bulb used for Fig. 2 in the present paper.

Another difference is that the relative frequency shift in the numerical results presented in Fig. 13 of Ref. 2 is of order 10^{-2}, but it is only of order 10^{-3} in the corresponding simulations presented by Mortell and Seymour.[8] When evaluated for the resonator shape considered by Mortell and Seymour,[8] the model developed in the present paper predicts a frequency shift of the same order (10^{-3}) and in the same direction (downward) as they obtained. However, when the 1D code is used to model the same resonator, the frequency shift is of the same order but it is in the opposite direction (upward). If the 1D code is then run with the number of harmonics used in the calculation limited to just 2 or 3 (to be consistent with the asymptotic methods used both in the present paper and by Mortell and Seymour[8]), the order of the frequency shift remains the same, and the direction is the same as predicted by the two asymptotic methods (downward). From these results it may be inferred that a frequency shift of order 10^{-3} is beyond the accuracy of asymptotic methods that account for only the lowest order nonlinear effects on the resonance frequency.

ACKNOWLEDGMENT

This work was supported by the Office of Naval Research.

[1]C. C. Lawrenson, B. Lipkens, T. S. Lucas, D. K. Perkins, and T. W. Van Doren, "Measurements of macrosonic standing waves in oscillating closed cavities," J. Acoust. Soc. Am. **104**, 623–636 (1998).

[2]Yu. A. Ilinskii, B. Lipkens, T. S. Lucas, T. W. Van Doren, and E. A. Zabolotskaya, "Nonlinear standing waves in an acoustical resonator," J. Acoust. Soc. Am. **104**, 2664–2674 (1998).

[3]W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes in Fortran 77*, 2nd ed. (Cambridge University Press, Cambridge, 1992).

[4]R. R. Erickson and B. T. Zinn, "Modelling of finite amplitude acoustic waves in closed cavities using the Galerkin method," J. Acoust. Soc. Am. **113**, 1863–1870 (2003).

[5]Y.-D. Chun and Y.-H. Kim, "Numerical analysis for nonlinear resonant oscillations of gas in axisymmetric closed tubes," J. Acoust. Soc. Am. **108**, 2765–2774 (2000).

[6]M. F. Hamilton, Yu. A. Ilinskii, and E. A. Zabolotskaya, "Linear and nonlinear frequency shifts in acoustical resonators with varying cross sections," J. Acoust. Soc. Am. **110**, 109–119 (2001).

[7]M. F. Hamilton, Yu. A. Ilinskii, and E. A. Zabolotskaya, "Reply to "Comment on 'Linear and nonlinear frequency shifts in acoustical resonators with varying cross sections' [J. Acoust. Soc. Am. **110**, 109–119 (2001)]" (L)," J. Acoust. Soc. Am. **124**, 3386–3389 (2008).

[8]M. P. Mortell and B. R. Seymour, "Nonlinear resonant oscillations in closed tubes of variable cross-section," J. Fluid Mech. **519**, 183–199 (2004).

[9]R. T. Beyer, *Nonlinear Acoustics* (Acoustical Society of America, New York, 1997), Sec. 3.1.

Acoustic microstreaming around an isolated encapsulated microbubble

Xiaozhou Liu[a)]
Key Laboratory of Modern Acoustics, Ministry of Education, Institute of Acoustics, Nanjing University, Nanjing, Jiangsu 210093, China

Junru Wu[b)]
Department of Physics, University of Vermont, Burlington, Vermont 05405, USA and Key Laboratory of Modern Acoustics, Ministry of Education, Institute of Acoustics, Nanjing University, Nanjing, Jiangsu 210093, China

(Received 23 January 2008; revised 6 December 2008; accepted 14 December 2008)

An analytical theory has been developed to calculate microstreaming velocity inside and outside an encapsulated microbubble (EMB) in a viscous liquid produced by its oscillations driven by an ultrasound field, taking account of two predominant modes of the EMB's motion: a monopole (pulsation) and a dipole (translational harmonic vibrations). Analytical expressions of radial as well as tangential stresses are derived near the shell of the EMB. Numerical calculations in parameter regimes applicable to sonoporation are presented. For the calculation the following parameters unless specified otherwise are used: $f=1$ MHz, $r_0=2$ μm, $\kappa=1.4$, $\rho_L=1000$ kg/m^3, $\rho_s=1100$ kg/m^3, $P_0=100$ kPa, $\mu_s=0.05$ Pa s, $\mu_L=0.001$ Pa s, $\sigma_1=0.04$ N/m, $\sigma_2=0.005$ N/m, and $G_s=15$ MPa. The calculated results show that the streaming velocity and stresses near an EMB are functions of the mechanical properties of shell and gas. Overall, the streaming velocity and stresses for an EMB are found to be greater than those for a similar size free bubble under the same ultrasound excitation. This finding is consistent with the existing theory of acoustic streaming of an oscillating bubble near a boundary given by Nyborg (1958) [J. Acoust. Soc. Am. **30**, 329–339].
 [DOI: 10.1121/1.3075552]

PACS number(s): 43.25.Nm, 43.25.Yw [OAS] Pages: 1319–1330

I. INTRODUCTION

Encapsulated microbubbles (EMBs) have been used not only as contrast agents in diagnostic ultrasonic imaging clinically but also in various therapeutic applications (Wu and Nyborg, 2006). One commercially available type of EMB is the Optison® (GE Healthcare, Princeton, NJ, USA); it is a FDA (Food and Drug Administration of the USA) approved ultrasound (US) imaging contrast agent and contains micron-size denatured albumin microspheres filled with octafluoropropane (an inert gas). The thickness of albumin shells is less than 30 nm. The microbubble concentration is $(5-8)\times10^8$/ml, and its size has a distribution with the mean radius of the microbubbles being 1–2.25 μm. The maximum radius is 16 μm, and 93% of the Optison® bubbles have radii less than 5 μm according to the specification provided by the manufacturer. In targeting drug and gene delivery *in vitro* and *in vivo* experiments, it has been demonstrated that EMBs excited by moderate intensity US can increase membrane permeability of nearby cells. These nearby cells remain viable while allowing the intra-cellular transport of therapeutic drugs and antibodies (Unger *et al.*, 2001, 2002; Miller, 2006; Tachibana and Tachibana, 2006; Wu, 2006). This process is now called reparable sonoporation (Wu and Nyborg, 2008).

It was determined (Bao *et al.*, 1997; Ward *et al.*, 1999, 2000) that the spatial peak acoustic pressure amplitude as low as 0.1–0.2 MPa of 1 or 2 MHz US assisted by EMBs' (Optison®) presence in a suspension of cells may be able to cause reparable sonoporation. Ward *et al.* (2000) determined that the interaction between oscillating EMBs and cells is a short-range one; the closer the EMB locates to a cell, the stronger the interaction. Tran *et al.* (2007) performed an *in vitro* experiment using the ruptured-patch-clamp whole-cell technique. They demonstrated the hyperpolarization of the membrane of a marked cell (mammary breast cancer cell line MDA-MB-231) during sonoporation (1 MHz, 0.15 MPa negative peak US; Sono Vue microbubbles). The hyperpolarization of a biological cell means an above-normal increase in the trans-membrane voltage. They concluded that US activated oscillations of EMBs modify the electrophysiologic cell activities by their "cellular massage" action and thus enhance the cell's permeability for macroparticle uptake. Cellular massage here may mean the actions on a cell through the moderate shear stress generated by nearby oscillating bubbles; this type of "massage" is presumably to be related to microstreaming around an EMB.

Known biological effects of US on a human are thermal and mechanical ones (NCRP, 2002). The former is due to absorption of US by tissue, and the latter is usually associated with the acoustic cavitation that is particularly important in the presence of EMBs. EMBs may become nuclei of acoustic cavitation, which is broadly defined as any acousti-

[a)]Author to whom correspondence should be addressed. Electronic mail: xzliu@nju.edu.cn
[b)]On sabbatical leave.

cally driven bubble activity. There are two types of acoustic cavitations: (1) inertial acoustic cavitation and (2) noninertial acoustic cavitation. Inertial and noninertial cavitations were formally called the transient and stable cavitations, respectively (Wu and Nyborg, 2008; NCRP, 2002; Flynn, 1964). If the acoustic intensity is sufficiently high, an EMB will first grow in size and then rapidly collapse when inertial cavitation takes place. If noninertial cavitation occurs, an EMB in a liquid is forced to oscillate with a relatively small to moderate displacement in the presence of an acoustic field. Generally speaking when applied US peak acoustic pressure is of the order of the magnitude 0.1–0.2 MPa during the frequency from 1 to 2 MHz, EMBs go through the noninertial cavitation. Accompanying the noninertial cavitation, relatively stable microstreaming usually occurs near an oscillating EMB (Gormley and Wu, 1998).

Microstreaming is a bulk nonperiodic motion of fluid; it is a nonlinear phenomenon, which is directly related to the second-order terms of the sound field (Nyborg, 1958, 1965; Wu and Du, 1997). Nyborg (1958, 1965) studied streaming field near a free gas bubble resting on a boundary. Wu and Du (1997) derived the microstreaming field for an isolated free gas bubble excited by an ultrasonic field under the condition that the radius of a bubble was much smaller than the US wavelength and showed that streaming was more vigorous inside than outside of the bubble. Marmottant and Hilgenfeldt (2003) experimentally showed that gentle free-bubble oscillations were sufficient to achieve rupture of lipid membranes. They believed that microstreaming generated by free-bubble oscillation was the main cause of the effect and calculated the microstreaming velocity field from Longuet-Higgins theory. It was also reported that the acoustic microstreaming occurred near a pulsating EMB of radius 10 μm in a 160 kHz standing wave field, which was observed by Gormley and Wu (1998). The experimental observed streaming patterns are shown in Fig. 1.

In this paper, we have extended our study to a microstreaming field for an isolated EMB in a viscous fluid excited by an ultrasonic field under the same condition: the equilibrium radius of an encapsulated bubble is much smaller than the wavelength of the sound wave. When US of angular frequency ω propagates in a viscous liquid, boundary layers form at the interface between the liquid and the shell as well as between shell and the core gas of the EMB. The boundary layer thickness δ is equal to $\sqrt{2\mu/\rho\omega}$. Associated with microstreaming, there are two types of shear stress in the boundary layer of an oscillatory EMB. The first one is called the alternating (ac) shear stress related with ac fluid flow, and the other is called the direct (dc) shear stress caused by nonoscillatory microstreaming; both take place in the boundary layers.

The mathematical procedure of this work is as follows: We first obtain solutions for the first-order flow (ac flow) velocity near an oscillating EMB. We then derive the second-order nonoscillatory (dc) flow, i.e., acoustic microstreaming. Based on the above two solutions, i.e., oscillatory and nonoscillatory flow velocities, we further calculate their associated stress of a vibrating EMB. Examples of numerical calculations are also given to show that the stress generated by both ac and dc flows in boundary layers may play roles in applications of sonoporation.

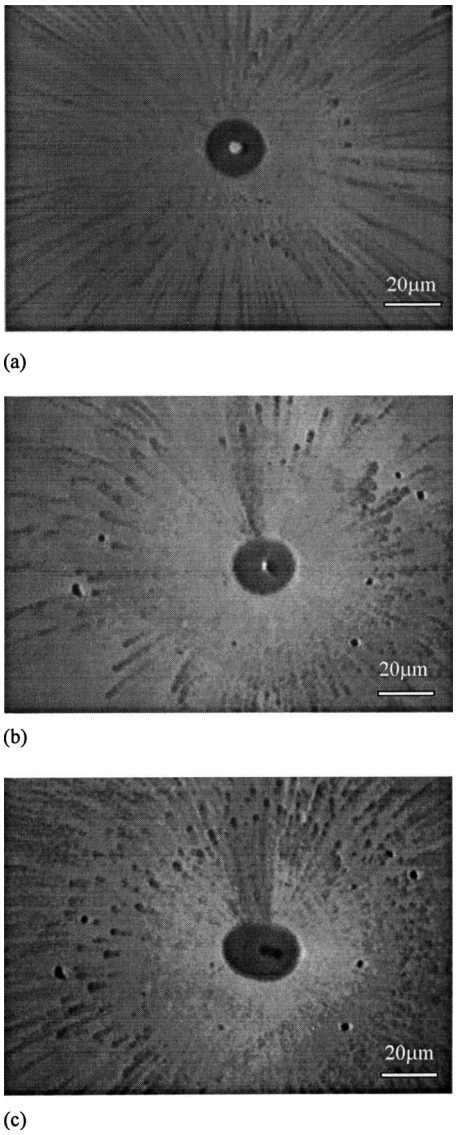


FIG. 1. Acoustic streaming pattern near an Albunex® sphere of 20 μm diameter in a 160 kHz standing wave field. (a) Acoustic streaming pattern generated by the sphere, which was pulsating (monopole motions). As acoustic pressure amplitude of the standing wave was increased, the sphere experienced both monopole and dipole oscillations, as shown in (b) and (c). The acoustic pressure amplitude of the standing wave was higher in (c) than in (b). It seemed that there were two jets leaving from the bubble in both (b) and (c); the spacing between the two jets were greater in (c) than in (b). This is the indication that the amplitude of translational (dipole) oscillation in (c) is greater than in (b). The streaming velocity was estimated to be 50–100 μm/s near the sphere.

II. THEORY

A. Dynamic motion of EMBs

The configuration of a reference frame is illustrated in Fig. 2, in which an isolated encapsulated microbubble is situated at the origin of the spherical coordinate (r, θ, ϕ) of a uniform and isotropic viscous liquid medium. A plane traveling acoustic wave propagates along the positive z direction. Due to viscosity, thin boundary layers are formed at liquid-

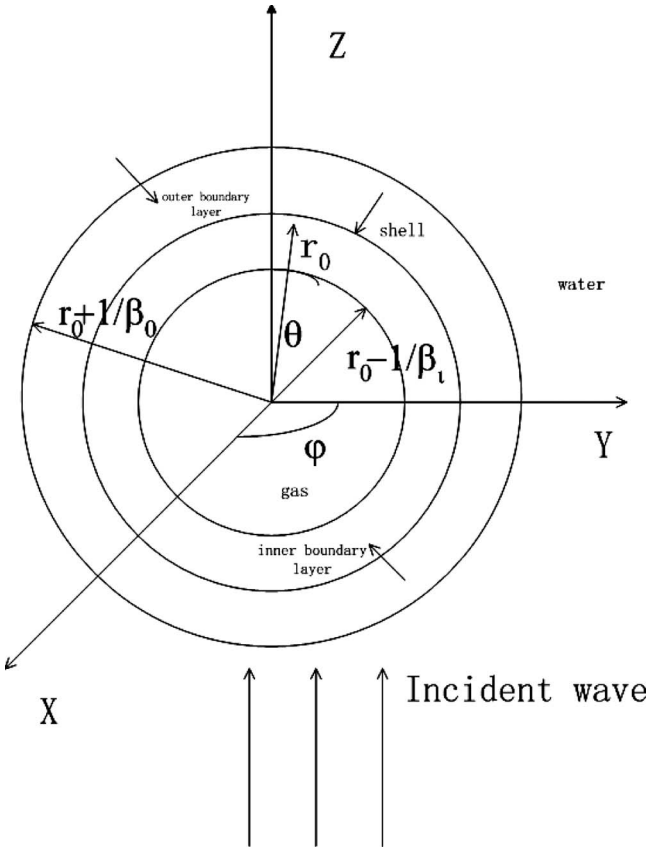


FIG. 2. The configuration of the system.

shell (outer boundary layer) and gas-shell (inner boundary layer) interfaces. The thicknesses of the inner and outer boundary layers are, respectively, of the order of β_i^{-1} and β_o^{-1} (the subscripts i and o denote inside and outside the bubble, respectively), where $\beta_i=\sqrt{\omega/2v_i}$, $\beta_0=\sqrt{\omega/2v_o}$, $v_i=\mu_i/\rho_{0i}$, and $v_0=\mu_0/\rho_{0o}$.

The main physical difference between a free gas bubble and an EMB such as Optison® is that the EMB has a human albumin layer (the shell of the EMB) that can make the bubble stable in liquids such as the body fluid and blood for a relatively long time [its half-life is 1.3 ± 0.69 min (mean $\pm$ SD)] (Henao *et al.*, 2006). Also the elastic shell changes the resonance response and behavior of the bubble. The following equation may be used to describe nonlinear oscillation of an EMB (Church, 1995):

$$r_1\dot{U}_1\left[1+\left(\frac{\rho_L-\rho_s}{\rho_s}\right)\frac{r_1}{r_2}\right]+U_1^2\left[\frac{3}{2}+\left(\frac{\rho_L-\rho_s}{\rho_s}\right)\times\left(\frac{4r_2^3-r_1^3}{2r_2^3}\right)\frac{r_1}{r_2}\right]=\frac{1}{\rho_s}\left[P_{G.\mathrm{eq}}\left(\frac{r_{01}}{r_1}\right)^{3\kappa}-p_\infty(t)-\frac{2\sigma_1}{r_1}-\frac{2\sigma_2}{r_2}-4\frac{U_1}{r_1}\left(\frac{V_s\mu_s+r_1^3\mu_L}{r_2^3}\right)-4\frac{V_sG_s}{r_2^3}\left(1-\frac{r_{e1}}{r_1}\right)\right],\tag{1}$$

where r_1 is the radius of the gas-filled cavity and U_1 is the radial velocity of inner interface. ρ_L and ρ_S are the density of liquid and shell, respectively, κ is polytropic exponent, $P_\infty(t)=P_0-P_A\sin\omega t$, P_0 is atmospheric pressure, P_A is the acoustic pressure amplitude, $P_{G,\mathrm{eq}}$ is the equilibrium pressure in the cavity, r_{e1} is the unstrained equilibrium position of inner interface, $V_s=r_{02}^3-r_{01}^2$, $r_{02}=r_{01}+r_s$, and r_s is the initial thickness of the shell, where the subscript 0 indicates conditions at time $t=0$. Further in Eq. (1), σ_1 and σ_2 are the tension at inner and outer interfaces, respectively, and G_s is the modulus of rigidity of shell. For megahertz frequencies the viscous damping dominates (Coakley and Nyborg, 1978), and therefore only this mechanism has been considered here. The flexural membrane forces of shell are negligible in our calculation.

Let $x=(r-r_{01})/r_{01}$; Eq. (1) can be rewritten as follows under the linear approximation (Church, 1995):

$$\ddot{x}+\delta_d\dot{x}+\omega_r^2x=P_A(\rho_sr_{01}^2\alpha)^{-1},\tag{2}$$

where δ_d and ω_r are damping constant and resonance angular frequency and given by (Church, 1995)

$$\delta_d=4\left[\frac{V_s\mu_s+r_{01}^3\mu_L}{r_{02}^3}\right](\rho_sr_{01}^2\alpha)^{-1},\tag{3}$$

$$\omega_r=\sqrt{(\rho_sr_{01}^2\alpha)^{-1}\left\{3\kappa P_0-\frac{2\sigma_1}{r_{01}}-\frac{2\sigma_2r_{01}^3}{r_{02}^4}+4\frac{V_sG_s}{R_{02}^3}\left[1+Z\left(1+\frac{3r_{01}^3}{r_{02}^3}\right)\right]\right\}}.\tag{4}$$

Here

$$\alpha=\left[1+\left(\frac{\rho_L-\rho_s}{\rho_s}\right)\frac{r_{01}}{r_{02}}\right]$$

and

$$Z=\left(\frac{2\sigma_1}{r_{01}}+\frac{2\sigma_2}{r_{02}}\right)\left(\frac{r_{02}^3}{V_s}\right)(4G_s)^{-1}.$$

Let $x=x_0\sin(\omega t+\phi)$; from Eq. (2), it is found that x_o and the phase are given by

$$x_0=\frac{P_A}{\rho_s\alpha\omega_0^2r_{01}^2}\chi(\Omega,\delta),\tag{5}$$

$$\phi=\arctan\left[\frac{\delta\Omega}{\Omega^2-1}\right],\tag{6}$$

and

$$\chi(\Omega,\delta)=[1/\sqrt{(1-\Omega^2)^2+\delta^2\Omega^2}]. \tag{7}$$

Here $\Omega=f/f_r=\omega/\omega_r$, where f_r and ω_r are the resonance frequency and resonance angular frequency of the EMB, respectively. At resonance, $\Omega=f/f_r=\omega/\omega_r=1$ and $\chi(\Omega,\delta)=1/\delta$. The resonance frequency $f_r=\omega_r/2\pi$ depends on properties of the gas, the liquid, and the shell, as given by Eq. (4).

B. The first-order flow (ac flow) velocity

The r and θ components of the relevant first-order velocity inside and outside a bubble can be written as follows (Wu and Du, 1997):

$$u_{1ri}=-ju_0A_0\frac{\chi k_i r}{3\lambda\sigma}+3ju_0A_1\cos\theta+6u_0A_1g_i\frac{e^{-jh_i(r_1-r_{01})}}{h_i r}\cos\theta,$$

$$u_{1\theta i}=-3ju_0A_1\sin\theta+3u_0A_1g_ije^{-jh_i(r_1-r_{01})}\sin\theta,$$

$$u_{1ro}=-ju_0\frac{A_0\chi k_o r}{3\sigma^2\lambda}+3ju_0A_1\cos\theta-6u_0A_1g_o\frac{e^{jh_o(r_2-r_{02})}}{h_o r}\cos\theta,$$

$$u_{1\theta o}=-3ju_0A_1\sin\theta+3ju_0A_1g_oe^{jh_o(r_2-r_{02})}\sin\theta, \tag{8}$$

where $A_0=1$, $A_1=-j$, $h_i=(1+j)\beta_i$, $h_o=(1+j)\beta_o$, and χ is given by Eq. (7).

The g_i and g_o are to be determined by the boundary conditions: at the shell, $r=r_{01}$, $r=r_{02}$,

$$u_{1\theta i}=u_{1\theta o}=0. \tag{9}$$

Therefore we can obtain

$$g_i=g_o=1. \tag{10}$$

We can further calculate the radial and tangential stresses due to the first-order flow velocity as follows:

$$\sigma_{1ri}=\mu_i\left(\frac{\partial u_{1ri}}{\partial r}\right),$$

$$\sigma_{1rio}=\mu_0\left(\frac{\partial u_{1ro}}{\partial r}\right),$$

$$\sigma_{1\theta i}=\mu_i\left(\frac{1}{r}\frac{\partial u_{1ri}}{\partial\theta}+\frac{\partial u_{1\theta i}}{\partial r}-\frac{u_{1\theta i}}{r}\right),$$

$$\sigma_{1\theta o}=\mu_o\left(\frac{1}{r}\frac{\partial u_{1ro}}{\partial\theta}+\frac{\partial u_{1\theta o}}{\partial r}-\frac{u_{1\theta o}}{r}\right). \tag{11}$$

Particularly, at $r=r_{01}$ and $r=r_{02}$, we obtain the radial and tangential stresses due to the first-order flow velocity,

$$\sigma_{1ri}=\mu_i\left[-ju_0A_0\frac{\chi k_i}{3\lambda\sigma}-6u_0A_1g_i\frac{1}{h_i r_{01}^2}\cos\theta-6ju_0A_1g_i\frac{1}{r_{01}}\cos\theta\right], \tag{12}$$

$$\sigma_{1r\theta i}=\mu_i\left[-\frac{6u_0A_1g_i\sin\theta}{h_i r_{01}}-3u_0A_1g_ij\sin\theta}{r_{01}}+3u_0A_1g_ih_i\sin\theta\right], \tag{13}$$

$$\sigma_{1ro}=\mu_o\left[-ju_0\frac{A_0\chi k_0}{3\sigma^2\lambda}+6u_0A_1g_o\frac{1}{h_o r_{02}^2}\cos\theta-6ju_0A_1g_o\frac{1}{r_{02}}\cos\theta\right], \tag{14}$$

$$\sigma_{1\theta o}=\mu_o\left[\frac{6u_0A_1g_o\sin\theta}{h_o r_{02}}-3ju_0A_1g_o\sin\theta}{r_{02}}-3h_0u_0A_1g_o\sin\theta\right]. \tag{15}$$

C. Acoustic streaming

The streaming velocities inside and outside an EMB satisfy the following equations (Wu and Du, 1997):

$$\frac{\partial^2 u_{2\theta i}}{\partial r^2}+\frac{2}{r}\frac{\partial u_{2\theta i}}{\partial r}\approx f_{\theta i}(r,\theta), \tag{16}$$

where $f_{\theta i}(r,\theta)=f_{\theta i0}+f_{\theta i1}$,

$$f_{\theta i0}=\frac{\chi u_0^2k_ir_1g_i\beta_ie^{-\beta_i(r_{01}-r_1)}}{2v_i\sigma\lambda}\times[\cos\beta_i(r_{01}-r_1)-\sin\beta_i(r_{01}-r_1)]\sin\theta, \tag{17}$$

$$f_{\theta i1}=\frac{9u_0^2g_i\beta_ie^{-\beta_i(r_{01}-r_1)}}{2v_i}\times[\cos\beta_i(r_{01}-r_1)+\sin\beta_i(r_{01}-r_1)]\cos\theta\sin\theta, \tag{18}$$

$$\frac{\partial^2 u_{2\theta o}}{\partial r^2}+\frac{2}{r}\frac{\partial u_{2\theta o}}{\partial r}\approx f_{\theta o}(r,\theta), \tag{19}$$

$$f_{\theta o}(r,\theta)=f_{\theta o0}+f_{\theta o1}, \tag{20}$$

$$f_{\theta o0}=\frac{-\chi u_0^2k_0r_2\beta_0e^{-\beta_0(r_{02}-r_2)}}{2v_0\sigma^2\lambda}\times[\cos\beta_0(r_2-r_{02})-\sin\beta_0(r_2-r_{02})]\sin\theta, \tag{21}$$

$$f_{\theta o1}=\frac{-9u_0^2g_0\beta_0e^{-\beta_0(r_2-r_{02})}}{2v_0}\times[\cos\beta_0(r_2-r_{02})+\sin\beta_0(r_2-r_{02})]\cos\theta\sin\theta. \tag{22}$$

According to Appendix B in Wu and Du (1997), the θ com-

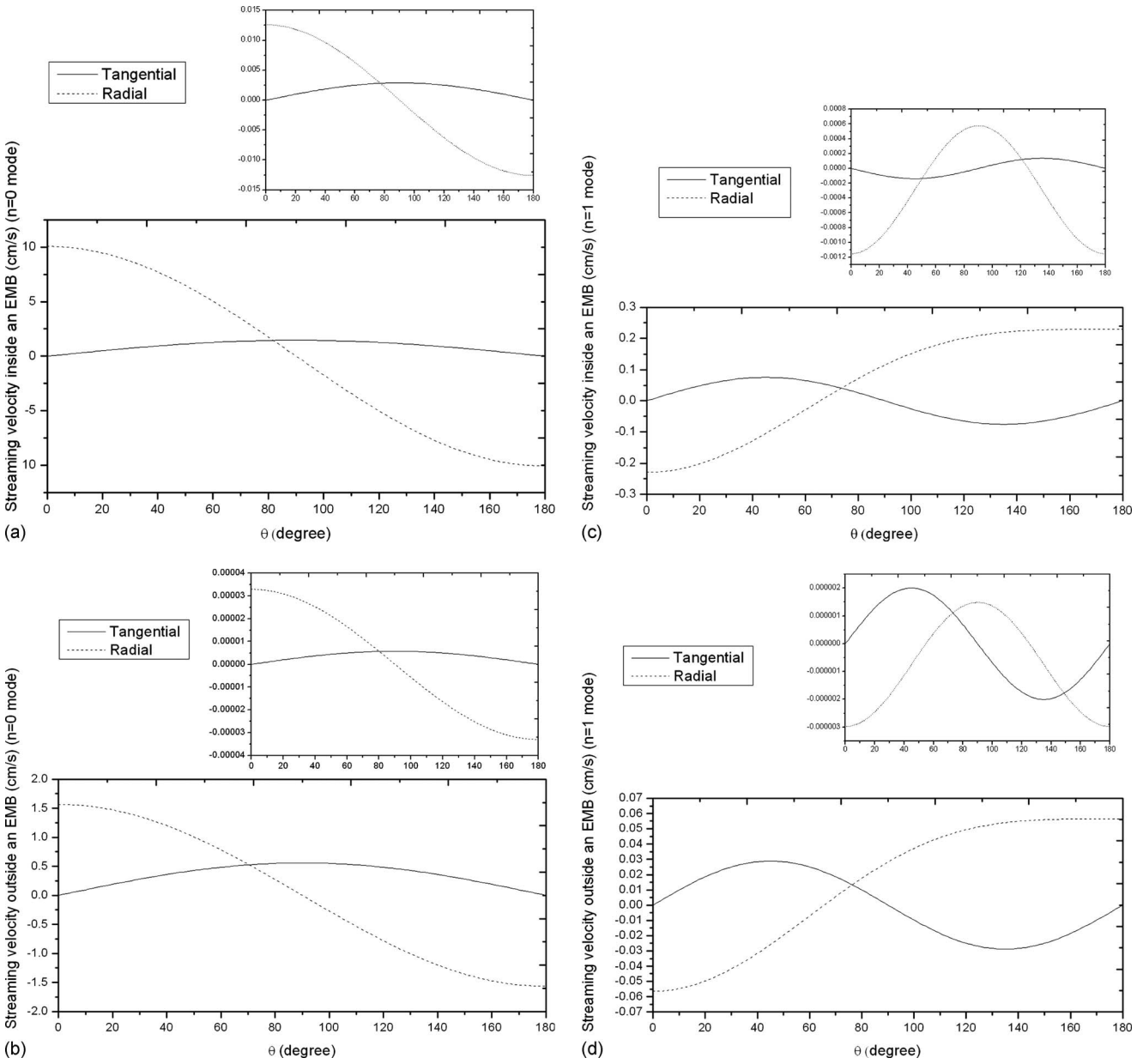


FIG. 3. Streaming velocity versus θ for an oscillating EMB and the inserts are the similar plot for an oscillating free bubble. (a) Streaming velocity inside ($r=r_0-0.5/\beta_i$) an EMB for $n=0$ mode, (b) streaming velocity outside ($r=r_0+0.5/\beta_o$) an EMB for $n=0$ mode, (c) streaming velocity inside ($r=r_0-0.5/\beta_i$) an EMB for $n=1$ mode, and (d) streaming velocity outside ($r=r_0+0.5/\beta_o$) an EMB for $n=1$ mode.

ponent of the streaming velocity inside the EMB can be calculated as

$$u_{2\theta_i} = u_{2\theta_{i0}} + u_{2\theta_{i1}}, \tag{23}$$

where

$$u_{2\theta_{i0}} = \frac{\chi u_0^2 k_i g_i}{2v_i \sigma \lambda} \sin\,\theta e^{-\beta_i(r_{01}-r_1)} r_{02}^2 \left(\frac{r_{01}}{r_1} - 1\right) \sin\,\beta_i(r_{01}-r_1) \tag{24}$$

for $n=0$ mode,

$$u_{2\theta_{i1}} = \frac{9u_0^2 g_i}{2v_i} \cos\,\theta \sin\,\theta \left(r_{01} - \frac{r_{01}^2}{r_1}\right) \times [e^{-\beta_i(r_{01}-r_1)} \cos\,\beta_i(r_{01}-r_1) - 1] \tag{25}$$

for $n=1$ mode, and for outside the EMB,

$$u_{2\theta_o} = u_{2\theta_{o0}} + u_{2\theta_{o1}}, \tag{26}$$

where

$$u_{2\theta o0} = \frac{\chi u_0^2 g_0 k_0}{2\lambda \sigma^2 v_0} \sin\,\theta e^{-\beta_0(r-r_0)} r_{02}^2 \left(1 - \frac{r_{02}}{r_2}\right) \sin\,\beta_0(r_2 - r_{02}) \tag{27}$$

for $n=0$ mode and

$$u_{2\theta o1} = \frac{9u_0^2 g_o}{2v_o} \cos\,\theta \sin\,\theta \left(r_{02} - \frac{r_{02}^2}{r_2}\right) \times [1 - e^{-\beta_{02}(r-r_0)} \cos\,\beta_0(r_2 - r_{02})] \tag{28}$$

for $n=1$ mode.

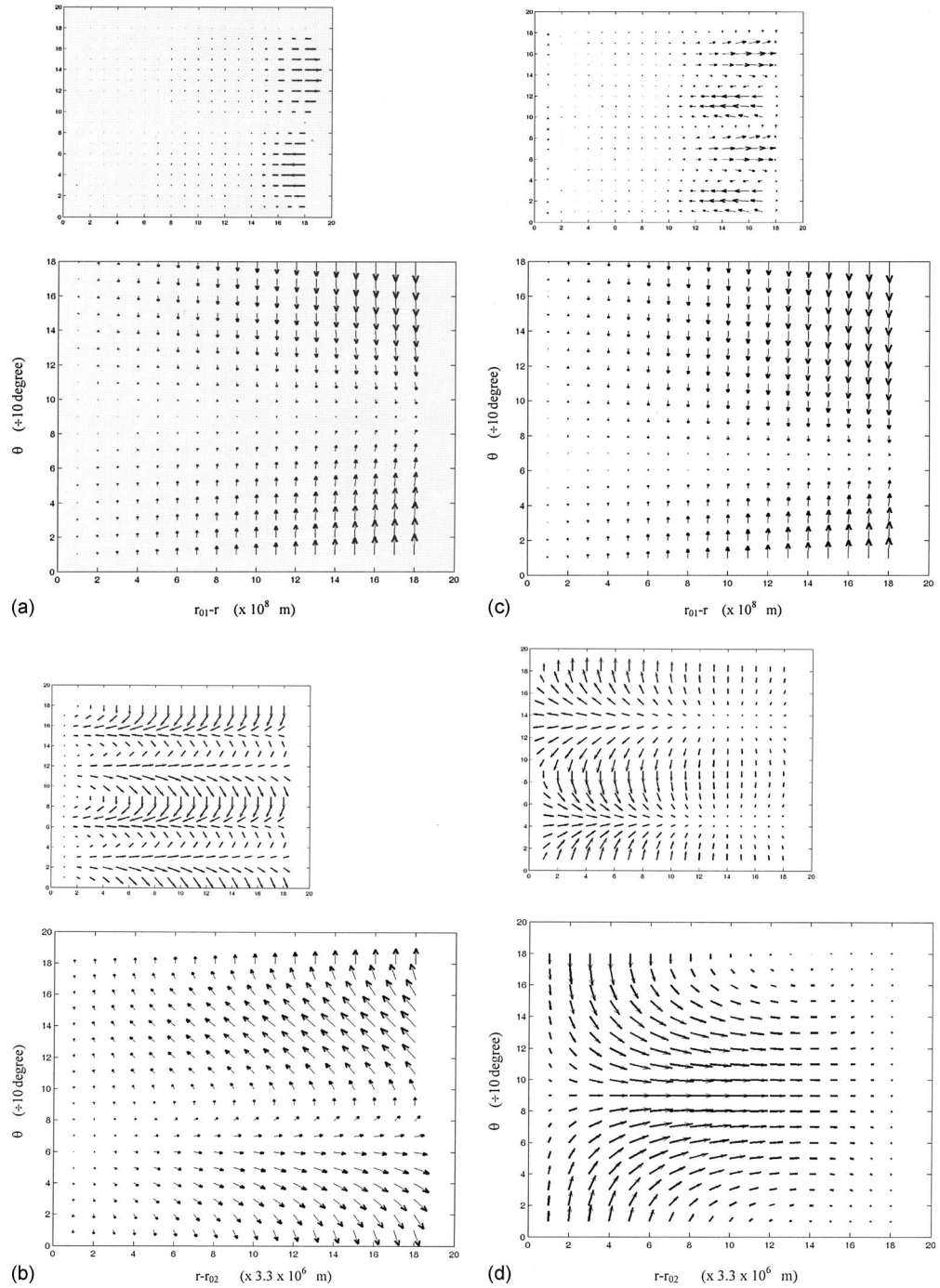


FIG. 4. Illustration of streaming lines generated by the $n=0$ mode: (a) $n=0$, inside the EMB, (b) $n=0$, outside the EMB, (c) $n=1$, inside the EMB, and (d) $n=1$, outside the EMB; the inserts are plots for a free bubble. The choice of magnification of the horizontal axis is arbitrary just to make the scale more visible.

The streaming velocity r-component inside and outside the EMB, u_{2ri} and u_{2ro}, can be evaluated by using $\nabla \cdot u_2 = 0$ and satisfying $u_{2ri} = u_{2ro} = 0$ at the boundaries $r = r_{01}$ and $r = r_{02}$. They are given by

$$u_{2r_i} = u_{2r_{i0}} + u_{2r_{i1}}, \tag{29}$$

where

$$u_{2r_{i0}} = \frac{\chi u_0^2 k_i g_i}{v_i \sigma \lambda} \frac{\cos\theta}{2\beta_i r_1^2} r_{01}^2 e^{-\beta_i (r_{01}-r_1)} \times [(r_{01}-r_1)\sin\beta_i (r_{01}-r_1) + (r_{01}-r_1) \times \cos\beta_i (r_{01}-r_1)] \tag{30}$$

for $n=0$ mode and

$$u_{2r_{i1}} = -\frac{9u_0^2 g_i r_{01}}{4v_i r_1^2}(2\cos\theta - \sin^2\theta)\left\{\frac{1}{\beta_i} e^{-\beta_i (r_{01}-r_1)} \times [(r_{01}-r_1)\sin\beta_i (r_{01}-r_1) - (r_{01}-r_1) \times \cos\beta_i (r_{01}-r_1)] - (r_{01}-r_1)^2\right\} \tag{31}$$

for $n=1$ mode.

The r-component streaming outside of the EMB, we have

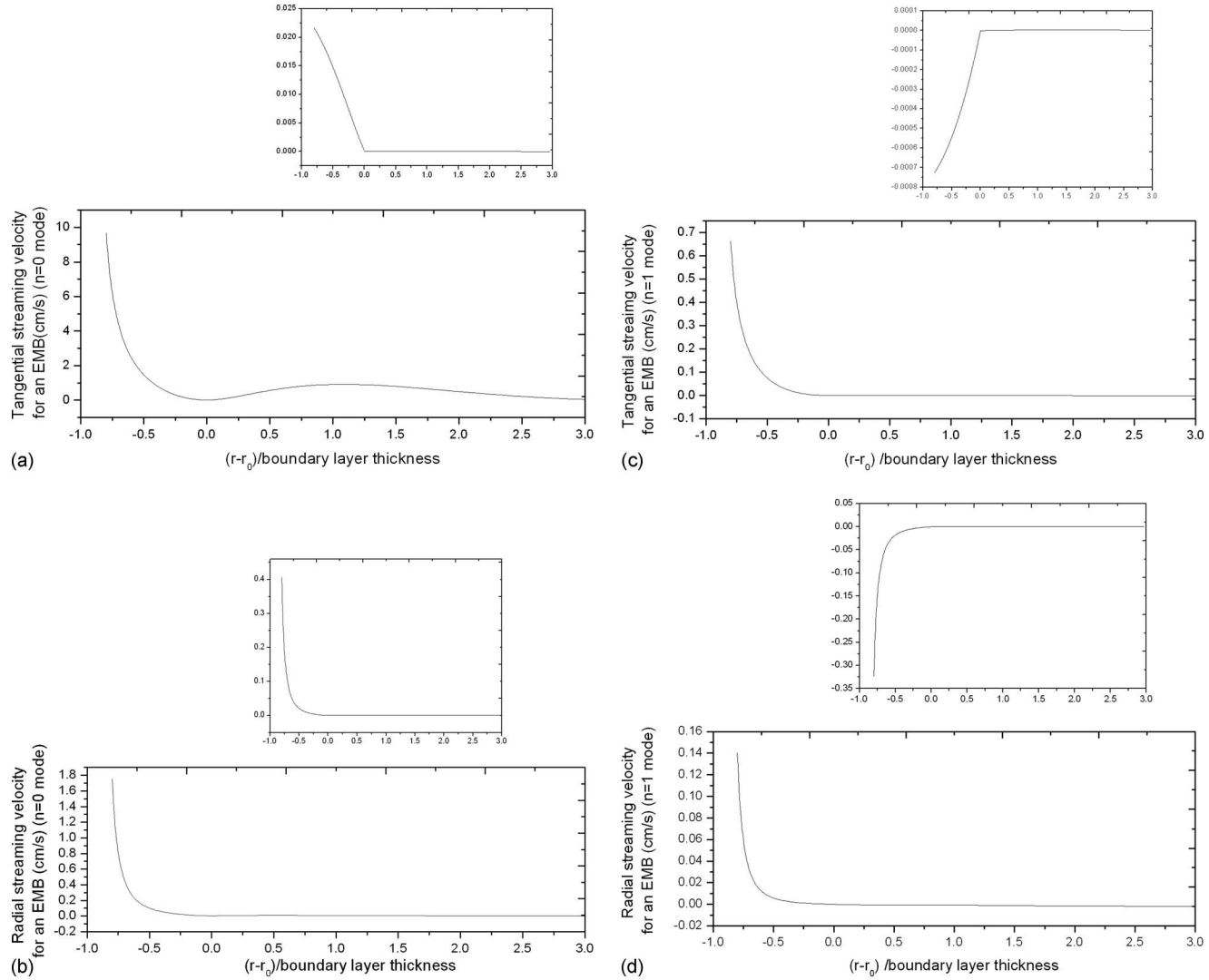


FIG. 5. Streaming velocity versus r at an angle where streaming velocity reaches maximum; the inserts are for a free bubble. (a) Tangential streaming velocity for $n=0$ mode, (b) radial streaming velocity for $n=0$ mode, (c) tangential streaming velocity for $n=1$ mode, and (d) radial streaming velocity for $n=1$ mode.

$$u_{2r_o} = u_{2r_{o0}} + u_{2r_{o1}}, \tag{32}$$

where

$$u_{2r_{o0}} = \frac{\chi u_0^2 g_0 k_0}{2\lambda\sigma^2 v_0 \beta_0 r_2^2} r_{02}^2 \cos\theta e^{-\beta_0(r_2-r_{02})} \times[(r_2-r_{02})\sin\beta_0(r_2-r_{02}) + (r_2-r_{02}) \times\cos\beta_0(r_2-r_{02})] \tag{33}$$

for $n=0$ mode and

$$u_{2r_{01}} = -\frac{9u_0^2 g_o r_{02}}{4v_o r_2^2}(2\cos\theta - \sin^2\theta)\left\{(r_2-r_{02})^2 - \frac{1}{\beta_0}e^{-\beta_0(r_2-r_{02})}[(r_2-r_{02})\sin\beta_0(r_2-r_{02}) - (r_2-r_{02})\cos\beta_0(r_2-r_{02})]\right\} \tag{34}$$

for $n=1$ mode.

Therefore the radial and tangential stresses due to streaming are as follows:

$$\sigma_{2ri} = \mu_i\left(\frac{\partial u_{2ri}}{\partial r}\right),$$

$$\sigma_{2rio} = \mu_0\left(\frac{\partial u_{2ro}}{\partial r}\right),$$

$$\sigma_{2\theta i} = \mu_i\left(\frac{1}{r}\frac{\partial u_{2ri}}{\partial\theta} + \frac{\partial u_{2\theta i}}{\partial r} - \frac{u_{2\theta i}}{r}\right),$$

$$\sigma_{2\theta o} = \mu_o\left(\frac{1}{r}\frac{\partial u_{2ro}}{\partial\theta} + \frac{\partial u_{2\theta o}}{\partial r} - \frac{u_{2\theta o}}{r}\right). \tag{35}$$

Thus,

$$\sigma_{2ri} = -\frac{\chi u_0^2 k_i g_i}{v\sigma\lambda}\frac{\cos\theta}{2\beta_i r_1^2} r_{01}^2 e^{-\beta_i(r_{01}-r_1)}[(r_{01}-r_1)\sin\beta_i(r_{01}-r_1) + (r_{01}-r_1)\cos\beta_i(r_{01}-r_1)] - \frac{9u_0^2 g_i r_{01}}{4v_i r_1^2}(2\cos\theta - \sin^2\theta)\left\{\frac{1}{\beta_i}e^{-\beta_i(r_{01}-r_1)}[(r_{01}-r_1)\sin\beta_i(r_{01}-r_1) - (r_{01}-r_1)\cos\beta_i(r_{01}-r_1)]\right.$$

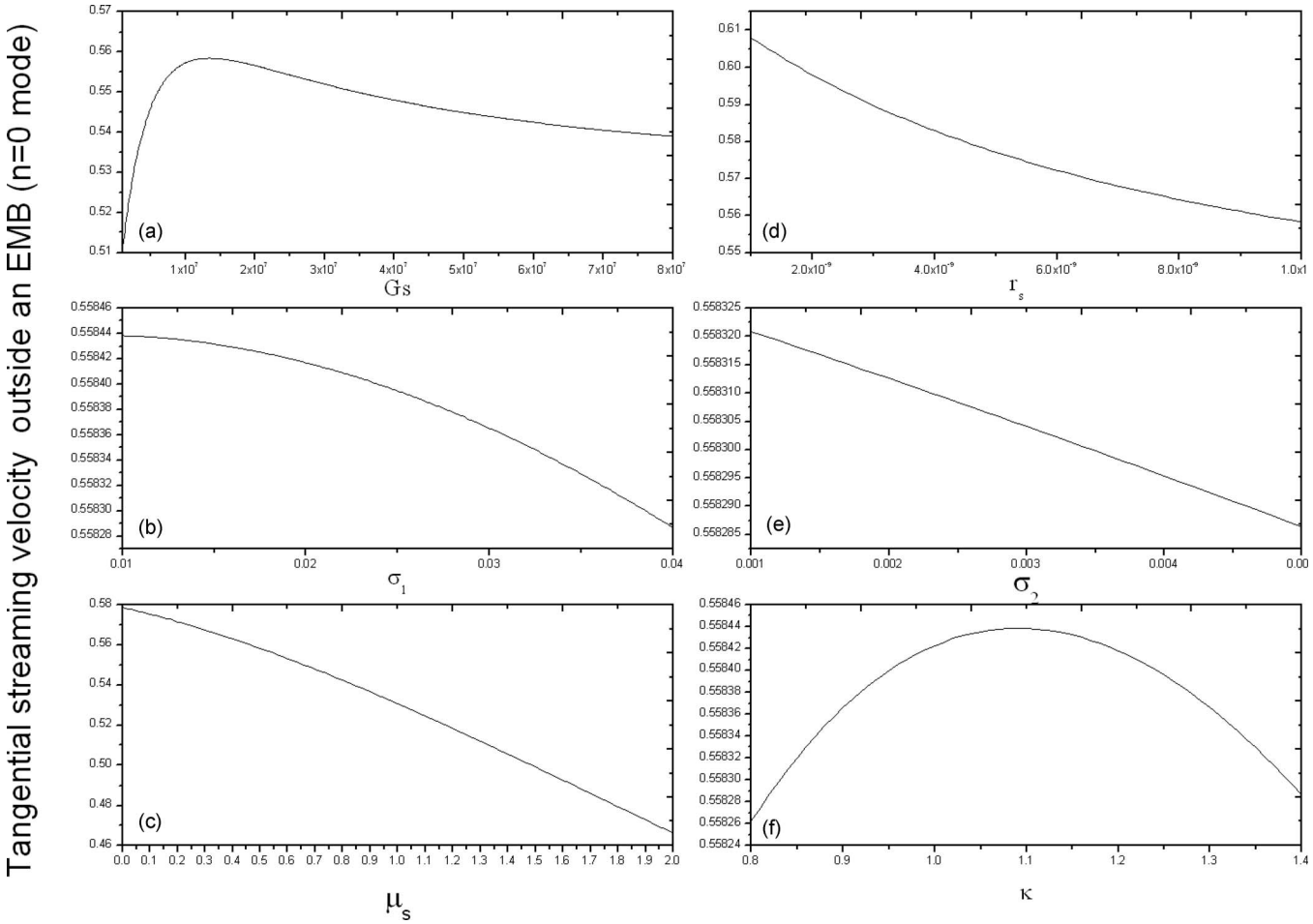


FIG. 6. Streaming velocity changes with the shell and gas mechanical parameters, (a) G_s, (b) r_s, (c) σ_1, (d) σ_2, (e) μ_s, and (f) κ.

$$- (r_{01} - r_1)^2 \Bigg\}, \tag{36}$$

$$\sigma_{2ro} = -\frac{\chi u_0^2 g_0 k_0}{2\lambda\sigma^2 v_0 \beta_0 r_2^2} r_{02}^2 \cos\theta e^{-\beta_0(r_2 - r_{02})}[(r_2 - r_{02}) \times \sin\beta_0(r_2 - r_{02}) + (r_2 - r_{02})\cos\beta_0(r_2 - r_{02})] - \frac{9u_0^2 g_o r_{02}}{4v_o r_2^2}(2\cos\theta - \sin^2\theta)\Bigg\{(r_2 - r_{02})^2 - \frac{1}{\beta_0} e^{-\beta_0(r_2 - r_{02})}[(r_2 - r_{02})\sin\beta_0(r_2 - r_{02}) - (r_2 - r_{02})\cos\beta_0(r_2 - r_{02})]\Bigg\}, \tag{37}$$

$$\sigma_{2\theta i} = \mu_i\Bigg\{\frac{\chi u_0^2 k_i g_i}{2v\sigma\lambda}\sin\theta r_{01}^2\Bigg[\beta e^{-\beta(r_{01} - r_1)}\left(\frac{r_{01}}{r_1} - 1\right) \times \sin\beta(r_{01} - r_1) + e^{-\beta(r_{01} - r_1)}\left(-\frac{r_{01}}{r_1^2}\right)\sin\beta(r_{01} - r_1) + e^{-\beta(r_{01} - r_1)}\left(\frac{r_{01}}{r_1} - 1\right)\beta\cos\beta(r_{01} - r_1)\Bigg]\Bigg\} + \frac{9u_0^2 g_i}{2v_i}\cos\theta\sin\theta\Bigg\{\left(\frac{r_{01}^2}{r_1^2}\right)[e^{-\beta(r_{01} - r_1)}\cos\beta(r_{01} - r_1) - 1] + \left(r_{01} - \frac{r_{01}^2}{r_1}\right)[\beta e^{-\beta(r_{01} - r_1)}\cos\beta(r_{01} - r_1) - \beta e^{-\beta(r_{01} - r_1)}\sin\beta(r_{01} - r_1)]\Bigg\}, \tag{38}$$

$$\sigma_{2\theta o} = \mu_0\Bigg\{\frac{-\chi u_0^2 g_0 k_0}{2\lambda\sigma^2 v_0}\sin\theta r_{02}^2\Bigg[-\beta e^{-\beta(r_2 - r_{02})}\left(1 - \frac{r_{02}}{r_2}\right) \times \sin\beta(r_2 - r_{02}) - e^{-\beta(r_2 - r_{02})}\left(\frac{r_{02}}{r_2^2}\right)\sin\beta(r_2 - r_{02}) + e^{-\beta(r_2 - r_{02})}\beta\left(1 - \frac{r_{02}}{r_2}\right)\cos\beta(r_2 - r_{02})\Bigg]\Bigg\} + \mu_0\Bigg\{-\frac{9u_0^2 g_o}{2v_o}\cos\theta\sin\theta\Bigg\{\frac{r_{02}^2}{r_2^2}[1 - e^{-\beta(r_2 - r_{02})} \times \cos\beta(r_2 - r_{02})] + \left(r_{02} - \frac{r_{02}^2}{r_2}\right) \times [-\beta e^{-\beta(r_2 - r_{02})}\cos\beta(r_2 - r_{02}) + \beta e^{-\beta(r_2 - r_{02})}\sin\beta(r_2 - r_{02})]\Bigg\}\Bigg\}. \tag{39}$$

It can be easily seen that the tangential stresses at $r = r_{01}$ and $r = r_{02}$ are zero.

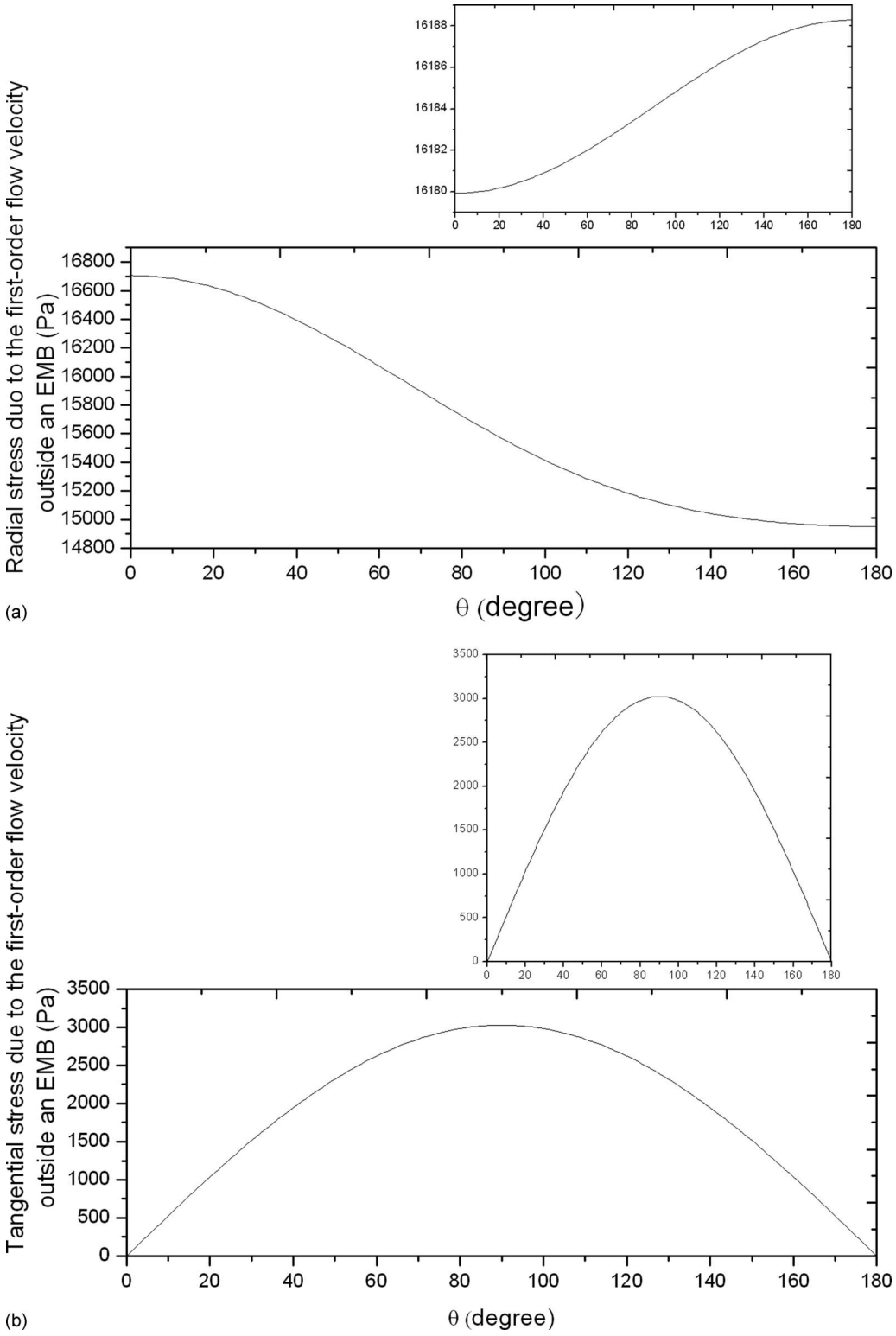


FIG. 7. Stress due to the first-order ac flow velocity at $r=r_{02}$ of an EMB: (a) radial stress and (b) tangential stress; the inserts are for a free bubble.

III. COMPUTATIONAL RESULTS

Based on Eqs. (24)–(39), we can calculate the streaming velocity inside and outside an EMB. As an example, we consider a case, $f=1$ MHz, $r_0=2$ μm, $\kappa=1.4$, $\rho_L=1000$ kg/m^3, $\rho_s=1100$ kg/m^3, $P_0=100$ kPa, $\mu_s=0.05$ Pa s, $\mu_L=0.001$ Pa s, $\sigma_1=0.04$ N/m, $\sigma_2=0.005$ N/m, and $G_s=15$ MPa (these parameters were found for Albunex®, Church, 1995). The thickness of the shell is 10 nm.

Figures 3(a)–3(d) are plots of tangential streaming velocity and radial streaming velocity versus θ for an EMB and a free bubble due to $n=0$ and $n=1$ modes, respectively. We observe that (1) the streaming velocity inside the EMB is greater than that of outside and (2) the streaming velocity corresponding to $n=0$ mode is greater than that of $n=1$ mode. The tangential streaming velocity of $n=0$ for both inside and outside the EMB reaches the maximum at θ

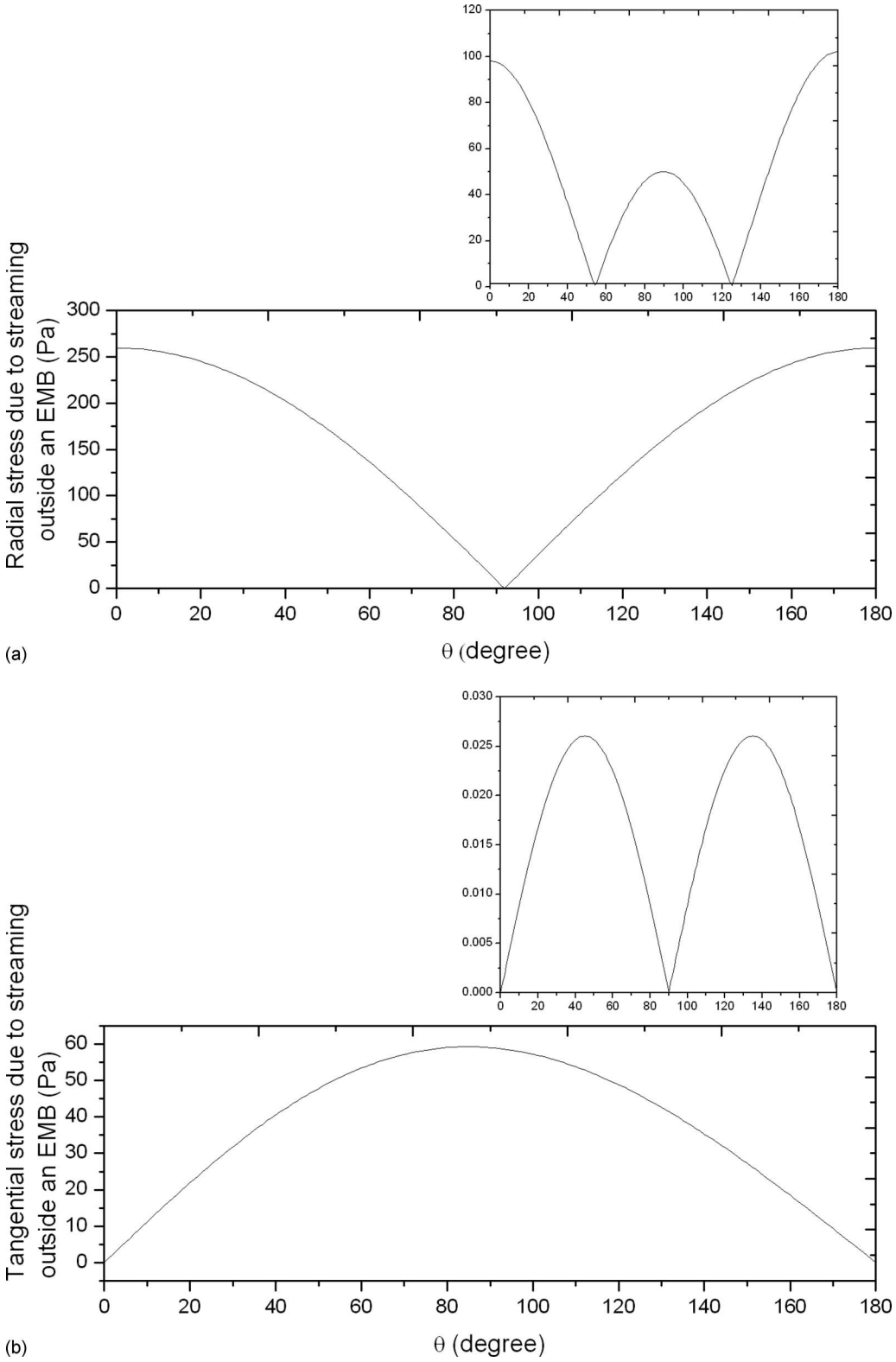


FIG. 8. Stress due to streaming of an EMB: (a) radial stress at $r=r_{02}$ and (b) tangential stress at $r=r_{02}+2/\beta_o$; the inserts are for a free bubble.

$=90°$, i.e., in the direction perpendicular to the wave propagation direction, while the radial streaming velocity of $n=0$ for both inside and outside the EMB reaches the maximum at $\theta=0°$, i.e., in the direction parallel to the wave propagation direction. For $n=1$ mode, both tangential streaming velocity and radial streaming velocity are of the same order of magnitude either inside or outside the EMB, but they are out of phase with each other. The magnitude of the tangential streaming velocity reaches the maximum at $\theta=45°$ and $\theta=135°$. For $n=0$ mode, the changes of streaming velocity for a free bubble are the same as those for an EMB; for $n=1$ mode, in the period of π, the radial streaming velocity changes sign twice for a free bubble while it changes sign once for an EMB.

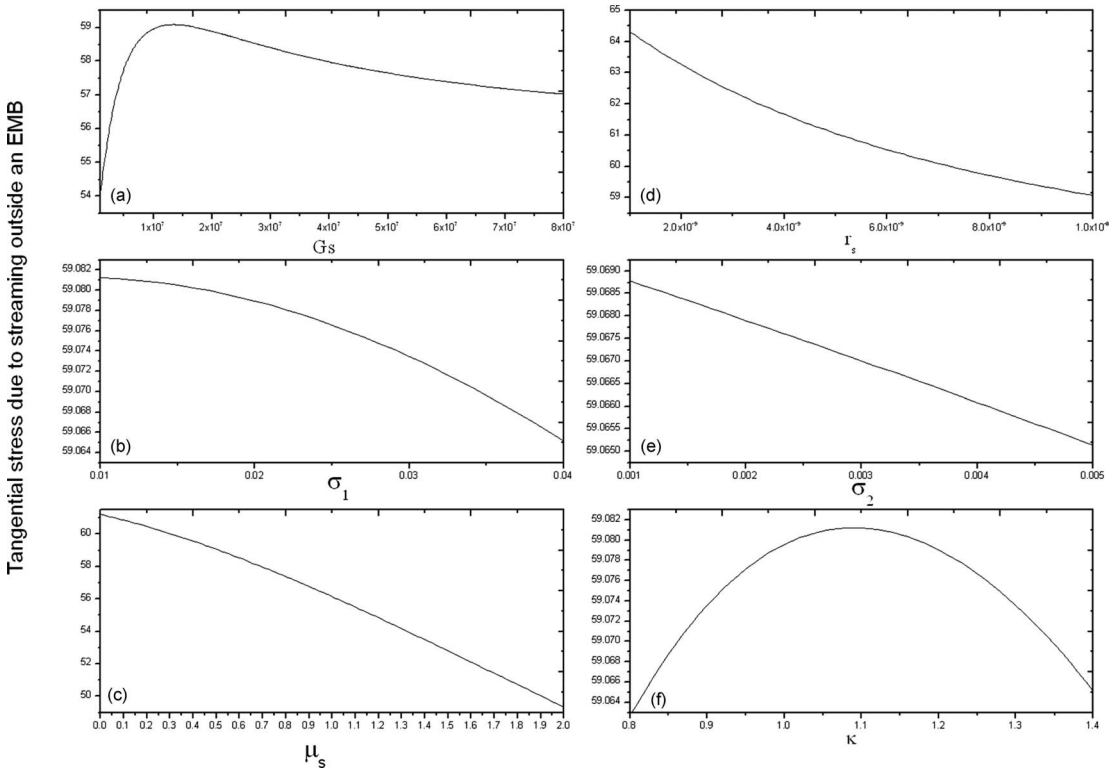


FIG. 9. Shear stress due to streaming changes with the shell and gas mechanical parameters: (a) G_s, (b) r_s, (c) σ_1, (d) σ_2, (e) μ_s, and (f) κ.

The streaming lines for $n=0$ and $n=1$ modes for EMB and free bubble are also plotted in Fig. 4, respectively, for illustration purposes. It can be observed the streaming line inside of the EMB changes direction at $\theta=90°$, while outside of the EMB the streaming line changes direction at $\theta=180°$ for $n=0$ mode. For $n=1$, the streaming line changes direction at $\theta=90°$ and $\theta=180°$ for both inside and outside of the EMB. Because of different boundary conditions, the streaming lines in an EMB are quite different from those in free bubble.

Figures 5(a)–5(d) are plots of tangential streaming velocity and radial velocity versus r for an EMB and a free bubble for $n=0$ and $n=1$ modes, respectively. The angle θ in each plot is chosen in such a way that tangential streaming or radial velocity reaches the local maximum, as shown in Figs. 3(a)–3(d). As expected from those figures, the following observations can be found: (1) As in a free gas bubble case, both tangential and radial streaming velocities inside an EMB are greater than outside an EMB. (2) Both the tangential and the radial streaming velocity are zero at $r=r_{01}$ and $r=r_{02}$. (3) Both tangential streaming velocity and radial velocity due to $n=0$ mode is greater than those of $n=1$ mode.

Figure 6 is a plot of the tangential streaming velocity of $n=0$ mode outside an EMB at $\theta=90°$ using the shell and gas mechanical property parameters as variables. We find that the streaming velocity is maximum when G_s is about 1.34 MPa and decreases with the increase in r_s, σ_1, σ_2, or μ_s. The relationship between screaming velocity and the polytropic exponent is interesting; it seems to reach the maximum when κ is near 1.09. However, the streaming velocity of $n=1$ mode is not affected by the shell and gas parameters significantly (not shown).

Figure 7 contains the radial and tangential stresses at outer interface ($r=r_{02}$) due the first-order ac velocity, respectively, for an EMB and a free bubble. The radial stress is much higher than the tangential stress. The calculation results show that the radial stress for an EMB is larger than that for a free bubble and the tangential stress for an EMB is about the same as that for a free bubble when they are driven by the same US.

Figure 8 shows plots of the radial stress at outer interface ($r=r_{02}$) and tangential stress near the outer interface ($r=r_{02}+2/\beta_0$) due to the streaming, respectively, for an EMB and a free bubble. The overall radial stress is greater than tangential stress. Near $\theta=90°$, tangential stress and radial stress reach maximum and minimum (zero), respectively, for an EMB. Three peaks of the radial stress and two peaks of the tangential stress for a free bubble are shown.

Figure 9 shows plots of tangential stress due to streaming outside an EMB at $\theta=90°$ that changes with the shell and gas mechanical parameters. The trend is similar as that of tangential streaming velocity of $n=0$ outside an EMB at $\theta=90°$.

IV. CONCLUSIONS AND DISCUSSIONS

We have extended a previously developed analytical theory to calculate microstreaming velocity inside and outside an isolated EMB in a viscous liquid produced by its scattered sound field, taking account of two predominant modes of the bubble's motion: a monopole (pulsation) and a dipole (translational harmonic vibrations). Based on the theory, analytical expressions of radial as well as tangential stresses are also derived near the shell of an EMB. The radial

stress generated by the first-order term (ac flow) is much higher than tangential or radial stress generated by streaming (dc flow). Since ac stress changes direction periodically according to the sound periodicity while dc stress does not change direction with time, it is the latter which plays predominant role to destroy the cell membrane in a lysis of red blood cells experiment (Rooney, 1970, 1972). In a reparable sonoporation experiment, when a cell is present inside of the outer boundary layer of an oscillating EMB wholly or partially, it may experience both radial and tangential stresses, therefore the so-called cellular massage action may take place (Tran *et al.*, 2007).

In the above-mentioned example of calculation, we used $f=1$ MHz, $r_0=2$ μm, and $P_0=0.1$ MPa since these parameters were used in several sonoporation *in vitro* experiments. For $r_0=2$ μm, 1 MHz is lower than the linear resonance frequency $f_r=3.02$ MHz for an EMB and the response of the EMB is relatively small; thus the linear acoustics may apply (Wu *et al.*, 2003). If we let $f=f_r$, the results calculated would be higher. For example, the maximal radial stress due to streaming of an EMB at $r=r_{02}$ is 489 Pa at $f=3.02$ MHz while it is 260 Pa at $f=1$ MHz, and the maximal tangential stress of an EMB at $r=r_{02}+2/\beta_o$ is 156 Pa at $f=3.02$ MHz while it is 59 Pa at $f=1$ MHz. On the other hand, the physical model used here is an ideal case. In an *in vitro* experimental condition, an EMB is not going to be isolated; it may have other EMBs and cells nearby; thus the damping coefficient should be higher and consequently the stress should be lower. In any case, the calculated stress values generated by microstreaming (dc flow) are in the range of the stress that can cause reparable sonoporation (Wu, 2002).

Shell and gas mechanical properties, especially G_s and the thickness of the shell, may influence the streaming and stress results. The variation of G_s and the thickness of the shell may change the resonance frequency of an EMB and also may influence the oscillation amplitude of an EMB. For the parameters used, compared with a free bubble, the oscillation amplitude of an EMB is smaller since the driving frequency used in our calculation is closer to the resonance frequency of a free bubble. However, it is interesting to note that the streaming velocity and the stress induced by the streaming for an EMB are actually greater than those for a free bubble; the enhancement here is caused by the boundary condition. The boundary condition for a free bubble is that the shear velocity and shear stress are continuous ($g_i=0.994$ and $g_o=-0.004\,69$) (Wu and Du, 1997). However, the boundary condition for an EMB is that the shear velocity and tangential shear stress are approximately zero, so $g_i=1$ and $g_o=1$. Our finding is consistent with the results of Nyborg (1958). In his paper, Nyborg (1958) showed that the presence of a solid boundary near an oscillating free bubble enhances the streaming velocity. The boundary conditions used have assumed that the shell of an EMB is more solidlike as far as the streaming is concerned. We believe that this assumption is a reasonable one. Future study will focus on the acoustic microstreaming considering the interaction among EMBs and the interaction between EMBs and cells or the blood vessel walls.

ACKNOWLEDGMENTS

This work is partially supported by the National Natural Science Foundation of China (Grant No. 10674066) and State Key Laboratory of Acoustics.

Bao, S., Thrall, B. D., and Miller, D. L. (**1997**). "Transfection of a reporter plasmid into cultured cells by sonoporation in vitro," Ultrasound Med. Biol. **23**, 953–959.

Coakley, W. T., and Nyborg, W. L. (**1978**). In *Cavitation: Dynamics of Gas Bubbles in Ultrasound: Its Applications in Medicine and Biology, Part I*, edited by F. J. Fry (Elsevier Scientific, Amsterdam).

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

Flynn, H. G. (**1964**). In *Physical Acoustics: Principles and Methods*, edited by W. P. Mason (Academic, New York), Vol. **1**, pp. 57–172.

Gormley, G., and Wu, J. (**1998**). "Observation of acoustic streaming near Albunex® spheres," J. Acoust. Soc. Am. **104**, 3115–3118.

Henao, E., Hodge, M., Felkai, D., McCollum, C., Noon, G., Lin, P., Lumsden, A., and Bush, R. (**2006**). "Contrast-enhanced Duplex surveillance after endovascular abdominal aortic aneurysm repair: Improved efficacy using a continuous infusion technique," J. Vasc. Surg. **43**, 259–264.

Marmottant, P., and Hilgenfeldt, S. (**2003**). "Controlled vesicle deformation and lysis by single oscillating bubbles," Nature (London) **423**, 153–156.

Miller, D. L. (**2006**). "Ultrasound-mediated gene therapy," in *Emerging Therapeutic Ultrasound*, edited by J. Wu and W. L. Nyborg (World Scientific, Hackensack, NJ).

NCRP (**2002**). *Exposure Criteria for Medical Diagnostic Ultrasound: II. Criteria Based on All Known Mechanism* (National Council on Radiation Protection and Measurements, Bethesda, MD).

Nyborg, W. L. (**1958**). "Acoustic streaming near a boundary," J. Acoust. Soc. Am. **30**, 329–339.

Nyborg, W. L. (**1965**). "Acoustic streaming," in *Physical Acoustics*, edited by W. Mason (Academic, New York), Vol. **IIB**, pp. 266–331.

Rooney, J. A. (**1970**). "Hemolysis near an ultrasonically pulsating gas bubble," Science **169**, 869–871.

Rooney, J. A. (**1972**). "Shear as a mechanism for sonically induced biological effects," J. Acoust. Soc. Am. **52**, 1718–1724.

Tachibana, K., and Tachibana, S. (**2006**). "Emerging technologies using ultrasound for drug delivery," in *Emerging Therapeutic Ultrasound*, edited by J. Wu and W. L. Nyborg (World Scientific, Hackensack, NJ).

Tran, T. A., Roger, S., Le Guennec, J. Y., Tranquart, F., and Bouakaz, A. (**2007**). "Effect of ultrasound activated microbubbles on the cell electrophysiological properties," Ultrasound Med. Biol. **33**, 158–163.

Unger, E. C., Hersh, E., Vannan, M., Matsunaga, T. O., and MaCreery, T., (**2001**). "Local drug and gene delivery through microbubbles," Prog. Cardiovasc. Dis. **44**, 45–54.

Unger, E. C., Matsunaga, T. O., and McCreery, T. (**2002**). "Therapeutic applications of microbubbles," Eur. J. Radiol. **42**, 160–168.

Ward, M., Wu, J., and Chiu, J. F. (**1999**). "Ultrasound-induced cell lysis and sonoporation enhanced by contrast agents," J. Acoust. Soc. Am. **105**, 2951–2957.

Ward, M., Wu, J., and Chiu, J. F. (**2000**). "Experimental study of the effects of Optison® concentration on sonoporation *in vitro*," Ultrasound Med. Biol. **26**, 1169–1175.

Wu, J. (**2002**). "Theoretical study on shear stress generated by microstreaming surrounding contrast agents attached to living cells," Ultrasound Med. Biol. **28**, 125–129.

Wu, J. (**2006**). in *Emerging Therapeutic Ultrasound*, edited by J. Wu and W. L. Nyborg (World Scientific, Hackensack, NJ).

Wu, J., and Du, G. (**1997**). "Streaming generated by a bubble in an ultrasound field," J. Acoust. Soc. Am. **101**, 1899–1907.

Wu, J., Pepe, J., and Dewitt, W., III (**2003**). "Nonlinear behaviors of contrast agents relative to diagnostic and therapeutic applications," Ultrasound Med. Biol. **29**, 555–562.

Wu, J., and Nyborg, W. L. (**2006**). *Emerging Therapeutic Ultrasound* (World Scientific, Hackensack, NJ).

Wu, J., and Nyborg, W. L. (**2008**). "Ultrasound, cavitation bubbles and their interaction with cells," Adv. Drug Delivery Rev. **60**, 1103–1116.

Model of coupled pulsation and translation of a gas bubble and rigid particle

Todd A. Hay,[a] Mark F. Hamilton, Yurii A. Ilinskii, and Evgenia A. Zabolotskaya
Applied Research Laboratories, The University of Texas at Austin, Austin, Texas 78713-8029

(Received 24 July 2008; revised 19 December 2008; accepted 9 January 2009)

A model of the interaction of a spherical gas bubble and a rigid spherical particle is derived as a coupled system of second-order differential equations using Lagrangian mechanics. The model accounts for pulsation and translation of the bubble as well as translation of the particle in an infinite, incompressible liquid. The model derived here is accurate to order R^5/d^5, where R is a characteristic radius and d is the separation distance between the bubble and particle. This order is the minimum accuracy required to account for the interaction of the bubble and particle. Dependence on the size and density of the particle is demonstrated through numerical integration of the dynamical equations for both the free and forced response of the system. Numerical results are presented for models accurate to orders higher than R^5/d^5 to demonstrate the consequences of truncating the equations at order R^5/d^5.
 [DOI: 10.1121/1.3077216]

PACS number(s): 43.25.Yw [ROC] Pages: 1331–1339

I. INTRODUCTION

Theoretical models describing the interaction of two or many gas bubbles in liquid have been in development for decades.[1–11] In the present work, the theoretical framework for one such model,[11] previously utilized to describe bubble-bubble interaction, is applied to the interaction between a pulsating spherical bubble and a rigid spherical particle. Both the bubble and particle are free to translate.

There currently exist several models for the interaction between bubbles and immovable solid objects. Coakley and Nyborg[12] derived an expression for the time-averaged force generated between a spherical bubble pulsating infinitesimally and a rigid, immovable sphere. Their expression is recovered as a special case of the present model. Theoretical, numerical, and experimental analyses of bubble behavior near rigid boundaries of infinite extent have shown that a bubble in pulsation translates toward plane, convex, and concave boundaries.[13–17] Other investigations have focused on bubble motion near rigid and deformable, but also immovable spheres.[18,19] These latter studies have concentrated primarily on modeling the violent aspherical collapse of the bubble and have employed boundary integral or other numerical techniques. In contrast, here we present an analytical model valid for low amplitude spherical oscillations, but for which both the bubble and particle are free to translate.

The present investigation was motivated initially by a desire to account for the interaction of cavitation clusters and kidney stone fragments produced during shock wave lithotripsy.[20] However, the extreme conditions of lithotripsy require consideration of liquid compressibility, cluster dynamics, and aspherical bubble deformation. While liquid compressibility corrections and dynamics of clusters containing arbitrary numbers of bubbles and particles are reported elsewhere,[21] the discussion here is limited for simplicity to a system with a single bubble and single particle under conditions where the liquid may be assumed incompressible. The assumption of incompressibility is valid if the wavelength of the acoustic excitation is much larger than the characteristic separation distance between the objects. In addition, the magnitudes of the radial and translational velocities in the system must be much less than the speed of sound in the liquid. Finally, it is assumed that the bubble remains spherical, and we thus ignore shape deformation requiring spherical harmonics of higher order than monopole (pulsation) and dipole (translation).

In comparison to bubble-bubble interaction, a higher-order model is required to describe bubble-particle interaction. The model equations must be accurate to at least order R^5/d^5, where R is a characteristic radius and d is the distance separating the bubble and particle. In contrast, the corresponding model equations for the coupled pulsation and translation of bubbles need only contain terms up to order R^2/d^2.[11] In practice, when the bubble and particle are in close proximity, a model with accuracy beyond R^5/d^5 is necessary to obtain numerical convergence. This issue will be addressed in detail.

This paper is structured as follows. The system energy and model equations are derived in Sec. II, and appropriate loss mechanisms are discussed in Sec. III. In Secs. IV and V, the effects of size and density of the particle are investigated for both free and forced responses of the system via numerical integration of the dynamical equations.

II. THEORY

The geometry of the problem is presented in Fig. 1. Coordinates with subscript 1 correspond to the gas bubble, and coordinates with subscript 2 correspond to the rigid particle. Both objects are assumed to be spherical at all times. Their positions are defined in relation to a fixed origin by the

[a]Electronic mail: haymaker@mail.utexas.edu

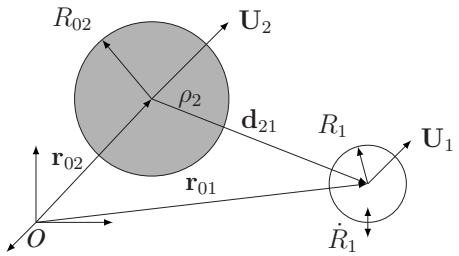


FIG. 1. Notation and geometry for bubble and particle with arbitrary translational motion.

vectors $\mathbf{r}_{0i}$, $i=1,2$. The instantaneous and equilibrium radii of the bubble are given by R_1 and R_{01}, respectively, while the radius of the particle is fixed at R_{02}. Motion is described by the radial velocity of the bubble $\dot{R}_1$ and translational velocities $\mathbf{U}_i=\dot{\mathbf{r}}_{0i}$, where dots over quantities indicate time derivatives. Lagrange's equations describing the dynamics of the system are

$$\frac{d}{dt}\left(\frac{\partial \mathcal{L}}{\partial \dot{R}_1}\right)=\frac{\partial \mathcal{L}}{\partial R_1}, \quad \frac{d}{dt}\left(\frac{\partial \mathcal{L}}{\partial \mathbf{U}_i}\right)=\frac{\partial \mathcal{L}}{\partial \mathbf{r}_{0i}}, \tag{1}$$

where $\mathcal{L}=\mathcal{K}-\mathcal{V}$ is the Lagrangian, $\mathcal{K}$ is the kinetic energy, and $\mathcal{V}$ is the potential energy of the system.

A. Potential energy

Potential energy is stored via compression or expansion of the bubble. In differential form, it is expressed as

$$d\mathcal{V}=(P_0-P_1)dV_1=4\pi(P_0-P_1)R_1^2dR_1,$$

where P_0 is atmospheric pressure and $V_1=\frac{4}{3}\pi R_1^3$ is the volume of the bubble. The pressure P_1 in the liquid just outside the bubble is taken to be

$$P_1=\left(P_0+\frac{2\sigma}{R_{01}}\right)\left(\frac{R_{01}}{R_1}\right)^{3\gamma}-\frac{2\sigma}{R_1}, \tag{2}$$

where γ is the ratio of specific heats and σ is surface tension. Additional effects that contribute to this pressure, such as heat transfer, gas diffusion, and condensation, are not considered here. Corrections for shear viscosity are discussed in Sec. III.

B. Kinetic energy

The motion of the liquid, the particle, and the gas inside the bubble all contribute to the kinetic energy of the system. However, the gas density is negligible compared to the liquid density, and kinetic energy associated with the motion of the gas may be neglected. The kinetic energy due to translational motion of the particle ($\mathcal{K}_{\text{part}}$) must be taken into account. The total kinetic energy is thus

$$\mathcal{K}=\mathcal{K}_{\text{part}}+\mathcal{K}_{\text{liq}}, \tag{3}$$

where $\mathcal{K}_{\text{liq}}$ is the kinetic energy of the surrounding liquid. The kinetic energy due to particle translation is simply

$$\mathcal{K}_{\text{part}}=\frac{1}{2}m_2U_2^2, \tag{4}$$

where m_2 is the mass of the particle.

The remaining component of the total kinetic energy, $\mathcal{K}_{\text{liq}}$, accounts for the motion of the inviscid incompressible liquid surrounding the bubble and particle. The liquid is assumed to be irrotational, such that its motion is described by a scalar velocity potential ϕ which satisfies Laplace's equation,

$$\nabla^2\phi=0. \tag{5}$$

The kinetic energy of the liquid is the integral

$$\mathcal{K}_{\text{liq}}=\frac{\rho}{2}\int_V|\nabla\phi|^2dV \tag{6}$$

over the volume surrounding the bubble and particle, where ρ is the liquid density. For a liquid at rest at infinity, Eq. (6) can be rewritten as[22]

$$\mathcal{K}_{\text{liq}}=-\frac{\rho}{2}\left(\int_{S_1}\phi(\mathbf{r}_1)\frac{\partial\phi(\mathbf{r}_1)}{\partial r_1}dS_1+\int_{S_2}\phi(\mathbf{r}_2)\frac{\partial\phi(\mathbf{r}_2)}{\partial r_2}dS_2\right), \tag{7}$$

where the surface S_i coincides with the bubble or particle wall, and $r_i=|\mathbf{r}_i|$ is the magnitude of the local coordinate vector $\mathbf{r}_i$ that defines position relative to the center of the bubble or particle.

Calculation of the kinetic energy thus requires knowledge of the velocity potential and its normal derivative on the surface of the bubble and particle. The velocity potential is expressed in the local coordinates of the bubble or particle. The normal derivative of the velocity potential on the surface of the bubble or particle is determined by the velocity boundary conditions

$$\left.\frac{\partial\phi}{\partial r_1}\right|_{S_1}=\dot{R}_1+\mathbf{U}_1\cdot\mathbf{n}_1, \tag{8}$$

$$\left.\frac{\partial\phi}{\partial r_2}\right|_{S_2}=\mathbf{U}_2\cdot\mathbf{n}_2, \tag{9}$$

where $\dot{R}_1$ accounts for pulsation of the bubble, $\mathbf{U}_i\cdot\mathbf{n}_i$ for translation, and $\mathbf{n}_i=\mathbf{r}_i/r_i$ is the unit vector in the direction of $\mathbf{r}_i$. Whereas Eq. (9) for the boundary condition on the surface of the particle is exact because there can be no shape deformation of a rigid particle, Eq. (8) ignores the possibility of shape deformation of the bubble wall by the absence of spherical harmonics beyond the monopole (pulsation) and dipole (translation).

There is no known expression for ϕ that satisfies the boundary condition on each sphere exactly. However, an approximate expression can be found that satisfies the boundary conditions to a desired order of R/d. We determine ϕ to the required accuracy by following the iterative approach used in Ref. 11. As shown in Sec. II C, it is necessary to obtain an expression for ϕ accurate to order R^5/d^5. This derivation is provided in the Appendix.

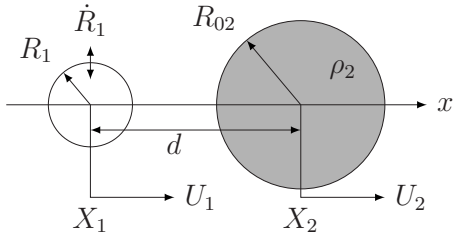


FIG. 2. Notation and geometry for bubble and particle with collinear translational motion.

The kinetic energy of the liquid is calculated by evaluating ϕ on the surfaces of the bubble and particle and substituting them into Eq. (7). The result of this calculation is

$$\mathcal{K}_{\text{liq}} = 2\pi\rho\left\{R_1^3\dot{R}_1^2 + \frac{1}{6}R_1^3U_1^2 + \frac{1}{6}R_2^3U_2^2 + \frac{R_2^3R_1^2\dot{R}_1}{d_{12}^2}(\mathbf{U}_2\cdot\mathbf{n}_{21}) + \frac{1}{2}\frac{R_1^3R_2^3}{d_{21}^3}[(\mathbf{U}_1\cdot\mathbf{U}_2) - 3(\mathbf{U}_1\cdot\mathbf{n}_{21})(\mathbf{U}_2\cdot\mathbf{n}_{21})] + \frac{1}{2}\frac{R_1^4R_2^3\dot{R}_1^2}{d_{21}^4} + \frac{1}{2}\frac{R_1^5R_2^3\dot{R}_1}{d_{21}^5}(\mathbf{U}_1\cdot\mathbf{n}_{12})\right\}. \tag{10}$$

At this point it is convenient to place the bubble and particle in the orientation shown in Fig. 2. The bubble and particle lie along the x axis such that the position vectors become $\mathbf{r}_{01}=X_1\mathbf{n}_x$ and $\mathbf{r}_{02}=X_2\mathbf{n}_x$, and the translational velocities are $\mathbf{U}_1=U_1\mathbf{n}_x$ and $\mathbf{U}_2=U_2\mathbf{n}_x$. Note that $\mathbf{d}_{12}=\mathbf{n}_x$, $\mathbf{d}_{21}=-\mathbf{n}_x$, and $d_{21}=d_{12}=d$. In these coordinates, the total kinetic energy is

$$\mathcal{K} = 2\pi\rho\left(R_1^3\dot{R}_1^2 + \frac{1}{6}R_1^3U_1^2 + \frac{1}{6}R_2^3U_2^2 + \frac{R_2^3R_1^2}{d^2}\dot{R}_1U_2 - \frac{R_1^3R_2^3}{d^3}U_1U_2 + \frac{R_1^4R_2^3}{2d^4}\dot{R}_1^2 + \frac{R_1^5R_2^3}{2d^5}\dot{R}_1U_1\right) + \frac{1}{2}m_2U_2^2. \tag{11}$$

C. Equations of motion

The equations of motion are obtained by substituting Eqs. (2) and (11) into Eqs. (1). The first of Eqs. (1) is the radial equation of motion for the bubble:

$$R_1\ddot{R}_1 + \frac{3}{2}\dot{R}_1^2 = \frac{P_1-P_0}{\rho} + \frac{1}{4}U_1^2 + \frac{3V_{02}}{8\pi}\left\{\frac{\dot{U}_2}{d^2} - \frac{U_2}{d^3}(U_1+2U_2) - \frac{R_1}{d^4}(R_1\ddot{R}_1 + 2\dot{R}_1^2) - \frac{R_1^2}{d^5}[R_1\dot{U}_1 + 4\dot{R}_1(U_1-U_2)]\right\}, \tag{12}$$

where P_1 is given by Eq. (2). The second of Eqs. (1) produces the translational equations of motion:

$$\dot{M}_1 = -F, \quad \dot{M}_2 = F, \tag{13}$$

$$M_1 = \frac{1}{2}\rho V_1U_1 - \frac{3}{2}\rho V_{02}\left(\frac{R_1^3}{d^3}U_2 - \frac{R_1^5}{d^5}\dot{R}_1\right), \tag{14}$$

$$M_2 = \left(\frac{1}{2}\rho+\rho_2\right)V_{02}U_2 - \frac{3}{2}\rho V_{02}\left(\frac{R_1^2}{d^2}\dot{R}_1 + \frac{R_1^3}{d^3}U_1\right), \tag{15}$$

$$F = 3\rho V_{02}\left(\frac{R_1^2}{d^3}\dot{R}_1U_2 + \frac{R_1^3}{d^4}U_1U_2 - \frac{R_1^4}{d^5}\dot{R}_1^2\right), \tag{16}$$

where M_1 and M_2 are the generalized momenta of the bubble and particle and F is the translational force that acts equally and oppositely on the bubble and particle. In Eqs. (12)–(16) it proved convenient for interpretation to introduce the volumes $V_1=\frac{4}{3}\pi R_1^3$ and $V_{02}=\frac{4}{3}\pi R_{02}^3$ and the particle density $\rho_2=m_2/V_{02}$. Except for the terms of order R^5/d^5 and those containing the particle mass, Eqs. (12)–(16) coincide with the corrected order R^4/d^4 equations published by Harkin *et al.*[7] describing coupled pulsation and translation of two bubbles after setting the radial velocity of the second bubble $\dot{R}_2$ to zero in their equation (see footnote 17 in Ref. 11 for discussion of the corrections).

It is straightforward, albeit tedious, to obtain expressions for the velocity potential and model equations to higher orders of accuracy than R^5/d^5. The steps outlined in the Appendix were automated with the aid of the MAXIMA computer algebra system to generate expressions for ϕ to arbitrary order in R^n/d^n. Harkin *et al.*[7] used a different approach to derive the velocity potential and model equations, which we also automated to arbitrary order R^n/d^n to verify the present method. The results are identical. For example, to order R^{10}/d^{10} the interaction force is given by

$$F = \frac{2}{5}\pi\rho R_1^2R_{02}^3\left[\frac{10}{d^3}U_2\dot{R}_1 + 15\frac{R_1}{d^4}U_1U_2 - 10\frac{R_1^2}{d^5}\dot{R}_1^2 - 25\frac{R_1^3}{d^6}\dot{R}_1U_1 - \frac{R_1}{d^7}(15R_1^3U_1^2 + 15R_{02}^3U_2^2 + 27R_1^3\dot{R}_1U_2 + 20R_1\dot{R}_1^2R_{02}^2) - 70\frac{R_1^3R_{02}^4}{d^8}\dot{R}_1U_1 - 2\frac{R_1^2}{d^9}(15R_{02}^4\dot{R}_1^2 + 30R_1^2R_{02}^2U_1^2 + 30R_1R_{02}^3U_1^2 - 18R_1^4\dot{R}_1^2 - 20R_1R_{02}^3\dot{R}_1U_2) + \frac{9}{2}\frac{R_1^3}{d^{10}}U_1(9R_1^4\dot{R}_1 + 10R_1R_{02}^3U_2 - 30R_{02}^4\dot{R}_1)\right]. \tag{17}$$

An analytic expression for the time-averaged interaction force $\langle F\rangle$ can be derived for infinitesimal bubble pulsation at an instant when the translational velocities of the bubble and particle are zero. At this instant $M_1=M_2=0$ and Eqs. (14) and (15) can be rearranged to obtain

$$U_2 = \frac{3\rho}{\rho+2\rho_2}\frac{R_1^2}{d^2}\dot{R}_1 + O\left(\frac{R^3}{d^3}\right), \tag{18}$$

$$U_1 = 3\frac{R_{02}^3}{d^3}U_2 = \frac{9\rho}{\rho+2\rho_2}\frac{R_1^2R_{02}^3}{d^5}\dot{R}_1 + O\left(\frac{R^6}{d^6}\right). \tag{19}$$

Substitution of Eqs. (18) and (19) into Eq. (17), assuming small periodic bubble pulsations with

$$R_1(t) = R_{01} + \xi_0 \sin \omega t, \quad \xi_0 \ll R_{01}, \tag{20}$$

and time-averaging over one acoustic cycle yields to leading order in ξ_0 the following contributions at fifth-, seventh-, and ninth-orders, respectively:

$$\langle F_5 \rangle = 4\pi\rho\left(\frac{\rho - \rho_2}{\rho + 2\rho_2}\right)\frac{R_{01}^4 R_{02}^3}{d^5}\omega^2\xi_0^2, \tag{21}$$

$$\langle F_7 \rangle = -4\pi\rho\frac{R_{01}^4 R_{02}^5}{d^7}\omega^2\xi_0^2, \tag{22}$$

$$\langle F_9 \rangle = -\frac{3}{5}\pi\rho[(15\rho - 24\rho_2)R_{01}^4 + (10\rho + 20\rho_2)R_{02}^4]\frac{R_{01}^4 R_{02}^3}{(\rho + 2\rho_2)d^9}\omega^2\xi_0^2. \tag{23}$$

There are no time-averaged contributions from terms at first through fourth orders, or from any even-order terms in R/d. Equation (21) is the expression derived by Coakley and Nyborg,[12,23] by an entirely different approach, for the force acting between a bubble and a rigid, stationary particle. Equation (21) indicates that, to leading order, for a particle having density greater than that of the liquid, $\rho_2 > \rho$, the time-averaged interaction force is negative, and the bubble and particle are attracted toward one another. For a less dense particle, $\rho_2 < \rho$, the bubble and particle repel. The result also demonstrates why it is essential to retain terms through fifth order in R/d in the equations of motion. Despite the restrictive conditions under which Eq. (21) is derived, the same trends are predicted by numerical solutions of Eqs. (12)–(16) even for large bubble pulsations and with the particle in motion.

While it is true that the influence of terms at higher orders, proportional to R^n/d^n, tends to decrease with increasing order n, the higher-order contributions to the translational force may not be negligible for sufficiently small separation distances. For a neutrally buoyant particle ($\rho_2 = \rho$) the contribution of the fifth-order terms, given by Eq. (21), vanishes. This seems to suggest that a seventh-order model is needed whenever considering neutrally buoyant particles, but we will show in Sec. IV A that the contribution of the higher-order terms is small unless the bubble and particle are extremely close. Note that the time-averaged contribution from the seventh-order force terms is always attractive, regardless of the relative particle density, while the contribution from the ninth-order terms is attractive except in the case of very dense, small particles.

III. SIMULATION PARAMETERS AND LOSSES

Inclusion of loss factors is guided by parameter values used for the numerical simulations. The equilibrium radius of the bubble was always taken to be $R_{01} = 100$ μm, the parameters for the surrounding liquid to be $\rho = 1000$ kg/m^3 and $\sigma = 0.073$ N/m, and the ambient pressure and gas constant to be $P_0 = 101$ kPa and $\gamma = 1.4$, respectively, corresponding to an air bubble in water at one standard atmosphere.

The natural frequency of the bubble in the absence of viscosity, commonly called the Minnaert frequency, is 32 kHz. For all simulations the bubble is either released from a non-equilibrium radius and allowed to ring down at its natural frequency, or it is driven at its natural frequency by an acoustic excitation. In this case it is sufficient for our purposes to introduce a damping term in the radial equation of motion that provides the correct attenuation at the natural frequency. Damping of the radial motion of the bubble is modeled via introduction of the quantity $-4(\eta_{\text{eff}}/\rho)\dot{R}_1/R_1$ on the right-hand side of Eq. (12).[24] Here, η_{eff} is an effective viscosity that is assigned the value $\eta_{\text{eff}} = 20\eta = 20$ mPa s, where $\eta = 1$ mPa s is the viscosity of water. This choice of η_{eff} approximates the total damping, due to heat transfer and radiation as well as viscosity, of a 100 μm bubble that pulsates at its natural frequency.[23] Harkin *et al.*[7] used the same value in their simulations of interacting bubbles with radii on the order of 100 μm. The translational velocity of the bubble is small, and viscous drag on the translational motion was taken into account by introducing the drag force $-4\pi\eta R_1 U_{r1}$ on the right-hand side of Eq. (13) for M_1, where U_{ri} is the translational velocity of the bubble (or particle) relative to the local velocity of the surrounding liquid.[11] Viscous drag on the translational motion of the particle was taken into account by introducing the Stokes drag term $-6\pi\eta R_2 U_{r2}$ on the right-hand side of Eq. (13) for M_2. Further discussion of drag is given in Ref. 11.

In the simulations that follow, the equations of motion were integrated numerically with a standard backward differentiation routine for different separation distances, acoustic excitations, and values of equilibrium radius and density of the particle. The case of free response, with no acoustic excitation, is considered first. The reason for investigating the free response is to avoid the primary Bjerknes forces produced by acoustic excitation and thus isolate the bubble-particle interaction forces, i.e., the secondary Bjerknes forces.

IV. FREE RESPONSE

The free response of the system is investigated by setting the initial bubble radius to a non-equilibrium value $R_1(0) \neq R_{01}$ and releasing it from rest. The particle is initially at rest in its equilibrium state. The simulations in this section were run with the higher-order extensions of Eqs. (12)–(16) that are accurate to order R^9/d^9, where the numerical solution converges, as will be demonstrated in Sec. IV A.

Presented in Fig. 3 are results for a particle of equilibrium radius $R_{02} = 2R_{01} = 200$ μm. Larger particles are used to emphasize interaction effects, because as seen from Eqs. (12) and (16), to leading order the interaction forces are proportional to the volume of the particle. The initial positions of the bubble and particle are, respectively, $X_1(0) = 0$ and $X_2(0) = 2.5R_{02} = 500$ μm, when the bubble is released from its initial radius $R_1(0) = 120$ μm, a value 20% greater than its equilibrium radius. Responses are shown for three different particle densities: $\rho_2 = 500$, 1000, and 2000 kg/m^3 ($\rho_2/\rho = 0.5$, 1, or 2). To aid the reader, the indices $i = 1$ and $i = 2$ on

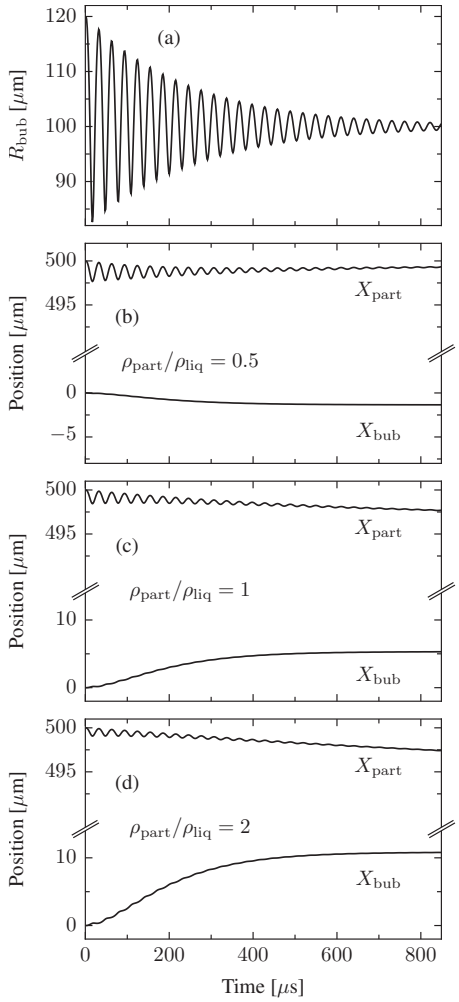


FIG. 3. Response of a bubble with equilibrium radius 100 μm and initial radius 120 μm with rigid particles of different densities. Bubble and particle initially separated by 500 μm. (a) Radial response of the bubble (independent of ρ_{part}/ρ_{liq}). (b) Positions for a light particle (ρ_{part}/ρ_{liq}=0.5). (c) Positions for a neutral particle (ρ_{part}/ρ_{liq}=1). (d) Positions for a heavy particle (ρ_{part}/ρ_{liq}=2).

parameters shown in the figures are replaced with the subscripts "bub" and "part," respectively, and the density of the liquid is denoted by ρ_{liq}.

The radial response of the bubble, which is independent of the density ratio ρ_{part}/ρ_{liq} to within graphical resolution, is shown in Fig. 3(a). Figures 3(b)–3(d) display the positions of the bubble and particle for the three density ratios (note the split vertical axes). Consistent with the approximate analytical result for the time-averaged interaction force given by Eqs. (21)–(23), the bubble and particle repel for ρ_{part}/ρ_{liq} =0.5, while they attract for ρ_{part}/ρ_{liq}=2 and ρ_{part}/ρ_{liq}=1, although the attractive force is less when the particle is neutrally buoyant. The attractive force in the case of a neutrally buoyant particle is due only to the inclusion of the higher-order terms which, as Eqs. (22) and (23) show, are both attractive for this set of parameters. If instead the order R^5/d^5 model equations were used, there would be virtually no translation in the case of a neutrally buoyant particle. Note that whereas the bubble is observed to come to rest within the time frames shown, the inertia of the particle causes it to drift for considerably longer times.

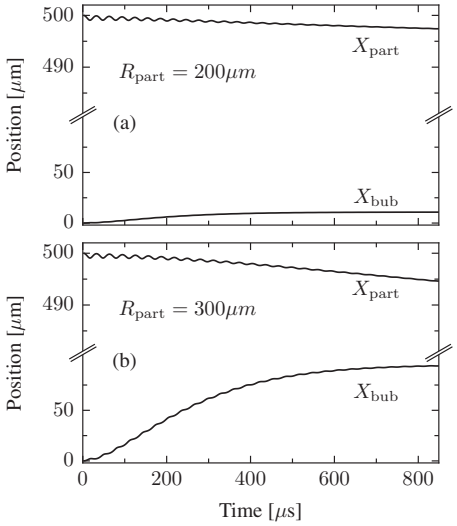


FIG. 4. Response of a bubble with equilibrium radius 100 μm and initial radius 120 μm with rigid heavy (ρ_{part}/ρ_{liq}=2) particles of different radii. Bubble and particle initially separated by 500 μm. (a) Particle radius R_{02} =200 μm. (b) Particle radius R_{02}=300 μm.

The dependence on particle size is illustrated in Fig. 4. Figure 4(a) is repeated from Fig. 3(d) (on an expanded axis), and Fig. 4(b) shows the effect of increasing the particle radius by 50%. Equation (16) indicates that the interaction force will increase for larger particles, although the actual increase is stronger than the R_{02}^3 dependence suggested by Eq. (16) because of the inclusion of the higher-order terms in the simulation model.

A. Truncation error

The order R^5/d^5 model presented in Sec. II C is accurate to the minimum order in R/d required to account for the interaction force between the bubble and a rigid particle. As described previously, the steps outlined in Sec. II were automated with the aid of a computer algebra system to generate models of arbitrary accuracy in R^n/d^n for the case of two interacting spheres. Here we discuss higher-order models of a bubble interacting with a rigid particle. Equation sets accurate up to and including order R^{15}/d^{15} were solved numerically at various separation distances to determine the influence of higher-order terms on the system dynamics. Figure 5 shows cases of different initial separation distances for heavy (ρ_{part}/ρ_{liq}=2) and light (ρ_{part}/ρ_{liq}=0.5) rigid particles of radius R_{02}=200 μm. The bubble and particle were initially separated by $d(0)$=0.5 mm [Fig. 5(a)] or 1.5 mm [Figs. 5(b) and 5(c)] (i.e., 2.5 or 7.5 particle radii).

The time-averaged contributions of terms of higher order than R^5/d^5 in the interaction force are attractive for these particle parameters [see Eqs. (22) and (23)]. Therefore, we expect the inclusion of these terms to result in greater translational motion in the case of a heavy particle [Fig. 5(c)], but less translation in the case of a light particle [Fig. 5(b)]. This prediction is confirmed by the simulations. As Fig. 5(a) shows, at large separation distances it is sufficient to truncate the model at n=5 [Eqs. (12)–(16)]. However, as the bubble and particle move closer together the contribution of the

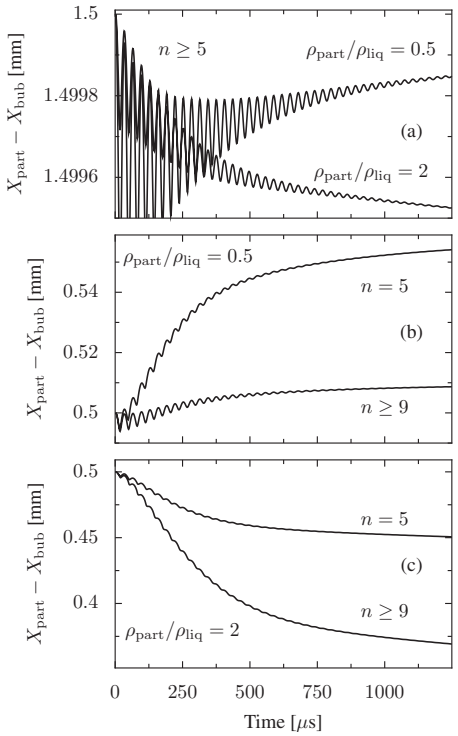


FIG. 5. Numerical solution of models accurate to various orders of R/d for a bubble of equilibrium radius 100 μm and initial radius of 120 μm with rigid particles of radius 200 μm and various densities. (a) Bubble and particle initially separated by 1.5 mm for models of order R^5/d^5 and higher for heavy (ρ_{part}/ρ_{liq}=2) and light (ρ_{part}/ρ_{liq}=0.5) particles. (b) Bubble and a light (ρ_{part}/ρ_{liq}=0.5) particle initially separated by 0.5 mm for models accurate to order R^5/d^5 or R^9/d^9 and higher. (c) Bubble and a heavy (ρ_{part}/ρ_{liq} =2) particle initially separated by 0.5 mm for models accurate to order R^5/d^5 or greater than R^9/d^9.

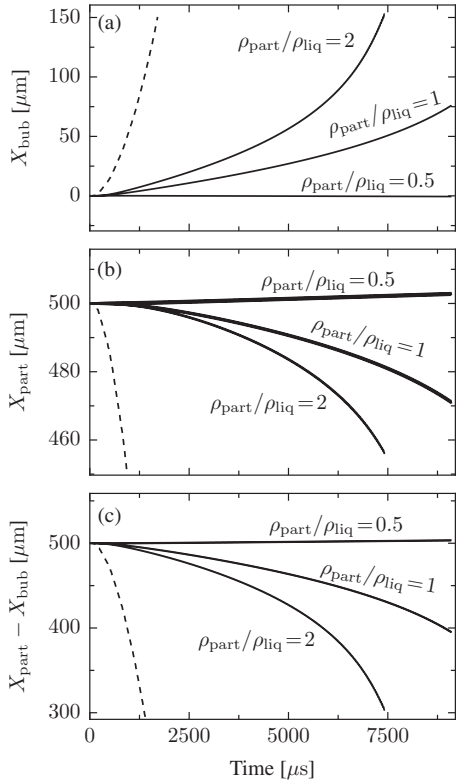


FIG. 6. Translational motion of a bubble of equilibrium radius 100 μm interacting with a second bubble of radius 100 μm (dashed line) or rigid particles of radius 200 μm (solid lines) and density ratios ρ_{part}/ρ_{liq}=0.5, 1, and 2 excited by a sinusoidal plane wave source of amplitude 1 kPa located at −10 cm along the x axis. Initial separation distance 500 μm. (a) Bubble position. (b) Particle position. (c) Separation distance between bubble and particle.

higher-order terms becomes significant. In the cases considered here the solution converges numerically at order n=9 for both light [Fig. 5(b)] and heavy [Fig. 5(c)] particles.

V. FORCED RESPONSE

The system may also be driven by an acoustic source. The source was included in the model using the method of Ilinskii *et al.*[11] To summarize, the kinetic energy due to a pulsating sphere with prescribed motion is interpreted in terms of p_{src} and $\mathbf{u}_{src}$, the pressure and particle velocity of an external source. The source kinetic energy is calculated to order R^5/d^5 and is thus consistent with the free-response model derived in Sec. II. This additional kinetic energy is added to the system kinetic energy, Eq. (10). The same parameter set outlined in Sec. IV was used, with an initial separation distance between the bubble and particle of 500 μm and a sinusoidal acoustic pressure of amplitude p_0 =1 kPa and frequency of 32 kHz (equal to the natural frequency of the bubble). Because the model is spatially one-dimensional, the acoustic wave radiated by the source must be planar and propagate along the x axis. While the inclusion of source terms is an acknowledgment of finite liquid compressibility, in this case the acoustic wavelength is much larger than the separation distance between the bubble and particle. Therefore the liquid may be assumed to be locally incompressible from the standpoint of the bubble or particle.

While source interaction terms were included to order R^5/d^5, terms accounting for bubble-particle interaction were included up to R^9/d^9 as in Sec. IV. Simulation results for the source parameters used in the present section suggest that source interaction terms are negligible at R^4/d^4, and therefore omission of the higher-order source terms will not affect the system dynamics.

Figure 6 shows the positions and separation distances between the bubble and particle for the three density ratios (ρ_{part}/ρ_{liq}=0.5,1,2) when the source is located at −10 cm along the x axis, and Fig. 7 shows results for a source located at +10 cm. In both figures the bubble position is plotted in part (a), the particle position in part (b), and the distance separating the bubble and particle in part (c). For comparison, results for a two-bubble system (with a second bubble of radius 100 μm replacing the particle) are shown with dashed lines. Note that in both figures translation in the two-bubble system is much greater than in the bubble-particle system. This is because the secondary Bjerknes force acting between two oscillating bubbles is of order R^2/d^2, three orders larger than the translational force in the bubble-particle system.

Notice that the direction of particle translation [Figs. 6(b) and 7(b)] is determined by the particle density [repulsive for a sufficiently light particle but attractive otherwise; recall Eqs. (21)–(23)]. However, the radiation force (primary Bjerknes force) exerted on the bubble by the acoustic field has a non-zero time average. Therefore, the direction of

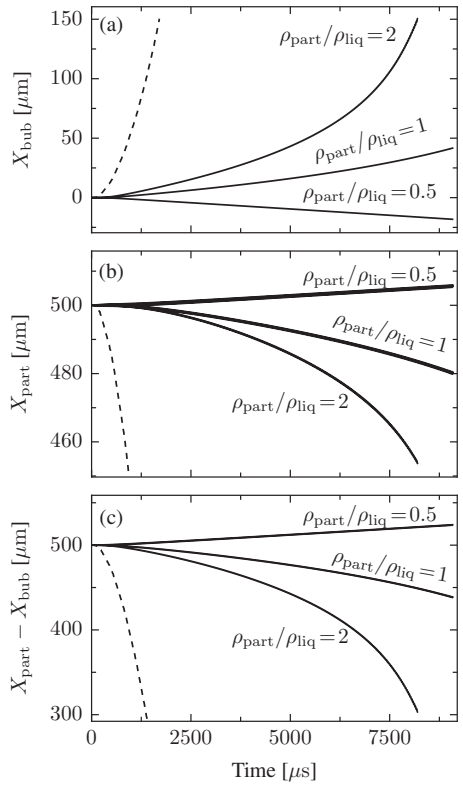


FIG. 7. Translational motion of a bubble of equilibrium radius 100 μm interacting with a second bubble of radius 100 μm (dashed line) or rigid particles (solid lines) of radius 200 μm and density ratios ρ_{part}/ρ_{liq}=0.5, 1, and 2 excited by a sinusoidal plane wave source of amplitude of 1 kPa located at +10 cm along the x axis. Initial separation distance 500 μm. (a) Bubble position. (b) Particle position. (c) Separation distance between bubble and particle.

bubble translation [Figs. 6(a) and 7(a)] is also influenced by the propagation direction of the acoustic excitation. In Fig. 6, with the source located at −10 cm, the radiation force tends to push the bubble toward the particle. This force, in combination with the higher-order interaction terms, causes the bubble to move toward the particle [Fig. 6(a)] for neutrally-buoyant and heavy particles, while in the case of a light particle the bubble remains nearly stationary. In Fig. 7, with the source located at +10 cm, the radiation force tends to push the bubble away from the particle. For the light particle its effect is sufficient to overcome the attractive force of the higher-order contributions [see Eqs. (21)–(23)] and cause overall repulsion of the bubble and particle. However, for heavy and neutrally buoyant particles, forces due to bubble-particle interaction dominate, resulting in attraction [Fig. 7(c)].

These results suggest that bubble-particle interaction forces are certainly important for accurate description of free-response translational dynamics. However, in the case of a forced system the bubble-particle translational forces may be overshadowed by the primary Bjerknes forces from an acoustic source of sufficient amplitude.

VI. SUMMARY

Lagrangian mechanics were used to derive dynamical equations describing the interaction of a gas bubble and a rigid particle in an incompressible liquid. The derivations were carried out through order R^5/d^5 because this is the minimum accuracy required to account for the interaction force. Dynamical models accurate to higher orders of R^n/d^n were generated via a computer algebra system to illustrate consequences of truncating the model at order R^5/d^5. In the case of a forced response, the presence of an external plane wave source was included to an order consistent with the model at order R^5/d^5. The system dynamics exhibit dependence on the density and size of the particle as well as the direction of source propagation.

ACKNOWLEDGMENTS

This work was supported by a National Science Foundation Graduate Research Fellowship and by NIH Grant No. DK070618.

APPENDIX: DERIVATION OF THE VELOCITY POTENTIAL

Here an expression for the velocity potential, which satisfies Eq. (5) to order R^5/d^5 subject to the boundary conditions [Eqs. (8) and (9)], is derived. The resulting expression may then be substituted into Eq. (7) to obtain an expression for the kinetic energy of the liquid. The total velocity potential is expressed as

$$\phi(\mathbf{r}_i) = \phi^r(\mathbf{r}_i) + \phi^t(\mathbf{r}_i), \tag{A1}$$

where ϕ^r is the component due to radial pulsation of the bubble and ϕ^t is the component due to translation. First, an expression for the radial component of the velocity potential is obtained. Although the particle is rigid and therefore does not pulsate, for the moment we will proceed as though it pulsates radially with velocity $\dot{R}_2$. We then specialize the result for a rigid particle. For an isolated pulsating sphere with radius R_i and wall velocity $\dot{R}_i$ in an infinite liquid, the radial component of the velocity potential a distance r_i from the center of the sphere is

$$\phi^r_{0i}(\mathbf{r}_i) = -\frac{R_i^2}{r_i}\dot{R}_i. \tag{A2}$$

An initial approximation of the radial velocity potential for an infinite liquid containing two pulsating spheres is the sum

$$\phi^r_0 = \phi^r_{01} + \phi^r_{02}. \tag{A3}$$

While Eq. (A3) satisfies Eq. (5), it does not obey the boundary condition on either sphere. For example, let the velocity potential be expressed in terms of the coordinates of the bubble ($\mathbf{r}_i=\mathbf{r}_1$), and calculate the liquid velocity at the bubble wall for comparison with Eq. (8). The velocity potential corresponding to the particle, $\phi^r_{02}(\mathbf{r}_2)$, can be expressed in terms of $\mathbf{r}_1$ by expanding ϕ^r_{02} in a Taylor series evaluated at the bubble wall:

$$\phi_{02}^r(\mathbf{r}_2) = \phi_{02}^r(\mathbf{d}_{21} + \mathbf{r}_1)$$
$$\simeq \phi_{02}^r(\mathbf{d}_{21}) + \nabla_{\mathbf{r}_2}\phi_{02}^r(\mathbf{r}_2)|_{\mathbf{r}_2=\mathbf{d}_{21}} \cdot \mathbf{r}_1$$
$$= -\frac{R_2^2}{d_{21}}\dot{R}_2 + \frac{R_2^2}{d_{21}^2} r_1 \dot{R}_2(\mathbf{n}_{21} \cdot \mathbf{n}_1), \tag{A4}$$

where $\mathbf{d}_{21}=\mathbf{r}_{01}-\mathbf{r}_{02}$ is the vector pointing from the center of the particle to the center of the bubble (recall Fig. 1), $d_{21}=|\mathbf{d}_{21}|$, and $\mathbf{n}_{21}=\mathbf{d}_{21}/d_{21}$. In Eq. (A4) and all subsequent expansions the series is truncated after the linear (dipole) term because all higher-order terms are orthogonal to the monopole and dipole terms in the integrand of Eq. (7), and therefore cannot contribute to the kinetic energy, as a result of including only monopole and dipole terms in boundary conditions [Eqs. (8) and (9)].

Evaluating the velocity at the bubble wall due to ϕ_0^r,

$$\left.\frac{\partial \phi_0^r}{\partial r_1}\right|_{S_1} = \left.\left(\frac{\partial \phi_{01}^r}{\partial r_1} + \frac{\partial \phi_{02}^r}{\partial r_1}\right)\right|_{S_1} = \dot{R}_1 + \frac{R_2^2}{d_{21}^2}\dot{R}_2(\mathbf{n}_{21} \cdot \mathbf{n}_1), \tag{A5}$$

and comparing with Eq. (8) shows that the error in satisfying the boundary condition on the radial motion is of order R^2/d^2 [the dipole term $\mathbf{U}_i \cdot \mathbf{n}_i$ in Eq. (8) will be satisfied by ϕ^t in Eq. (A1)]. To counteract the second term in Eq. (A5) a correction ϕ_1^r is added to ϕ_0^r which satisfies Eq. (5) as well as

$$\left.\frac{\partial \phi_1^r}{\partial r_1}\right|_{S_1} = -\left.\frac{\partial \phi_0^r}{\partial r_1}\right|_{S_1} + \dot{R}_1 = -\frac{R_2^2}{d_{21}^2}\dot{R}_2(\mathbf{n}_{21} + \mathbf{n}_1). \tag{A6}$$

The appropriate correction, applied to both the bubble and particle and expressed in terms of the local coordinate $\mathbf{r}_1$ of the bubble, is

$$\phi_1^r = \phi_{11}^r(\mathbf{r}_1) + \phi_{12}^r(\mathbf{d}_{21} + \mathbf{r}_1)$$
$$\simeq \phi_{11}^r(\mathbf{r}_1) + \phi_{12}^r(\mathbf{d}_{21}) + \nabla_{\mathbf{r}_2}\phi_{12}^r(\mathbf{r}_2)|_{\mathbf{r}_2=\mathbf{d}_{21}} \cdot \mathbf{r}_1$$
$$= \frac{R_2^2 R_1^3}{2d_{21}^2 r_1^2}\dot{R}_2(\mathbf{n}_{21} \cdot \mathbf{n}_1) + \frac{R_1^2 R_2^3}{2d_{21}^4}\dot{R}_1 + \frac{R_1^2 R_2^3}{d_{21}^5} r_1 \dot{R}_1(\mathbf{n}_{21} \cdot \mathbf{n}_1). \tag{A7}$$

Examining the velocity at the bubble wall due to terms ϕ_0^r and ϕ_1^r,

$$\left.\frac{\partial}{\partial r_1}(\phi_0^r + \phi_1^r)\right|_{S_1} = \dot{R}_1 + \frac{R_1^2 R_2^3}{d_{21}^5}\dot{R}_1(\mathbf{n}_{21} \cdot \mathbf{n}_1), \tag{A8}$$

shows that the error in satisfying the boundary condition for the radial motion is of order R^5/d^5, and therefore another correction term, ϕ_2^r, must be added. The appropriate correction is

$$\phi_2^r = \phi_{21}^r(\mathbf{r}_1) + \phi_{22}^r(\mathbf{d}_{21} + \mathbf{r}_1)$$
$$= \frac{R_1^5 R_2^3}{2d_{21}^5 r_1^2}\dot{R}_1(\mathbf{n}_{21} \cdot \mathbf{n}_1) + O\left(\frac{R^7}{d^7}\right). \tag{A9}$$

The velocity at the bubble wall, based on the sum of ϕ_0^r with correction terms ϕ_1^r and ϕ_2^r, is

$$\left.\frac{\partial}{\partial r_1}(\phi_0^r + \phi_1^r + \phi_2^r)\right|_{S_1} = \dot{R}_1 + O\left(\frac{R^8}{d^8}\right). \tag{A10}$$

Comparison with Eq. (8) shows that the expression is accurate to the desired order R^5/d^5. The final form of the velocity potential for the radial motion, expressed in the local coordinates of the bubble and consisting of terms ϕ_0^r, ϕ_1^r, and ϕ_2^r, is

$$\phi^r(\mathbf{r}_1) = -\frac{R_1^2}{r_1}\dot{R}_1 - \frac{R_2^2}{d_{21}}\dot{R}_2 + \frac{R_2^2}{d_{21}^2} r_1 \dot{R}_2(\mathbf{n}_{21} \cdot \mathbf{n}_1)$$
$$+ \frac{R_2^2 R_1^3}{2d_{21}^2 r_1^2}\dot{R}_2(\mathbf{n}_{21} \cdot \mathbf{n}_1) - \frac{R_1^2 R_2^3}{2d_{21}^4}\dot{R}_1$$
$$+ \frac{R_1^2 R_2^3}{d_{21}^5} r_1 \dot{R}_1(\mathbf{n}_{21} \cdot \mathbf{n}_1) + \frac{R_1^5 R_2^3}{2d_{21}^5 r_1^2}\dot{R}_1(\mathbf{n}_{21} \cdot \mathbf{n}_1). \tag{A11}$$

The corresponding expression in terms of the coordinates of the particle, $\mathbf{r}_2$, is obtained simply by exchanging the subscripts 1 and 2 in all quantities. Velocity potential expressions for our specific case of a rigid particle with radius R_{02} and zero radial velocity are then obtained from Eq. (A11) by setting $R_2=R_{02}$ and $\dot{R}_2=0$.

The translational component of the velocity potential is obtained in a similar fashion. For an isolated translating sphere with radius R_i and translational velocity $\mathbf{U}_i$, the velocity potential at a distance r_i from the center of the sphere is

$$\phi_{0i}^t(\mathbf{r}_i) = -\frac{R_i^3}{2r_i^2}\mathbf{U}_i \cdot \mathbf{n}_i. \tag{A12}$$

As with the radial component, the individual contributions from each sphere are summed and then the contribution from the particle is expressed in terms of the local coordinate of the bubble:

$$\phi_0^t = \phi_{01}^t(\mathbf{r}_1) + \phi_{02}^t(\mathbf{d}_{21} + \mathbf{r}_1)$$
$$\simeq \phi_{01}^t(\mathbf{r}_1) + \phi_{02}^t(\mathbf{d}_{21}) + \nabla_{\mathbf{r}_2}\phi_{02}^t(\mathbf{r}_2)|_{\mathbf{r}_2=\mathbf{d}_{21}} \cdot \mathbf{r}_1$$
$$= -\frac{R_1^3}{2r_1^2}(\mathbf{U}_1 \cdot \mathbf{n}_1) - \frac{R_2^3}{2d_{21}^2}\mathbf{U}_2 \cdot \mathbf{n}_{21}$$
$$+ \frac{R_2^3}{2d_{21}^3} r_1[3(\mathbf{U}_2 \cdot \mathbf{n}_{21})(\mathbf{n}_{21} \cdot \mathbf{n}_1) - \mathbf{U}_2 \cdot \mathbf{n}_1]. \tag{A13}$$

Evaluating the translational velocity at the wall of the bubble,

$$\left.\frac{\partial \phi_0^t}{\partial r_1}\right|_{S_1} = \mathbf{U}_1 \cdot \mathbf{n}_1 + \frac{R_2^3}{2d_{21}^3}[3(\mathbf{U}_2 \cdot \mathbf{n}_{21})(\mathbf{n}_{21} \cdot \mathbf{n}_1) - \mathbf{U}_2 \cdot \mathbf{n}_1], \tag{A14}$$

and comparing with Eq. (8) shows that the error is of order R^3/d^3. A correction term, ϕ_1^t, which satisfies Eq. (5) as well as the condition

$$\left.\frac{\partial \phi_1^t}{\partial r_1}\right|_{S_1} = -\left.\frac{\partial \phi_0^t}{\partial r_1}\right|_{S_1} + \mathbf{U}_1 \cdot \mathbf{n}_1$$
$$= -\frac{R_2^3}{2d_{21}^3}[3(\mathbf{U}_2 \cdot \mathbf{n}_{21})(\mathbf{n}_{21} \cdot \mathbf{n}_1) - \mathbf{U}_2 \cdot \mathbf{n}_1], \tag{A15}$$

is needed. The appropriate correction, applied to both the bubble and particle and expressed in terms of $\mathbf{r}_1$, is

$$\phi_1^t = \phi_{11}^t(\mathbf{r}_1) + \phi_{12}^t(\mathbf{d}_{21} + \mathbf{r}_1)$$
$$\simeq \phi_{11}^t(\mathbf{r}_1) + \phi_{12}^t(\mathbf{d}_{21}) + \nabla_{\mathbf{r}_2}\phi_{12}^t(\mathbf{r}_2)|_{\mathbf{r}_2=\mathbf{d}_{21}} \cdot \mathbf{r}_1$$
$$= \frac{R_1^3 R_2^3}{4r_1^2 d_{21}^3}[3(\mathbf{U}_2 \cdot \mathbf{n}_{21})(\mathbf{n}_{21} \cdot \mathbf{n}_1) - \mathbf{U}_2 \cdot \mathbf{n}_1]$$
$$- \frac{R_2^3 R_1^3}{d_{21}^5}(\mathbf{U}_1 \cdot \mathbf{n}_{21}) + O\left(\frac{R^6}{d^6}\right). \tag{A16}$$

Evaluating the translational velocity at the wall of the bubble yields

$$\left.\frac{\partial \phi^t}{\partial r_1}\right|_{S_1} = \left.\frac{\partial}{\partial r_1}(\phi_0^t + \phi_1^t)\right|_{S_1} = \mathbf{U}_1 \cdot \mathbf{n}_1 + O\left(\frac{R^6}{d^6}\right), \tag{A17}$$

and therefore the expression for the translational velocity potential is accurate to the desired order of R^5/d^5. The final form of the translational velocity potential, consisting of ϕ_0^t and the correction term ϕ_1^t, is

$$\phi^t(\mathbf{r}_1) = -\frac{R_1^3}{2r_1^2}(\mathbf{U}_1 \cdot \mathbf{n}_1) - \frac{R_2^3}{2d_{21}^2}\mathbf{U}_2 \cdot \mathbf{n}_{21}$$
$$+ \frac{R_2^3}{2d_{21}^3} r_1[3(\mathbf{U}_2 \cdot \mathbf{n}_{21})(\mathbf{n}_{21} \cdot \mathbf{n}_1) - \mathbf{U}_2 \cdot \mathbf{n}_1]$$
$$+ \frac{R_1^3 R_2^3}{4r_1^2 d_{21}^3}[3(\mathbf{U}_2 \cdot \mathbf{n}_{21})(\mathbf{n}_{21} \cdot \mathbf{n}_1) - \mathbf{U}_2 \cdot \mathbf{n}_1]$$
$$- \frac{R_2^3 R_1^3}{d_{21}^5}(\mathbf{U}_1 \cdot \mathbf{n}_{21}). \tag{A18}$$

As with Eq. (A11), the expression for Eq. (A18) in terms of the coordinates of the particle, $\mathbf{r}_2$, is obtained by exchanging the subscripts 1 and 2. The total velocity potential ϕ, the sum of Eqs. (A11) and (A18), may now be substituted into Eq. (7) to obtain Eq. (10).

[1] O. V. Voinov and A. M. Golovin, "Lagrange equation for a system of bubbles of varying radii in a liquid of small viscosity," Fluid Dyn. **5**, 458–464 (1970).

[2] G. N. Kuznetsov and I. E. Shchekin, "Interaction of pulsating bubbles in a viscous liquid," Sov. Phys. Acoust. **18**, 466–469 (1973).

[3] E. A. Zabolotskaya, "Interaction of gas bubbles in a sound field," Sov. Phys. Acoust. **30**, 365–368 (1984).

[4] Yu. A. Ilinskii and E. A. Zabolotskaya, "Cooperative radiation and scattering of acoustic waves by bubbles in liquid," J. Acoust. Soc. Am. **92**, 2837–2841 (1992).

[5] T. Barbat, N. Ashgriz, and C. Liu, "Dynamics of two interacting bubbles in an acoustic field," J. Fluid Mech. **389**, 137–168 (1999).

[6] A. A. Doinikov, "Translational motion of two interacting bubbles in a strong acoustic field," Phys. Rev. E **64**, 026301 (2001).

[7] A. Harkin, T. J. Kaper, and A. Nadim, "Coupled pulsation and translation of two gas bubbles in a liquid," J. Fluid Mech. **445**, 377–411 (2001).

[8] N. Wang and P. Smereka, "Effective equations for sound and void wave propagation in bubbly fluids," SIAM J. Appl. Math. **63**, 1849–1888 (2003).

[9] A. A. Doinikov, "Mathematical model for collective bubble dynamics in strong ultrasound fields," J. Acoust. Soc. Am. **116**, 821–827 (2004).

[10] M. F. Hamilton, Yu. A. Ilinskii, G. D. Meegan, and E. A. Zabolotskaya, "Interaction of bubbles in a cluster near a rigid surface," ARLO **6**, 207–213 (2005).

[11] Yu. A. Ilinskii, M. F. Hamilton, and E. A. Zabolotskaya, "Bubble interaction dynamics in Lagrangian and Hamiltonian mechanics," J. Acoust. Soc. Am. **121**, 786–795 (2007).

[12] W. T. Coakley and W. L. Nyborg, "Cavitation dynamics of gas bubbles," in *Applications in Ultrasound: Its Applications in Medicine and Biology*, edited by F. J. Fry (Elsevier, New York, 1978), pp. 77–159.

[13] M. S. Plesset and R. B. Chapman, "Collapse of an initially spherical vapour cavity in the neighbourhood of a solid boundary," J. Fluid Mech. **47**, 283–290 (1971).

[14] J. R. Blake, B. B. Taib, G. Doherty, and Y. Tomita, "Transient cavities near boundaries. Part 1. Rigid boundary," J. Fluid Mech. **170**, 479–497 (1986).

[15] J. R. Blake and D. C. Gibson, "Cavitation bubbles near boundaries," Annu. Rev. Fluid Mech. **19**, 99–123 (1987).

[16] J. R. Blake, P. B. Robinson, A. Shima, and Y. Tomita, "Interaction of two cavitation bubbles with a rigid boundary," J. Fluid Mech. **255**, 707–721 (1993).

[17] Y. Tomita, P. B. Robinson, R. P. Tong, and J. R. Blake, "Growth and collapse of cavitation bubbles near a curved rigid boundary," J. Fluid Mech. **466**, 259–283 (2002).

[18] H. Miao and S. M. Gracewski, "Response of an ultrasonically excited bubble near a fixed rigid object," ARLO **6**, 144–150 (2005).

[19] S. M. Gracewski, H. Miao, and D. Dalecki, "Ultrasonic excitation of a bubble near a rigid or deformable sphere: Implications for ultrasonically induced hemolysis," J. Acoust. Soc. Am. **117**, 1440–1447 (2005).

[20] Yu. A. Pishchalnikov, O. A. Sapozhnikov, M. R. Bailey, J. C. Williams, Jr., R. O. Cleveland, T. Colonius, L. A. Crum, A. P. Evan, and J. A. McAteer, "Cavitation bubble cluster activity in the breakage of kidney stones by lithotripter shockwaves," J. Endourol **17**, 435–446 (2003).

[21] T. A. Hay, "A model of the interaction of bubbles and solid particles under acoustic excitation," Ph.D. dissertation, The University of Texas at Austin, Austin, TX (2008).

[22] H. Lamb, *Hydrodynamics*, 6th ed. (Dover, New York, 1993), Chaps. 46 and 81.

[23] T. G. Leighton, *The Acoustic Bubble* (Academic, San Diego, 1994).

[24] E. A. Zabolotskaya, Yu. A. Ilinskii, G. D. Meegan, and M. F. Hamilton, "Modifications of the equation for gas bubble dynamics in a soft elastic medium," J. Acoust. Soc. Am. **118**, 2173–2181 (2005).

An analytical model for turbulence scattered rays in the shadow zone for outdoor sound propagation calculation

Yiu Wai Lam
Acoustics Research Centre, University of Salford, Salford M5 4WT, United Kingdom

(Received 26 June 2008; revised 2 December 2008; accepted 9 January 2009)

In outdoor sound propagation, an inherent problem of the ray tracing method is its inability to determine the sound pressure level in the shadow zone, where geometrical rays do not penetrate. This is a serious problem in a turbulent atmosphere where significant sound energy will be scattered into the shadow. Empirical corrections that are determined from measurements or numerical simulations are limited to situations within the bounds of the empirical corrections. This paper describes a different approach where the ray tracing model is modified analytically into a scattered ray model. Rays are first diffracted from the shadow boundary, which is determined by the geometrical ray paths. The diffracted rays are then scattered by turbulence in their way to the receiver. The amount of scatter is determined from turbulence statistics that are determined from a Gaussian turbulence model. Most of the statistics are determined analytically except one element, which is determined empirically from numerical simulations. This turbulence scattered ray model is shown to have good accuracy against calculations based on the parabolic equation, and against previously published measurement data. It was found that the agreement is good both with and without turbulence, at distance up to 2 km from the shadow boundary.
 [DOI: 10.1121/1.3076928]

PACS number(s): 43.28.Gq, 43.28.Js, 43.50.Vt [JWP] Pages: 1340–1350

I. INTRODUCTION

The calculation of outdoor sound propagation is an important element in environmental noise assessments. Since the outdoor environment is complex, it is necessary for such calculations to take into account a large number of environmental factors, such as ground attenuation and atmospheric effects. A number of significant advancements have been made in the past few decades, and there now exist several accurate methods for calculation.[1–3] The parabolic equation (PE) method is particularly useful for long range sound propagation because of its accuracy and its ability to take into account range dependent ground and atmospheric conditions. Recent advancements in accelerating the calculation[4] also greatly enhanced its practical applicability. However, despite these advances, the method is still expensive to use at high frequencies, and is not readily usable in complex surroundings such as in urban cities where sharp changes in topography, meteorology, or material conditions frequently occur. For calculations in such surroundings, the ray tracing method[5] provides a more practical alternative approach. For example, in the current noise mapping exercise in Europe, most of the calculations are based on ray models.

Ray models in outdoor sound propagation come in different forms.[6–9] All are based on geometrical considerations. The variations come from the different degrees of approximation they place on the wave behavior along the propagation path. The ray tube approach and the semi-analytical ray model take into account more of the wave behavior, but can suffer from singularity effects such as caustics. The heuristic ray approach and its variant, which only use the ray paths to provide information on the phase and wave coherence, are easier to use, and can provide acceptable accuracy for engineering calculations. However, these ray models suffer from the same inherent problem that they cannot determine the propagation of sound into the shadow zone. By nature of their geometrical assumption, it is not possible for rays to penetrate into the shadow. There are various suggestions to overcome this problem. If the atmosphere can be assumed to be non-turbulent, range independent, and has a linear vertical sound speed profile, then the residue series approach[10] can be used to calculate the diffraction of sound into the shadow. However, this solution requires the determination of the roots of a complex function, which is difficult to obtain accurately deep in the shadow. Moreover, the inability to account for turbulence is a serious limitation for calculations involving real atmospheres. In realistic environments, a more practical approach is to apply empirical adjustments to the sound level in the shadow zone. For example, a recent suggestion[11] uses corrections that are determined empirically from a large number of numerical simulations. However, purely empirical corrections are fundamentally limited by the conditions under which they are developed. In contrast, an analytical approach will be more flexible, and could provide a better understanding of the underlying problem.

This paper describes the development of an analytical approach based on ray tracing to calculate the propagation of sound into the shadow zone. The basic ray tracing model is based on a discretized implementation of the heuristic model.[8] In an upward refracting case where a shadow is formed, the diffraction of sound into the shadow is assumed here to be analogous to the geometrical diffraction by an equivalent barrier in a homogeneous atmosphere. The scattering into the shadow zone due to turbulence is calculated

from statistics that are determined from a Gaussian turbulence model of the atmosphere. The accuracy of the model will be tested against the PE using published cases in literature.

II. THE RAY TRACING MODEL

Our basic ray tracing is based largely on a numerical approach. Once launched, rays are traced individually through the atmosphere to find the trajectory of the ray that forms the shadow boundary. The overall sound speed profile in the atmosphere can be arbitrary. In our implementation, the atmosphere is divided into a grid of small rectangular cells. The cell division is chosen such that the change in the sound speed within each cell is small enough to be approximated by a linear profile, so that an analytical solution[8] can be used within each cell to determine the ray propagation trajectory. The rule of thumb is to set the grid size at 1/10 of the length scale of the fastest changing condition. In practice, in the absence of turbulence, the grid size is typically 1 m in the vertical direction to accommodate the expected logarithmic shape of the sound speed profile near ground. The grid size in the horizontal direction is set at 5 m to accommodate likely range dependent variations of the terrain and meteorological conditions. However, when simulating the scattering of rays due to atmospheric turbulence, the grid size is set at 1/10 of the smallest length scale of the turbulence structure. Since the Gaussian turbulence spectrum used in this paper has a length scale of 1.1 m, the grid size used is 0.11 m. The ground terrain, which can be undulating, determines the lower boundary of the computational grid. Once the reflection geometry is worked out by the numerical ray tracing, the effect of the ground at each reflection can then be approximated by the Weyl–van der Pol formulation.[12,13] The height of the upper boundary of the computational grid is set using the rule of thumb of 1/10 of the maximum horizontal range.

In a downward refracting case where there are multiple rays passing through a sampling area, the effect of turbulence can be approximated by the reduction in coherence between the rays using the mutual coherence function.[8] However, as is common in all geometrical ray tracing methods, the procedure will fail to predict the sound pressure in the shadow zone of an upward refracting atmosphere. By definition no geometrical rays can penetrate into the shadow. In literature, several solutions to this problem have been suggested. In the simple case of an atmosphere with a linear vertical sound speed profile and has no turbulence, the sound field can be represented by an analytical normal mode solution. The residue series solution[10] can then be derived for an upward refracting case. However, this is not a practical solution in view of its limitations concerning the sound speed profile and turbulence. The comparisons published by L'Espérance *et al.*[8] show that it can substantially over-estimate the attenuation in real life situations where turbulence is present. For realistic atmospheric conditions that include turbulence, a common approach is to introduce empirical scattering to the ray trajectory[9] or to use empirical estimates of the sound pressure.[11] The dependence on empirical values significantly limits the applicability of such approaches. For example, in the ray model of Heimann and Gross,[9] the effect of shadow diffraction is simulated by randomly changing the propagation direction of a ray when the ray is close to the ground, where diffraction is thought to be most important. Unfortunately, the physical mechanism for this ray diversion is not fully formulated. Instead, the implementation relies on empirical constants and assumed dependence on the height from the ground. Although the model worked well in some cases,[9] it is doubtful that it will work equally well in situations outside the range of conditions under which the empirical parameters were derived without a full understanding of the physics involved. In this paper, we propose to address this problem by means of an analytical ray diffraction model.

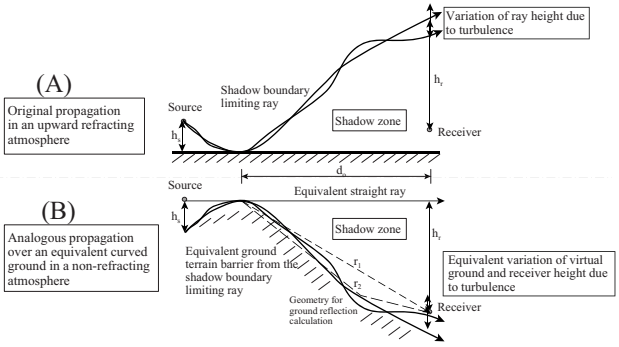


FIG. 1. Turbulence scattering of the analogous shadow boundary diffraction model. Upper half is the scattering of the original shadow boundary ray path over a flat ground. Lower half is the resulting scattering in the analogous model of a curved surface in a homogeneous atmosphere.

III. SHADOW BOUNDARY RAY DIFFRACTION MODEL

In here, we use a shadow ray diffraction model that is based on the analogy between a curved ray path in the presence of refraction, and a curved surface in a homogeneous atmosphere in which there is no refraction. This analogy has been used successfully in studies involving linear sound speed profiles and analogous cylindrical surfaces.[14] In here we apply the concept to an arbitrary sound speed profile and a generalized terrain with finite impedance cover. The analogy is illustrated in Fig. 1.

In the analogous virtual barrier configuration, the sound level at the receiver is calculated from the diffraction path from the barrier top to the receiver by applying simple thin barrier diffraction formulas. Here, one significant point to note is that the path difference cannot be calculated from the physical source location in the analogous barrier configuration. This is obvious if one considers the special case when the source is close to the barrier top (or the ground reflection point of the shadow boundary ray). An example is a strong upward refracting atmosphere where the shadow boundary is close to the source. In this case the path difference between the direct source to receiver path and the diffracted path over the barrier top is always small no matter how deep the receiver is in the shadow. This will give rise to only small attenuation even deep in the shadow. This is clearly incorrect. One would expect the attenuation to be higher with a stronger upward refracting atmosphere and a shadow boundary closer to the source. The explanation for this odd behav-

ior is that the analogy is based on geometrical ray acoustics. Since a ray is, in fact, an approximation of a plane wave, the "equivalent straight ray" in the analogous virtual barrier configuration should be considered as a plane wave. The propagation should therefore be from a source that is far away. Hence the path difference should be calculated with the source placed at infinity, which is equivalent to calculating the difference between r_1 and d_0 in Fig. 1.

In addition to the shadow boundary diffraction, the sound level at the receiver is also affected by attenuation due to ground reflection. This is clearly the case if one considers the sound level just before and after the shadow boundary. Before the boundary both direct and ground reflected rays are present and therefore the sound level will suffer ground attenuation, which can be as much as 20 dB at frequencies where ground attenuation is strongest. Just after the boundary, the shadowing (virtual barrier) effect is still small, and the attenuation will be of the order of only 5 dB (the usual line-of-sight barrier attenuation). Obviously, one would not expect the ground attenuation to just disappear when the receiver passes from one side to the other of the shadow boundary into the shadow. Hence it is necessary to correct for ground attenuation in the analogous virtual barrier model as well. As a first approximation, the ground attenuation is determined simply from the geometry of the geometrical reflection from the barrier top to the receiver. The local slope of the ground near the reflection point is used to correct for the ground reflection angle.

The prediction from this shadow boundary diffraction model is compared against a standard Crank–Nicolson PE[15,16] prediction using the benchmark case 3.[17] This benchmark case has a flat ground with finite impedance. The atmosphere is strongly upward refracting with a linear sound speed profile of $-0.1/\mathrm{s}$, but has no turbulence. The refraction is very strong. The shadow boundary at the receiver height is about 300 m from the source. The PE solution is a well established technique for outdoor sound propagation.[17] It has been shown to have very good accuracy in the benchmark cases.[17] It is therefore chosen to be the reference in this and subsequent comparisons. Figures 2(a), 2(b), and 2(c) show the comparisons at the frequencies of 10, 100, and 1000 Hz, respectively. The y-axis is in terms of the transmission loss (TL) as defined in the benchmark paper.[17] The lower limit of the y-axis is set at −80 dB as is set in the benchmark paper. The horizontal distance is shown to be 10 km, which covers a large range deep into the shadow. Also shown in the figures is the predicted sound pressure in free field in a still atmosphere (with air absorption). This is to allow the attenuation due to ground effect and atmospheric refraction alone to be easily identified. The figures show that the simple model proposed here for the shadow diffraction effect has good accuracy in this benchmark case. On the whole, the prediction agrees well with the prediction by the PE. The increases in the TL due to ground attenuation [most dominant at 100 Hz in Fig. 2(b)] and due to frequency [Fig. 2(c)] are well predicted. Given the simplicity of the shadow diffraction model, the agreement is quite remarkable. The simple diffraction model produces some artifacts at around the shadow boundary (at around 300–400 m). This is due to the simple ground correction used in the model. The equivalent ground is circular in this benchmark case. Since the ground reflection angle is determined by simple straight lines (see Fig. 1), the angle changes markedly near the shadow boundary (top of the circular arc), thus producing some sudden changes in the sound attenuation. A more elaborate diffraction model that takes into account the ground curvature could improve the accuracy. Nevertheless, the simple model is thought to be close enough to the PE solution to be usable in practice.

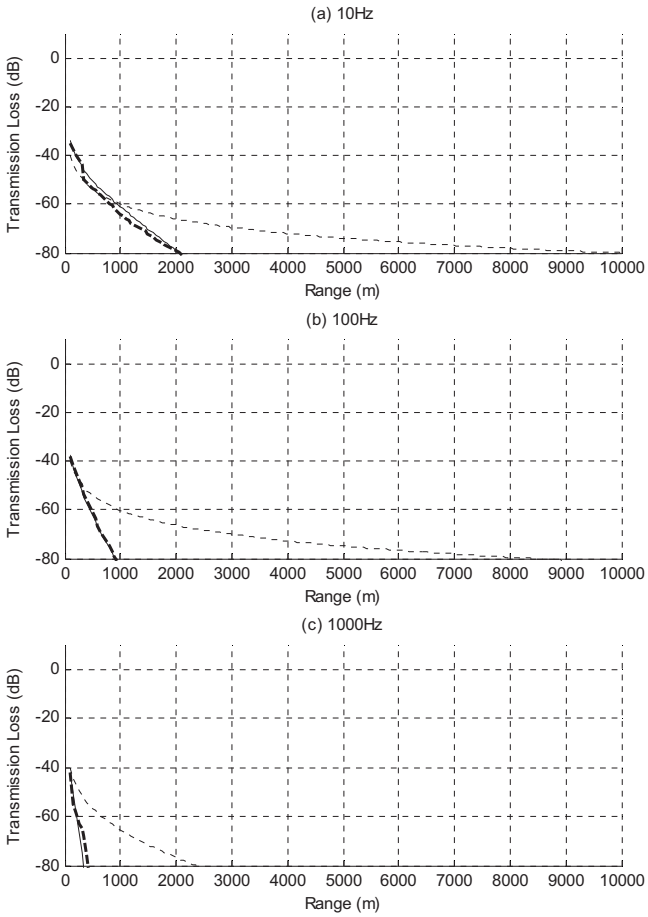


FIG. 2. TL versus distance predicted by the shadow boundary ray diffraction model (broken line) and PE (solid line) for the upward refracting benchmark case 3. The dotted line is for free field in a homogeneous still atmosphere.

IV. TURBULENCE SCATTERING

The ray based shadow boundary diffraction model of Sec. III provides a convenient basis for extending the calculation to account for the effect of turbulence scattering in the shadow zone. Turbulence can be simulated as temporal and spatial fluctuations of the sound speed along the propagation path. Let the refractive index be $n=c_0/c$. c_0 is a nominal constant sound speed, typically taken as the mean speed on the ground. c is the instantaneous sound speed at the position of interest. The effect of turbulence can be represented as the fluctuating part, μ, of the refractive index at the position of interest,

$$n = \bar{n} + \mu, \tag{1}$$

where $\bar{n}=c_0/\bar{c}$. $\bar{c}$ is the mean sound speed at that position, which is taken to be the value in the absence of turbulence. The fluctuation μ can be simulated by random realizations, which can be generated based on known models of the turbulence structure. Once μ is realized, Eq. (1) can be used to calculate the variation in the sound speed gradient due to turbulence within each of the grid cells in the numerical ray tracing algorithm. The ray trajectories will be scattered by these variations. When averaged over a large number of realizations, the result could be used to simulate the scattering of the sound energy by turbulence. For shadow zone calculation, which is the main interest of this paper, the scattering of the shadow boundary ray produces variations of the analogous equivalent diffraction geometry that will in turn change the attenuation detected at the receiver. However, even though ray tracing is faster than other numerical techniques such as the PE solution, performing a complete set of ray tracing for each random realization over a large number of realizations is still too time-consuming for practical use. We will therefore proceed to develop a model that can simulate the scattering of the rays at the receiver without tracing the rays through all the realizations.

A. Sound speed gradient due to turbulence

Since the numerical ray tracing assumes that the sound speed gradient is linear within each small grid cell in the propagation domain, we will need to relate the change in the gradient to the turbulence fluctuation μ. Rewriting Eq. (1) explicitly in terms of sound speeds, we have

$$\frac{c}{c_0} = \frac{1}{\frac{c_0}{\bar{c}} + \mu} \cong \frac{\bar{c}}{c_0}\left(1 - \frac{\bar{c}}{c_0}\mu\right) \quad \text{for } \mu \ll 1. \tag{2}$$

The vertical sound speed gradient is given by

$$\frac{\partial c}{\partial z} = \frac{\partial \bar{c}}{\partial z} - \left(\frac{\bar{c}^2}{c_0}\frac{\partial \mu}{\partial z} + \frac{2\mu\bar{c}}{c_0}\frac{\partial \bar{c}}{\partial z}\right). \tag{3}$$

In the absence of turbulence, the grid cells are assumed to be small enough that the vertical sound speed gradient within each is approximately linear. Therefore,

$$\bar{c} = \bar{c}_1(1 + a_0 z), \tag{4}$$

where $\bar{c}_1$ is the mean sound speed at the lower boundary of the cell, and a_0 is the linear gradient in the absence of turbulence. Generally μ is small and we may ignore the second term in the parentheses in Eq. (3) compared with the first term in the parentheses. Also, a_0 and the cell height are small enough such that $\bar{c}^2/c_0 \cong \bar{c}_1$. The effect of turbulence is therefore to modify the linear gradient by an extra term. The total gradient, a, is then given by

$$a = \frac{\partial c}{\partial z} \cong a_0 - \bar{c}_1\frac{\partial \mu}{\partial z}. \tag{5}$$

In the shadow diffraction calculation, the refracting atmosphere is replaced with a still atmosphere with a modified curved ground terrain. In this analogous case, a_0 and $\bar{c}_1$ in Eq. (5) are 0 and c_0, respectively.

The value of μ and its derivative can be determined for specific turbulence structures. Here we will use the simple Gaussian turbulence model.[18,19] Although the Gaussian model is not strictly correct for real atmospheres, it can be adapted to specific frequency ranges to give acceptable results.[20] Later, we will also show that the Gaussian model provides a convenient basis to simplify the simulation.

The random realization of $\mu(\mathbf{r})$ where the position vector in 2 dimension is (x,z) can be obtained by a spectral decomposition[21]

$$\mu(\mathbf{r}) = \sqrt{4\pi\Delta k}\sum_{n=1}^{N} \cos(\mathbf{k}_n \cdot \mathbf{r} + \alpha_n)\sqrt{F(\mathbf{k}_n)n\Delta k}, \tag{6}$$

where $\mathbf{k}_n=(k_{nx},k_{nz})=(k_n\cos\theta_n, k_n\sin\theta_n)$ and $k_n=n\Delta k$ for $n=1,2,\ldots,N$. N is the number of harmonics (or modes) used in the superposition, θ_n and α_n are independent random polar angles between 0 and 2π, and Δk is the spectral wavenumber resolution.

For a Gaussian turbulence model, the spectral density function depends on the standard deviations of the temperature and longitudinal wind velocity fluctuations, and the correlation length of the Gaussian spectrum ℓ. As an approximation, the standard deviation terms are commonly represented by a single μ_0^2, the square of the standard deviation of the fluctuating part of the index of refraction. We can then write

$$F(k_x,k_z) = \mu_0^2\frac{\ell^2}{4\pi}e^{-(k_x^2+k_z^2)\ell^2/4}. \tag{7}$$

The derivative of $\mu(\mathbf{r})$ is then

$$\frac{\partial \mu(\mathbf{r})}{\partial z} = -\sqrt{4\pi\Delta k}\sum_{n=1}^{N} k_{nz}\sin(\mathbf{k}_n \cdot \mathbf{r} + \alpha_n)\sqrt{F(\mathbf{k}_n)n\Delta k}. \tag{8}$$

This equation can be used to determine the sound speed gradient within a grid cell in the presence of turbulence according to Eq. (5).

B. Turbulence scattered rays

In the presence of turbulence, the shadow boundary ray is scattered as it propagates through the atmosphere. This is illustrated in the upper part of Fig. 1. Each random realization will produce a different scattered path. In principle, an analogous equivalent curved ground can be used to calculate the diffraction from each of this scattered boundary ray. However, this will be very time-consuming since a complete ray tracing will have to be done for each realization. Instead, we note that the dominant effect on the diffraction amplitude is due to the change in the height, denoted by h_r in Fig. 1, between the shadow boundary and the receiver. Hence our objective is to find a simple way to simulate this change in height due to turbulence.

Consider the propagation of a ray through a randomly realized turbulent atmosphere. The ray is traced numerically through a series of grid cells within each the sound speed gradient is approximately linear, as calculated by Eq. (5) in

the presence of turbulence. Note that the grid cell size in this case is set at 1/10 of the Gaussian correlation length ℓ to make sure that the randomly realized refraction index fluctuations are properly sampled.

It can be shown[8] analytically that the horizontal angle at which the ray enters the cell, ϕ_1, is related to the angle at which the ray exits the cell, ϕ_2, by the following equations:

$$\sin \phi_2 = \sin \phi_1 - \frac{x_2 - x_1}{R} \tag{9}$$

and

$$\cos \phi_2 = \cos \phi_1 + \frac{z_2 - z_1}{R}, \tag{10}$$

where (x_1, z_1) and (x_2, z_2) are, respectively, the coordinates of the entry and exit points. R is the radius of the ray curvature given by

$$R = \frac{1}{a \cos \phi_1}. \tag{11}$$

The change in height, $\Delta z = z_2 - z_1$, after the ray passes through this cell, is then

$$\Delta z = R(\cos \phi_2 - \cos \phi_1). \tag{12}$$

Now, let us consider the analogous curved ground diffraction model, in which the shadow boundary ray is equivalent to a straight ray in the absence of turbulence. In this model, the effective linear sound speed gradient in each of the grid cells is created by the turbulence alone, i.e., $a_0 = 0$ in Eq. (5), and is the sole cause of the scattering and the change in ray height. The ray horizontal angle will be small, especially in long range propagation, and we can approximate Eqs. (12) and (9) by

$$\Delta z \cong R\left(\frac{\phi_1^2 - \phi_2^2}{2}\right) \cong \Delta x\left(\frac{\phi_2 + \phi_1}{2}\right), \tag{13}$$

where $\Delta x = x_2 - x_1$. In the numerical ray tracing procedure, each cell has the same length Δx. For the ith cell, and for small horizontal angles, Eq. (9) gives

$$\phi_{i+1} \cong \phi_i - \frac{\Delta x}{R_i}. \tag{14}$$

Note that the entry horizontal angle into cell i is the same as the exit horizontal angle from cell $i-1$. Applying this formula recursively, and taking the initial horizontal angle of the ray at source to be ϕ_0, we have

$$\phi_{i+1} \cong \phi_0 - \sum_{j=1}^{i} \frac{\Delta x}{R_j}. \tag{15}$$

Substituting this into Eq. (13) for the ith cell gives the cumulative height change at the exit of cell i to be

$$\begin{aligned}\Delta z_{i+1} &\cong \Delta x \frac{(\phi_{i+1} + \phi_i)}{2} \\ &\cong \phi_0 + \frac{\Delta x}{2}\left(\sum_{j=1}^{i} \frac{-\Delta x}{R_j} + \sum_{j=1}^{i-1} \frac{-\Delta x}{R_j}\right) \\ &= \phi_0 - \Delta x^2\left(\sum_{j=1}^{i} \frac{1}{R_j} - \frac{1}{2R_i}\right).\end{aligned} \tag{16}$$

Again, taking the assumption that the horizontal angles are small, we can approximate Eq. (11) by $R_j \approx 1/a_j$. Equation (16) then becomes

$$\Delta z_{i+1} \cong \phi_0 - \Delta x^2\left(\sum_{j=1}^{i} a_j - \frac{a_i}{2}\right). \tag{17}$$

The gradient a_i is different for each cell. However, for a homogenous turbulence structure, all the a_i's are determined from the same turbulence model, Eqs. (5) and (8). They are all proportional to the turbulence strength represented by the square root of μ_0^2, i.e., the square of the standard deviation of the fluctuating part of the index of refraction for a Gaussian turbulence model. Hence we can write

$$a_j = q_j \sqrt{\mu_0^2}, \tag{18}$$

where the proportional factor q_j is determined from Eq. (8) for each cell. We can now rewrite Eq. (17) as

$$\Delta z_{i+1} \cong \phi_0 - \sqrt{\mu_0^2}\Delta x^2\left(\sum_{j=1}^{i} q_j - \frac{q_i}{2}\right). \tag{19}$$

This equation is independent of frequency. Moreover, all the variables in the equation except μ_0^2 are independent of the turbulence strength. For long range propagation, and for the analogous shadow boundary diffraction model, the initial horizontal ray angle ϕ_0 is small and close to zero. This means that the ray tracing simulation needs only be done once, for one turbulence strength, and the result can be applied to other turbulence strength through a simple scaling. Although the factor q_j is dependent on the Gaussian correlation length ℓ, in practice, ℓ is generally taken as a constant 1.1 m in a Gaussian model.

C. Changes in shadow ray height due to turbulence

Since we are using a Gaussian turbulence model to generate the random realizations, we expect the resulting change in height (Δz) to exhibit Gaussian behavior. Figure 3 shows the result of a simulation of the scattering of a horizontal ray through a Gaussian turbulence atmosphere with μ_0^2 set to 3×10^{-6}. The correlation length ℓ is set to 1.1 m in this and all subsequent calculations. The result is from 100 random realizations of the turbulent atmosphere. The size of each grid cell is set at 0.11 m, which is 1/10 of the Gaussian correlation length ℓ to make sure that the randomly realized refraction index fluctuations are properly sampled. The changes in the ray height due to turbulence scattering at a distance of 1 km were recorded. The cumulative probability distribution function (CDF) of the height change is plotted in the figure. It can be seen that the CDF looks similar to that of

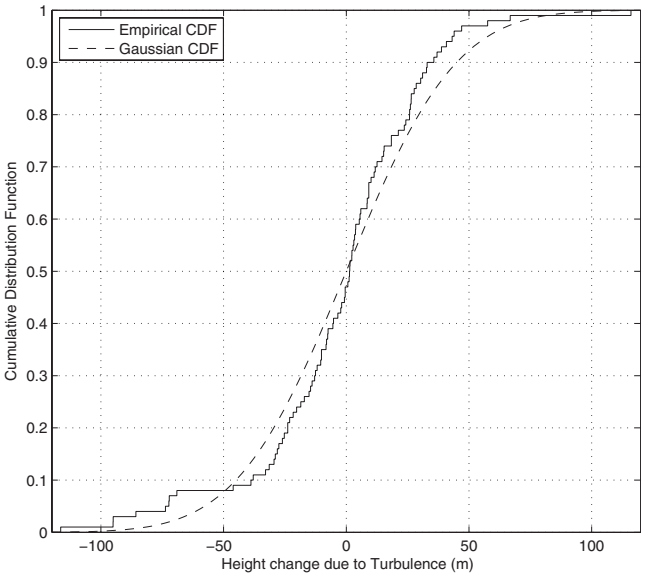


FIG. 3. The CDF of the simulated height change (solid line) of a horizontal ray after propagated 1 km through an atmosphere with a Gaussian turbulence structure using 100 random realizations. The dotted line is that of a standard Gaussian distribution.

an ideal Gaussian distribution of zero mean. From here on, we assume that we could take the distribution of height change (Δz) in Eq. (19) over many random realizations to be Gaussian with zero mean, and can be characterized by its standard deviation.

We now need to determine the dependence of Δz on range. This comes out from the term in parentheses on the right hand side of Eq. (19). Again, the propagation of a horizontal ray through 100 random realizations was used to simulate the height changes due to turbulence. The turbulence model is again Gaussian, with μ_0^2 set at 3×10^{-6}. The values of the standard deviation of the height change at every 20 m up to 10 km were calculated from the simulated realizations. Note that 0 m corresponds to the start of the shadow zone. Figure 4 shows the result. The standard deviation's dependency with range is smooth, and is almost linear on a log-log scale. The dependency is almost linear at short to medium range up to 1 km. At longer range the standard deviation grows faster and a higher order curve is needed. It is

FIG. 4. Dependence of the standard deviation of the scattered height statistics with range. The turbulence model used to generate the 100 random realizations is the same as that of Fig. 3.

TABLE I. Coefficients for the estimation of the standard deviation of height changes in Eq. (20).

Coefficient	a_{11}	a_{31}	a_{32}	a_{33}
Value	0.13	0.10	3.2×10^{-5}	-1.2×10^{-9}

apparent that, up to the range of 10 km, the dependency on range can be represented easily by simple regression curves. This is a useful result since a simple regression curve would allow us to calculate the standard deviation at any range without needing to perform time-consuming ray tracing through many random realizations again. Note that the size of each grid cell in the numerical ray tracing is again set at 0.11 m, 1/10 of the Gaussian correlation length ℓ, in order to sample the turbulence properly. Since the maximum range in the simulation is 10 km, this requires calculation of ray refractions over approximately 91 000 cells for each random realization. This is obviously time-consuming, and the ability to replace this with a simple regression formula is significant.

As a simple choice, we used a linear regression for the range from 0 to 1 km, and then a third order regression from 1 to 10 km. The resulting coefficients are listed in Table I. An equation for the standard deviation of the height change, Δz_{STD}, for near horizontal propagation ($\phi_0 \approx 0$) can now be written as

$$\Delta z_{\mathrm{STD}}(x) = \sqrt{\frac{\mu_0^2}{\mu_{\mathrm{ref}}^2}} a_{11} x \quad \text{for } 0 < x \le 1 \text{ km}$$

$$= \sqrt{\frac{\mu_0^2}{\mu_{\mathrm{ref}}^2}} (a_{31} x + a_{32} x^2 + a_{33} x^3) \quad \text{for } x > 1 \text{ km}. \tag{20}$$

The scaling, with respect to the turbulence strength, is taken from Eq. (19). The reference turbulence strength is set at $\mu_{\mathrm{ref}}^2 = 3\times10^{-5}$ to give more convenient values to the coefficients.

Equations (19) and (20) predict that Δz can be described by a Gaussian distribution and that it is proportional to $\sqrt{\mu_0^2}$. To test the dependence of the height changes on the strength of turbulence, ray tracing simulations were repeated over a range of μ_0^2 values from 1×10^{-6} to 7×10^{-5}. The standard deviations of the height changes at a number of distances from the shadow boundary are plotted against the square root of μ_0^2 in Fig. 5. It can be seen that the standard deviation does indeed scale very well with $\sqrt{\mu_0^2}$ at all the ranges. Also shown in the graph as dotted lines are the standard deviations that are calculated simply from Eq. (20). The prediction from the simple formula matches fairly well with the simulated result from the full numerical ray tracing. The error is largely less than 10%. In the 500 m case where the height change is small the error can go up to 30%. There is also a tendency for the simulated standard deviation to be slightly larger than that from Eq. (20) at the extremes: when the range is >2 km (very deep in shadow) and the turbulence strength is very high ($>5\times10^{-5}$). Overall, the simple equation is a good replacement of the time-consuming ray tracing simulation under realistic conditions.

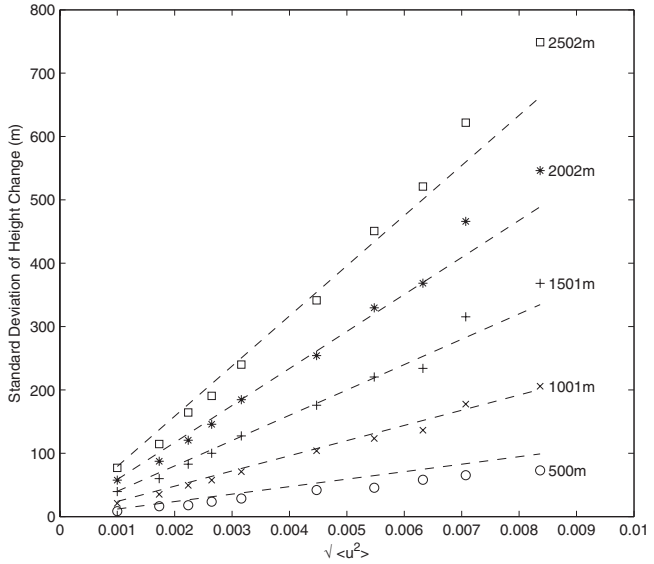


FIG. 5. Variation of the standard deviation of the simulated height changes against the square root of μ_0^2 of the Gaussian type turbulence. The symbols are results from numerical ray tracing using 100 random realizations for each turbulence strength. The dashed lines are direct calculations from the approximate formula.

V. TURBULENCE SCATTERED SHADOW RAY (TSSR) MODEL

With the results from Sec. IV, we can now calculate the shadow zone sound pressure in the presence of turbulence by the following steps.

(1) The limiting ray that defines the shadow boundary in an upward refracting atmosphere is first determined by a single numerical ray tracing process in the absence of turbulence.
(2) The shadow boundary geometry is replaced with an equivalent configuration of a straight horizontal ray passing just above the top of a curved ground with geometry defined by the original curve shadow boundary ray, as shown in Fig. 1. This horizontal ray effectively defines the line of sight in the modified ground geometry. In the absence of turbulence, the sound pressure level in the shadow zone is simply calculated as a barrier diffraction problem with an impedance ground.
(3) In the presence of turbulence, the turbulence structure is modeled as Gaussian, with the strength represented by μ_0^2. The correlation length of the Gaussian spectrum ℓ is assumed to be fixed and equal to 1.1 m. The effect of turbulence is to be represented by the scattering of the ray height above the receiver.
(4) Equation (20) allows the standard deviation of the ray height changes, Δz_{STD}, due to turbulence to be calculated at any range up to 10 km from the shadow boundary, and for any turbulence strength μ_0^2.
(5) Δz_{STD} is then used to generate random height changes from a Gaussian distribution with zero mean. A positive height change increases the height of the ray and therefore extends the height separation between the receiver and the line of sight. This will then increase the attenuation due to diffraction. Conversely, a negative height change decreases the height of the ray and therefore reduces the height separation between the receiver and the line of sight. The will then decrease the attenuation due to diffraction. With each height change, a new diffracted sound pressure level is calculated from the modified diffraction geometry.
(6) Finally, the attenuation values calculated from all the realizations are energy averaged to give the mean attenuation due to shadowing in the presence of turbulence.

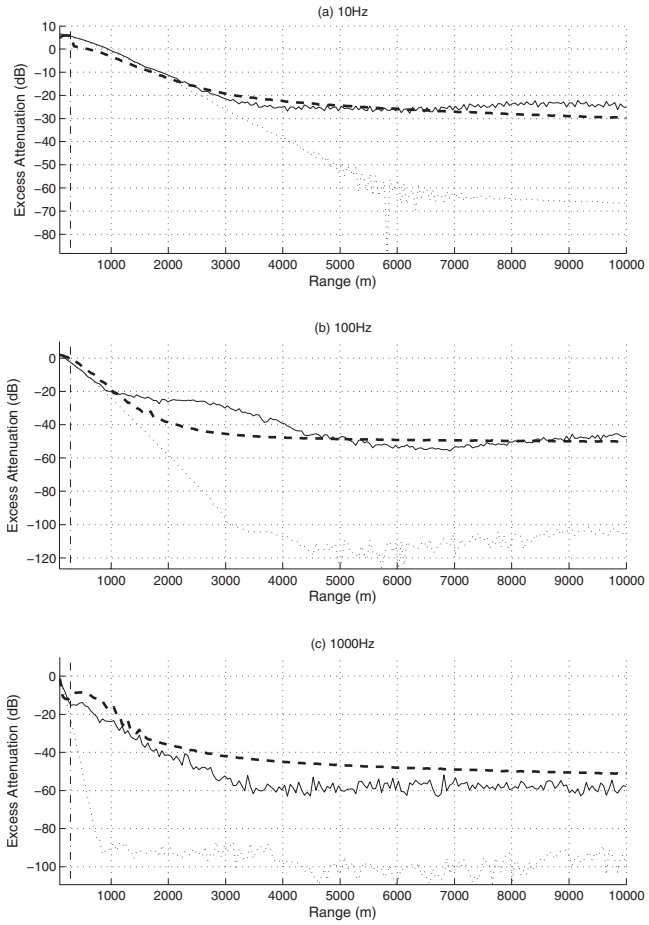


FIG. 6. Excess attenuation versus distance predicted by the TSSR model (broken line) and PE (solid line) for the modified upward refracting benchmark case 3 with turbulence added. A Gaussian turbulence model is assumed with a strength of $\mu_0^2=3\times10^{-6}$. The dotted line is the PE prediction for no turbulence. The vertical chained line indicates the estimated shadow boundary at 300 m.

For simplicity, we abbreviate this procedure as the TSSR model. To test the accuracy of the model, it was used to calculate the sound pressure level in the shadow zone of benchmark case 3 with the addition of turbulence. Since the published benchmark case result does not include turbulence, it is necessary to compare the result with a numerical calculation that has established accuracy. In this case, the standard PE solution with the addition of the randomization procedure of Gilbert *et al.*[1] for the simulation of turbulence is used as the reference. The same Gaussian turbulence structure is used in both the PE calculation and the ray tracing calculation. The result for a moderate turbulence strength of $\mu_0^2=3\times10^{-6}$ is shown in Fig. 6. The PE prediction with no turbu-

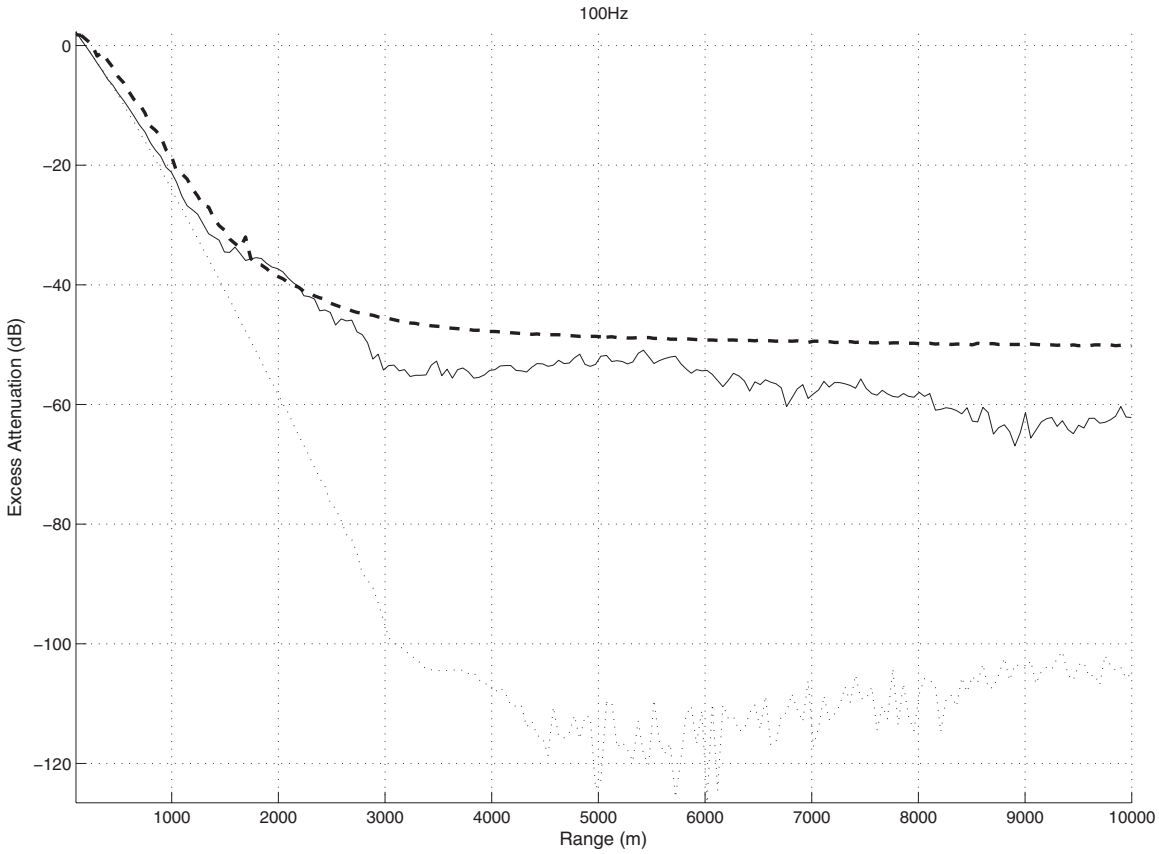


FIG. 7. Same as Fig. 6(b) but with a Gaussian correlation length of 4 m.

lence is also shown in the figure to let the effect of turbulence to be seen clearly. As mentioned before, this benchmark case is strongly upward refracting, with the shadow boundary starts at about 300 m from the source, as shown by the vertical chained line in the figure. Most of the range shown in the figure is therefore in the shadow zone. At the lowest frequency, 10 Hz, a sharp transition can be seen at the shadow boundary at around 300 m where the calculation changes from the normal ray tracing calculation before the shadow boundary to the TSSR calculation beyond the shadow boundary. The sharp change is of the order of 5 dB. This change can be easily explained. Before the shadow, the ground reflection detected by the ray tracing will be largely in phase with the direct sound at this low frequency. Hence the sound level increases by around 6 dB due to the ground. Immediately behind the shadow boundary, the calculation switches to the TSSR model since no rays arrive. Right at the boundary, the diffracted ray is at line of sight between source and the receiver. The attenuation given by the geometrical diffraction model at line of sight is about 5 dB. Hence a sudden increase in attenuation of about 5 dB is seen when the calculation switches to the TSSR model at the shadow boundary. Note that the 6 dB pressure level increase due to ground reflection is still retained as the TSSR model also accounts for the reflection by the ground behind the shadow boundary. Hence the drop is only 5 dB. If the ground effect is not included in the calculation behind the shadow, an 11 dB difference would have been seen. This shows the importance of accounting for the ground effect even in the shadow zone in this model. In principle, this sudden change can be avoided if the effect of diffraction is introduced gradually before the receiver hits the shadow boundary. However, since the ray tracing model is not aimed at low frequency applications, we have not made such adjustments in our current model. Further into the shadow, the TSSR prediction agrees very well with the PE prediction.

At the higher frequencies, 100 Hz and 1 kHz, the sudden change at the shadow boundary is no longer observable. This is because the ground reflection is no longer in phase with the direct sound at higher frequencies, and attenuation starts to appear even before the shadow. At 100 Hz, the TSSR prediction matches the PE prediction closely all the way up to about 1 km. From then on, the PE result shows an unexpected increase in level at ranges up to 4 km, then a slight decrease afterwards. This variation resembles some form of interference effect. Further testing of the PE suggests that this is related to the combination of the specific 100 Hz frequency and the particular Gaussian turbulence parameters used, and is exaggerated by the extremely large sound speed gradient of the benchmark case. For example, changing the Gaussian length scale from 1.1 to 4 m changes the shape notably, as shown in Fig. 7. Also, as shown later in Fig. 8, this spurious increase is not seen when a higher turbulence strength is used. Further investigation into this spurious behavior of the PE in combination with a Gaussian turbulence model is, however, outside the scope of this paper, although it could be an interesting subject for future work. Otherwise the overall trends of the excessive attenuation predicted by the two models appear to agree.

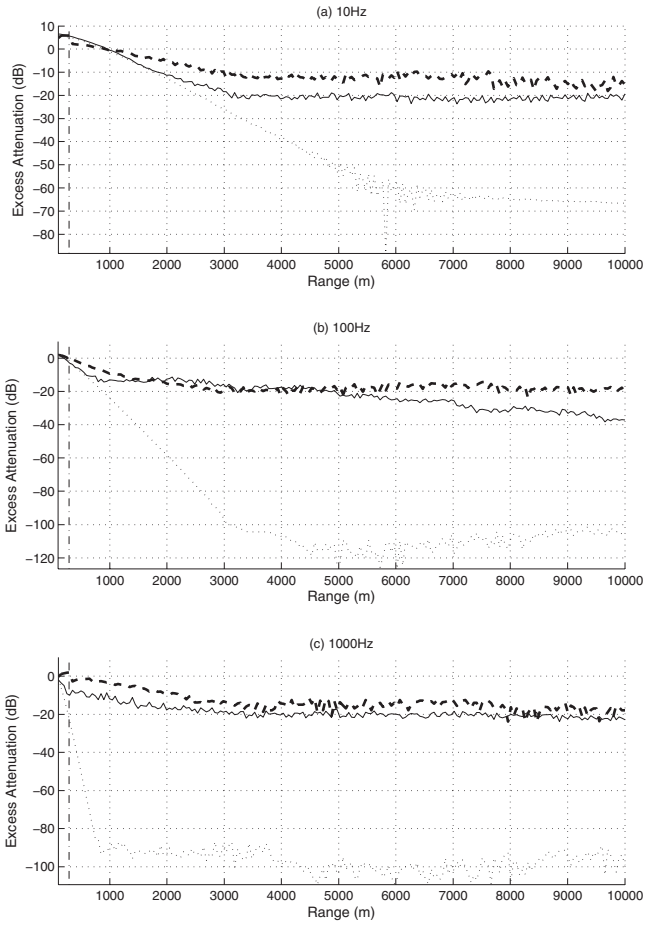


FIG. 8. Same as Fig. 6 but with $\mu_0^2=3\times10^{-5}$.

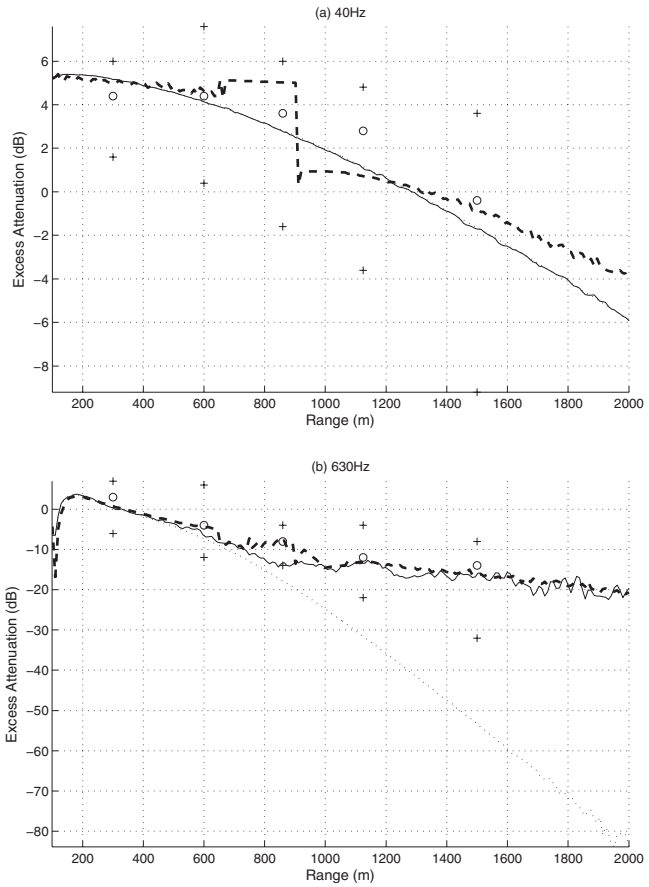


FIG. 9. Excess attenuation versus distance predicted by the TSSR model (thick broken line) and PE (solid line) against experimental data published by White and Gilbert (Ref. 22). The assumed sound speed gradient is $-0.034/\mathrm{s}$, and $\mu_0^2=1\times10^{-6}$. +, ○, and + represent the maximum, mean, and minimum experimental results of White and Gilbert (Ref. 22).

At 1 kHz, the TSSR predicts consistently smaller magnitude of attenuation compared with the PE prediction. The difference is around 5 dB up to 2 km, and increases to around 10 dB or more at longer ranges.

The effect of larger turbulence is shown in Fig. 8, where the turbulence strength is at a high level of $\mu_0^2=3\times10^{-5}$. This is considered to be the upper limit of turbulence that one would normally encounter. With this high level of turbulence, the attenuation in the shadow is severely limited, reaching just over -10 to -20 dB at 10 Hz, and around -20 dB at the higher frequencies even at 10 km into the shadow. At 10 Hz, the TSSR prediction is still fairly close to that of the PE model, but is consistently less than the PE attenuation from about 1 km onwards. At 100 Hz, the agreement is very good in the region up to about 5 km. The spurious increase in sound level between 1 and 4 km, seen earlier in the PE result in Fig. 6(b), at this frequency does not appear at this higher level of turbulence. From 5 km onwards, the TSSR predicted attenuation is again consistently less than that from the PE. A similar trend is also observed in the 1 kHz result. Overall, the results show that the TSSR prediction has good agreement with the PE prediction at short to medium range from the shadow boundary at all the test frequencies. At ranges further than 3.5 km, the TSSR predicts smaller attenuation values than the PE model.

As a further test of the reliability of the TSSR model, predictions were made and compared against the measured data of White and Gilbert.[22] These data had been used by L'Espérance *et al.*[8] in their development of a heuristic ray tracing model, although their comparison was not entirely successful due to the absence of a proper turbulence model in the shadow zone. Since the purpose of this paper is on the diffraction and turbulence effect in the shadow zone, only the upward refracting data are included in this comparison. The meteorological parameters used are the estimated values used by L'Espérance *et al.*[8] (see Figs. 11 and 12 of Ref. 8). The estimated linear sound speed gradients for the two upward refracting cases are $-0.034/\mathrm{s}$ and $-0.068/\mathrm{s}$. The corresponding values of μ_0^2 for the approximate Gaussian turbulence are 1×10^{-6} and 2×10^{-6}, respectively. The data set includes two frequencies—a low frequency of 40 Hz and a medium frequency of 630 Hz. The two cases are shown in Figs. 9 and 10. In both cases the TSSR model predictions agree well with the measured data, except at the shadow boundary at 40 Hz where a sharp change of about 5 dB is seen. This sharp change is due to the switch over at the boundary from a hard ground effect at low frequency to a diffraction calculation in the TSSR model. This has already been discussed in detail earlier. It is interesting to note that even with this sharp change, the predicted values are still within the minimum to maximum range of the measured values. This error is not seen at the higher frequency,

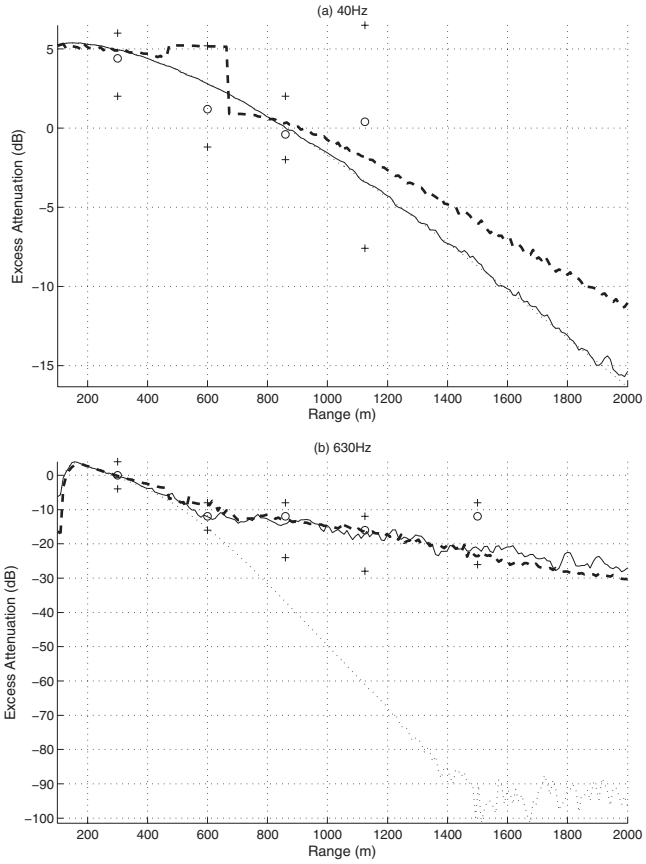


FIG. 10. Same as Fig. 9 but with sound speed gradient of $-0.068/\mathrm{s}$, and $\mu_0^2=2\times10^{-6}$.

630 Hz, where the agreement is very good. The agreement between the TSSR model and the PE model in these realistic cases is significantly better than that observed in the extreme benchmark cases shown in Figs. 6 and 8 within the corresponding horizontal ranges.

VI. CONCLUSIONS

In this paper, a simple analytical model has been developed for the scattering of rays due to turbulence in a shadow zone. The model is based on the assumption of a Gaussian turbulence structure, and makes use of the equivalence of ray propagation in a refractive atmosphere over flat ground and the propagation in a still atmosphere over a curved ground. The equations are derived analytically in all but one aspect. The equation for the standard deviation of the changes in ray height due to turbulence is obtained from statistics generated by numerical simulations. Nevertheless, this equation is shown to be valid for a horizontal distance of 10 km over a large range of turbulence strengths for an assumed Gaussian turbulence structure.

This simple analytical model has been tested in a simulated benchmark case and against previously published measured data at a variety of frequencies and turbulence strengths. The excess attenuation predicted by the model in the shadow zone has been shown to agree well with that predicted by a standard PE solution in all the test cases at distances up to 2 km from the shadow boundary. Right at the shadow boundary, there is an error of up to 5 dB at low frequency where the ground is effectively hard. This is due to the step change of attenuation from hard ground pressure doubling to a line-of-sight barrier screening effect in the model. This error does not appear at frequencies higher than 100 Hz in the test cases. Further into the shadow zone, from 2 to 10 km, the model tends to predict less excess attenuation than the PE.

The comparisons presented in this paper have shown that the TSSR model provides fairly accurate predictions of excess attenuation in the shadow zone at distances up to 2 km from the boundary under a variety of turbulence strengths. The formulation of the model is based on ray tracing, and is therefore well suited to such geometrical models, which up to now have problems dealing with attenuation in the shadow zone. The model can also be used as a simple, standalone prediction tool for the attenuation in a shadow zone due to turbulence.

[1]K. E. Gilbert, R. Raspet, and X. Di, "Calculation of turbulence effects in an upward-refracting atmosphere," J. Acoust. Soc. Am. **87**, 2428–2437 (1990).
[2]W. E. McBride, H. E. Bass, R. Raspet, and K. E. Gilbert, "Scattering of sound by atmospheric turbulence: Predictions in a refractive shadow zone," J. Acoust. Soc. Am. **91**, 1336–1340 (1992).
[3]A. L'Espérance, Y. Gabillet, and G. A. Daigle, "Outdoor sound propagation in the presence of atmospheric turbulence: Experiments and theoretical analysis with the fast field program algorithm," J. Acoust. Soc. Am. **98**, 570–579 (1995).
[4]K. E. Gilbert and X. Di, "A fast Green's function method for one-way sound propagation in the atmosphere," J. Acoust. Soc. Am. **94**, 2343–2352 (1993).
[5]A. D. Pierce, *Acoustics: An Introduction to Its Physical Principles and Applications* (McGraw-Hill, New York, 1981), pp. 371–413 and 469–478.
[6]E. M. Salomons, "Downwind propagation of sound in an atmosphere with a realistic sound-speed profile: A semianalytical ray model," J. Acoust. Soc. Am. **95**, 2425–2436 (1994).
[7]M. M. Boone and E. A. Vermaas, "A new ray-tracing algorithm for arbitrary inhomogeneous and moving media including caustics," J. Acoust. Soc. Am. **90**, 2109–2117 (1991).
[8]A. L'Espérance, P. Herzog, G. A. Daigle, and J. R. Nicolas, "Heuristic model for outdoor sound propagation based on an extension of the geometrical ray theory in the case of linear sound speed profile," Appl. Acoust. **37**, 111–139 (1992).
[9]D. Heimann and G. Gross, "Coupled simulation of meteorological parameters and sound level in a narrow valley," Appl. Acoust. **56**, 73–100 (1999).
[10]R. Raspet, G. E. Baird, and W. Wu, "The relationship between upward refraction above a complex impedance plane and the spherical wave evaluation for a homogeneous atmosphere," J. Acoust. Soc. Am. **89**, 107–114 (1991).
[11]J. Hofmann and K. Heutschi, "An engineering model for sound pressure in shadow zones based on numerical simulations," Acta Acust. **91**, 661–670 (2005).
[12]Y. W. Lam and M. R. Monazzam, "On the modeling of sound propagation over multi-impedance discontinuities using a semiempirical diffraction formulation," J. Acoust. Soc. Am. **120**, 686–698 (2006).
[13]C. F. Chien and W. W. Soroka, "Sound propagation along an impedance plane," J. Sound Vib. **43**, 9–20 (1975).
[14]A. Berry and G. A. Daigle, "Controlled experiments on the diffraction of sound by a curved surface," J. Acoust. Soc. Am. **83**, 2047–2058 (1988).
[15]K. E. Gilbert and M. J. White, "Application of the parabolic equation to sound propagation in a refracting atmosphere," J. Acoust. Soc. Am. **85**, 630–637 (1989).
[16]R. A. Sack and M. West, "A parabolic equation for sound propagation in two dimensions over any smooth terrain profile: The generalised terrain parabolic equation (GT-PE)," Appl. Acoust. **45**, 113–129 (1995).
[17]K. Attenborough, S. Taherzadeh, H. E. Bass, X. Di, R. Raspet, G. R. Becker, A. Güdesen, A. Chrestman, G. A. Daigle, A. L'Espérance, Y. Gabillet, K. E. Gilbert, Y. L. Li, M. J. White, P. Naz, J. M. Noble, and H. A. J. M. van Hoof, "Benchmark cases for outdoor sound propagation

models," J. Acoust. Soc. Am. **97**, 173–191 (1995).
[18]G. A. Daigle, J. E. Piercy, and T. F. W. Embleton, "Effects of atmospheric turbulence on the interference of sound waves near a hard boundary," J. Acoust. Soc. Am. **64**, 622–630 (1978).
[19]G. A. Daigle, J. E. Piercy, and T. F. W. Embleton, "Line-of-sight propagation through atmospheric turbulence near the ground," J. Acoust. Soc. Am. **74**, 1505–1513 (1983).
[20]D. K. Wilson, J. G. Brasseur, and K. E. Gilbert, "Acoustic scattering and the spectrum of atmospheric turbulence," J. Acoust. Soc. Am. **105**, 30–34 (1999).
[21]E. M. Salomons, *Computational Atmospheric Acoustics* (Kluwer, Dordrecht, 2001), pp. 215–227.
[22]M. J. White and K. E. Gilbert, "Application of the parabolic equation to the outdoor propagation of sound," Appl. Acoust. **27**, 227–238 (1989).

Flow-structure-acoustic interaction in a human voice model

Stefan Becker,[a] Stefan Kniesburges,[b] Stefan Müller,[c] and Antonio Delgado[d]
Institute of Fluid Mechanics (LSTM), Friedrich-Alexander University Erlangen-Nuremberg, Cauerstrasse 4, 91058 Erlangen, Germany

Gerhard Link[e] and Manfred Kaltenbacher[f]
Department of Sensor Technology (LSE), Friedrich-Alexander University Erlangen-Nuremberg, Paul-Gordan-Strasse 3/5, 91052 Erlangen, Germany

Michael Döllinger[g]
Department of Phoniatrics and Pediatric Audiology, University Hospital Erlangen, Medical School, Bohlenplatz 21, 91054 Erlangen, Germany

(Received 29 May 2008; revised 8 November 2008; accepted 11 December 2008)

For the investigation of the physical processes of human phonation, inhomogeneous synthetic vocal folds were developed to represent the full fluid-structure-acoustic coupling. They consisted of polyurethane rubber with a stiffness in the range of human vocal folds and were mounted in a channel, shaped like the vocal tract in the supraglottal region. This test facility permitted extensive observations of flow-induced vocal fold vibrations, the periodic flow field, and the acoustic signals in the far field of the channel. Detailed measurements were performed applying particle-image velocimetry, a laser-scanning vibrometer, a microphone, unsteady pressure sensors, and a hot-wire probe, with the aim of identifying the physical mechanisms in human phonation. The results support the existence of the Coanda effect during phonation, with the flow attaching to one vocal fold and separating from the other. This behavior is not linked to one vocal fold and changes stochastically from cycle to cycle. The oscillating flow field generates a tonal sound. The broadband noise is presumed to be caused by the interaction of the asymmetric flow with the downstream-facing surfaces of the vocal folds, analogous to trailing-edge noise.
 [DOI: 10.1121/1.3068444]

PACS number(s): 43.28.Ra, 43.70.Gr, 43.50.Nm [AK] Pages: 1351–1361

I. INTRODUCTION

The interaction between applied airflow and the innervated structure (i.e., oscillating vocal folds in the vocal tract, see Fig. 1) is the basis for human phonation and therefore for the resulting acoustic signal.[1] Increased subglottal pressure causes an airstream through the glottis and excites vocal fold oscillations. The airstream is modulated and forms the primary voice signal. Subsequently, the voice signal is filtered by the supraglottal tract and emitted as an acoustic signal through the mouth.[1]

Identifying and analyzing the basic physical mechanisms of sound production are of great interest for improving medical treatment and for developing artificial vocal fold implants in the future. On this account, many international groups are engaged in experimental and theoretical approaches in voice research, with the common objective of identifying the leading processes for the production of the basic tone in human phonation. Therefore it is essential to understand the behavior of the individual parts in this multifarious physic problem.

Experimental approaches can be divided into two main groups, those using rigid models[2] and those using oscillating models.[3,4]

The major goal for experimental workers using rigid models was the investigation of the flow field directly downstream of the vocal folds. Alipour *et al.*[2] measured the velocity distributions in rigid glottal models and compared them with the time-averaged velocity obtained in oscillating excised canine larynxes and also with numerical results. They observed an asymmetric flow field downstream of the glottal duct.

This asymmetry was also observed by Scherer *et al.*[5] and Shinwari *et al.*[6] who applied rigid vocal folds to investigate the pressure distribution of airflow within symmetric and oblique glottal ducts. They compared divergent cross sections at a steady inflow at different increasing transglottal pressures and visualized the flow through the synthetic laryngeal model. Both the visualization and the pressure measurements along the glottal duct showed asymmetric distributions even within the symmetric glottal duct. Hence the results supported the existence of the *Coanda effect*[7] for the symmetric and the oblique glottal shape.

Due to the unsteady character of the flow through the oscillating glottis Hofmans *et al.*[8] applied an unsteady flow

[a]Electronic mail: stefan.becker@lstm.uni-erlangen.de
[b]Electronic mail: stefan.kniesburges@lstm.uni-erlangen.de
[c]Electronic mail: stefan.mueller@lstm.uni-erlangen.de
[d]Electronic mail: antonio.delgado@lstm.uni-erlangen.de
[e]Electronic mail: gerhard.link@lse.eei.uni-erlangen.de
[f]Electronic mail: manfred.kaltenbacher@lse.eei.uni-erlangen.de
[g]Electronic mail: michael.doellinger@uk-erlangen.de

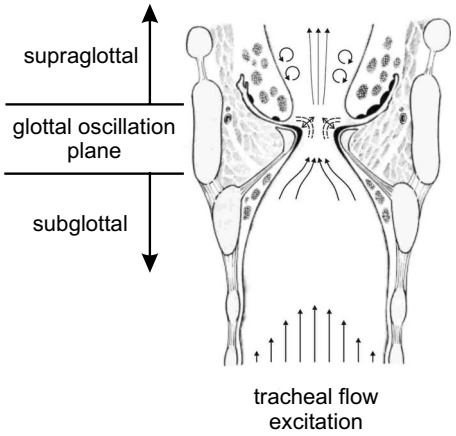


FIG. 1. Schematic of the vocal larynx.

passing through a rigid glottal constriction with three different pairs of vocal folds. During the experiments, the Coanda effect occurred and the laminar jet downstream of the constriction became turbulent, but these two effects developed very slowly. Therefore, for the dynamic process of oscillating vocal folds, the authors estimated a fully laminar and symmetric flow within and downstream of the vocal folds due to the unsteady flow behavior.

The conclusion of Hofmans *et al.*[8] that the Coanda effect and turbulence do not develop because of the long time needed was disproved partially by Erath and Plesniak[9,10] following a similar approach. They investigated static upscaled models of the vocal folds with symmetric divergent glottal ducts. In this work, the flow showed the Coanda effect for low divergent angles and furthermore indicated a bimodal flow state, i.e., the jet changed its attachment to a vocal fold in a stochastic manner between cycles. They concluded that this stochastic flow behavior causes the dipole nature of the basic tone of the human voice.

The rigid model approach could not, however, provide any certain information about the dynamic processes in the larynx during phonation. Thus, new experimental setups were developed, which included the oscillation of the vocal folds, either externally driven or as self-sustained oscillating vocal fold models.

The first approach was followed by Triep *et al.*[4] who developed an upscaled model of the vocal tract including externally driven synthetic vocal folds. The working fluid was water. Respecting the Reynolds and Strouhal numbers, the investigated flow field showed that the Coanda effect developed clearly. Additionally, large vortices were detected developing in the shear layer of the jet and shedding downstream. The flow field was determined as fully three dimensional. The vortex-shedding frequency was measured as five times higher than the fundamental frequency of the glottis displacement.

With a very similar setup Krane *et al.*[11] analyzed in more detail the unsteady behavior of the flow within an upscaled driven Plexiglas model, which also used water as the working fluid. They observed a delayed onset of flow after glottis opening and a flow breakdown before the total closure of the glottis occurred. Furthermore, they showed strong dependencies of the jet velocity, the volume flow, the vortex-shedding rate, and the amplitude on the frequency of the vocal fold oscillation behavior. They concluded that the oscillating flow is nonlinear in all phases of the vocal fold oscillations and so cannot be assumed to be quasisteady at any instant of the cycle. The results of Krane *et al.*[11] imply that all results obtained by using rigid vocal fold models for the flow field have to be regarded with skepticism concerning assumptions of the full fluid-structure-acoustic coupled process.

In this context Park and Mongeau[12] showed that the quasisteady assumption seems to be valid for over 70% of the oscillation cycle. Therefore they used a driven orifice model with typical cross sections of the glottal duct and compared the orifice discharge coefficients for steady and unsteady inflows. At the beginning and the end of a cycle the quasisteady assumptions became invalid because of an increasing influence of viscous effects and the large acceleration of the fluid and the synthetic vocal folds, which increased the unsteady behavior. However, Zhang *et al.*[13] measured the radiated sound pressure and found good agreement with the predicted sound pressure, which was calculated on the quasisteady assumption.

The above investigations show that for assumptions about the fully coupled process of human phonation rigid and driven models are only significant under certain conditions. Whereas the fluid-structure-acoustic interaction is totally neglected using rigid models, driven vocal folds fail to satisfy the energy balance between the three physical domains of fluid dynamics, structure dynamics, and acoustics. It is certainly essential to develop a test setup for investigating the human phonation to reproduce the fully coupled vocal fold vibration process including the acoustic sound production. Therefore, Thomson *et al.*[3] developed a synthetic flexible model of a vocal fold in a hemilaryngeal configuration. Applying a constant pressure gradient over the glottal constriction, the vocal fold vibrated with a frequency of about 89 Hz, which is almost in the range of typical human phonation frequencies.

In recent work, Thomson *et al.*[14] used a full glottis model with two opposing synthetic vocal folds of the same geometry and material as built before.[3] They were able to increase the oscillation frequency up to 120 Hz. In addition, they simulated the flow problem two dimensionally with a fully coupled numerical fluid-structure solver and obtained good agreement with the results of experiments. The numerical results allowed detailed analysis of the transfer of energy from the fluid to the moving structures by viscous forces and pressure distribution at the surface. They supported the assumptions that the periodic change from convergent to divergent shape of the glottal duct is necessary for self-sustained oscillations of the vocal folds.

Neubauer *et al.*[15] investigated the near flow field just downstream of synthetic vocal folds with the same physical properties, which were developed by Thomson *et al.*,[3] applying high-speed imaging, digital particle image velocimetry (DPIV) and statistical analysis, as well as principal component analysis (also called empirical orthogonal eigenfunctions[16,17]), which enabled them to determine coherent structures within a spatial-temporal dataset. They ob-

served that the Coanda effect was forced by an asymmetry of the glottal duct, caused by a released fixation of one vocal fold as the authors described. Furthermore the jet changed the vocal fold to which it is attached within one oscillation cycle. In addition, Neubauer *et al.*[15] detected a flapping of the jet, which they related to vortex production. This displayed the transition from laminar to turbulent flow and to vortex convection. They suggested that the flapping of the jet is mainly caused by the vortices produced in the shear layers of the jet.

The first aim of the present work is to reproduce the human phonation process conserving the energy balance between the fluid, structure, and acoustic domain. Therefore it was essential to produce self-sustained vocal fold vibrations within a fluid-structure interaction, which generate a detectable acoustic signal in the far field of the experimental setup. The authors are convinced that for imaging the human phonation process with an experimental setup, a strong tonal acoustic signal related to the basic tone of the human voice has to be produced by the flow-induced vibrations of the synthetic vocal folds.

Maintaining these basic conditions the second aim was to investigate the full fluid-structure-acoustic interaction within an experimental setup, which included synthetic vocal folds similar to the models developed by Thomson *et al.*[3] Therefore the process was investigated by high-speed visualization of the flow and the oscillating vocal folds. The flow field downstream of the artificial glottis was observed mainly by using DPIV measurement techniques.

In a further step correlations between flow and structure quantities and the acoustic pressure (AP) helped in identifying the basic mechanism of the sound generation. Three mechanisms important for sound production can be distinguished:

(I) the flow-induced structure vibration of the vocal folds,
(II) the resulting pulsating mass flow downstream of the glottis, and
(III) the stochastic turbulence in the flow region downstream of the vocal folds.

In this work, we neglected the first mechanism because of the very high damping properties of the vocal folds. With this experimental approach it was possible to identify the leading structures in the dynamic flow field. Further the correlation measurements gave a deeper insight in the processes of the human sound production. It was possible to identify the major acoustic source term.

II. EXPERIMENTAL SETUP

A. Physical parameters

To get a better understanding of the physical mechanisms of voice generation, a special test facility has been constructed. The basic design parameters of the synthetic vocal folds are identical to those in the models used by Thomson *et al.*[3] Table I provides the main characteristic parameters, stating the dimensions of the human larynx, the fluid parameters, and the dimensionless characteristic numbers.

TABLE I. Physical parameters for the investigation of the human voice.

Diameter of the larynx	$D=18-22$ mm
Maximal glottis opening	$w_{G,\max}=1-3$ mm
Pressure difference across the glottis	$\Delta p=500-5000$ Pa
Mean velocity	$U_{\text{mean}}=20-40$ m/s
Frequency	$f=100-200$ Hz
Reynolds number	$\text{Re}=\frac{U_{\text{mean}}w_{G,\max}}{\nu}=O(10^3)$
Mach number	$\text{Ma}=\frac{U_{\text{mean}}}{c}=O(10^{-1})$
Strouhal number	$\text{St}=\frac{fw_{G,\max}}{U_{\text{mean}}}=O(10^{-2})$
Young's modulus	$E=5-10$ kPa

B. Synthetic vocal folds

Figure 2 shows the design of the self-oscillating synthetic vocal folds. The models were made by casting a liquid polymer solution into a mold. The solution consists of a two-part polyurethane composition: cure polymer Evergreen™ 10 and Everflex™ for increased flexibility. The stiffness of the cured rubber was varied by adjusting the mixing ratio of the different compounds. The mixing ratio was set to 1: 3.33: 1 (Evergreen™ Part1: Everflex™: Evergreen™ Part2) to obtain synthetic folds with Young's modulus of $E=6.5$ kPa, which is in the range of natural human vocal folds.[3,18] However, the synthetic vocal folds, which were mounted in the test channel, vibrated only at very small amplitudes. Therefore, we additionally fabricated synthetic vocal folds with a small mass body cast into each tip of the folds. The spherical body consisted of lead with a weight of 0.433 g. Thereby the synthetic vocal folds have an inhomogeneous material distribution, which is also the case in real vocal folds.

C. Test channel

The geometric dimensions of the test channel were adapted to the human vocal tract. Figure 3 shows a schematic representation and also a photograph of the test rig. It consists of an unsteady mass flow controller,[19] which can deliver both a constant and a pulsating mass flow rate comparable to the human breathing process. A settling chamber with a nozzle is connected to the mass flow engine followed by the main test section, which is screwed over a flange. It contains a honeycomb smoothing grid, which has the function of set-

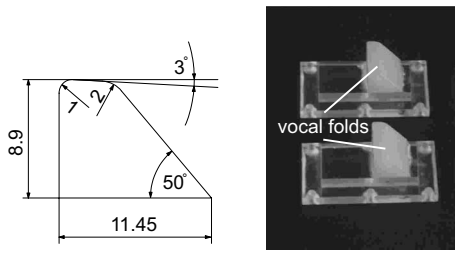


FIG. 2. Geometry of the synthetic vocal folds (dimensions in millimeters).

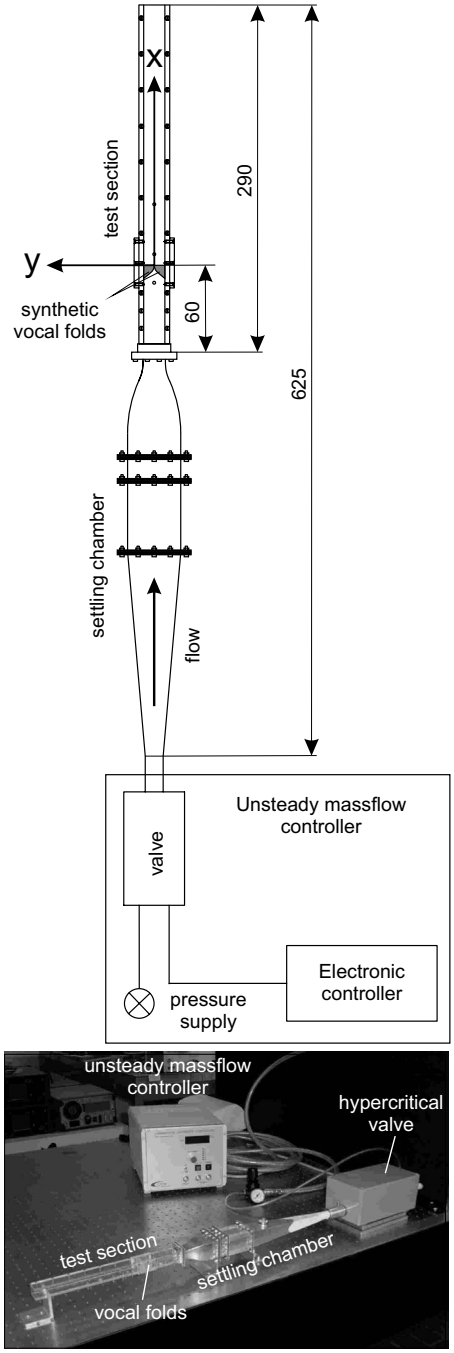


FIG. 3. Schematic representation and a photograph of the basic test rig (dimensions in millimeters).

tling down turbulence in the inflow of the main test channel produced by the flow engine. The main test section with a rectangular cross section of 15×17.8 mm^2 includes the synthetic vocal folds and consists of Plexiglas to allow optical access. Downstream of the glottal duct there is a rectangular extension of the same cross section as upstream with a length of 230 mm, which corresponds to a simplified model of the vocal tract. It guarantees that tracer particles do not disappear into the environment. Additionally the flow field develops in a more realistic way in comparison to the human physiology than a model without extension. The length of the synthetic vocal tract assured that the resonant back-coupling of standing acoustic waves can be neglected. Regarding the extension as a half open tube[20] the resonant frequencies developing a standing wave can be calculated by $f_n = nc_0/4L$ with the integer n, the speed of sound c_0, and the length of the tube L, which leads to the lowest resonant frequency of 373 Hz. This frequency is higher than the dominant frequencies in the present work by a factor of about 10. The synthetic vocal folds were fixed at three sides to the test channel with a two-component epoxy adhesive.

D. Measurement techniques

Several measurement techniques were used to investigate both the flow field behind the synthetic vocal folds and their oscillations, which are described in the following. Additionally, the coherence of all signals was calculated with the AP recorded by a microphone positioned outside the test channel.

1. Pressure measurement

The unsteady pressure distribution was measured 12 mm upstream (UPU) and 7 mm downstream (UPD) of the synthetic vocal folds. For this purpose, two pressure sensors of type *Sealed Gage 604* by Kulite[21] were used. Their signals were recorded at a sampling rate of 40 kHz. The sensors are mounted in the bottom plate, each in a hole with a diameter of 2.4 mm.

2. Optical measurements

By using the digital high-speed camera, *Speedcam Visario 1500*,[22] we performed measurements both for the flow field and the mechanical displacements of the vocal folds.

First, the flow field behind the vocal folds without the additional mass (so the oscillations are just a few micrometers) was recorded with a frame rate of 1 kHz and a resolution of 1538×1024 pixels. A Nd:YAG (yttrium aluminum garnet) laser with doubled frequency (wavelength $\lambda = 532$ nm) working in continuous mode was formed into a thin laser sheet by a cylindrical lens and coupled into the test channel via the outlet perpendicular to the synthetic glottis. Tracer particles (water droplets) produced by a common medical device for moistening the breathing air were inserted over a drilling in the divergent part of the settling chamber. The inner channel walls were cleaned after each measurement point because of adhesion of the water droplets, which obstructed the view on the laser sheet.

Second, the displacements of the vocal folds with additional mass were recorded with a frame rate of 4 kHz and a resolution of 768×512 pixels. To obtain information about the position within the oscillation cycle, the recordings were triggered by the unsteady pressure signal of the sensor upstream of the glottis. Because of the very high frame rate and the short exposure time for each picture, a special spotlight was applied to obtain very strong and glimmer-free light.

3. Particle-image velocimetry (PIV)

PIV measurements were applied to determine the flow field behind the oscillating vocal folds. For some basic explanations of the PIV measurement technique, we refer to Raffel *et al.*[23] Each of the two frequency-doubled Nd:YAG

lasers produced a thin laser sheet, which entered the test channel via the outlet toward the main flow direction. They were oriented perpendicular to the artificial glottis in the middle of the channel dividing the vocal folds in transversal direction. The lasers were operated in Q-switch mode. One pulse lasted between 5 and 10 ns and the lasers worked with a time delay of 2 μs. The generation of the tracer particles was done analogously to the high-speed visualization of the flow field except that the water was mixed with glycerin (mixing ratio water to glycerin 4000:1) to prevent the water droplets from evaporating. It was also necessary to clean the inner walls of the channel walls after each measurement point. A charge-coupled device camera was placed above the test channel to record the flow field just behind the synthetic vocal folds. The dimensions of the investigated area were 30.2×24.2 mm^2. The whole test system was controlled via the PIV processor, *FlowMap 2550* (Dantec[24]). The calculation of the vector plots of the flow velocity and the postprocessing was performed by the software program FLOWMANAGER (Dantec) and our own postprocessing tools based on MATLAB.

Since the observed phenomena indicate a periodically unsteady fluid-structure coupled problem, the PIV recordings were triggered by the unsteady pressure sensor, which was located 12 mm upstream of the artificial glottis. Thereby it was possible to determine the flow field at a certain phase angle within an oscillation cycle of the synthetic vocal folds.

4. Hot-wire anemometry/constant-temperature anemometry (CTA)

For the time-resolved determination of the turbulent velocity fluctuations, a one-wire probe was used. This probe was positioned in the test channel 35 mm behind the synthetic vocal folds in the center point of the cross section. The probe was operated in the constant-temperature mode, that is, the temperature of the wire and thus its electrical resistance are held constant. Therefore, the wire is inserted in a Wheatstone bridge circuit to readjust the resistance of the wire by adjusting the bridge. Applying a probe with a 5 μm diameter wire in the constant-temperature mode, a maximum fluctuation frequency of the flow velocity of 1.2 MHz can be resolved. The sampling rate during the measurements was 44 kHz. So, during each measurement, approximately 2.64×10^6 data points were recorded and stored in a file, which is equal to measurement time of 60 s. For more basic information about the CTA, especially the calibration of the system, see Ref. 25.

5. Laser-Doppler vibrometry (LDV)

To measure the oscillations of one synthetic vocal fold, the laser-Doppler vibrometer of Polytec[26] was used. The laser beam was coupled into the test channel via the outlet. The laser spot was located in the center of the frontal surface of one synthetic vocal fold. To improve the reflective properties of the synthetic fold, the frontal face was painted with a silver lacquer pen. The laser beam was reflected at the painted surfaces and interfered with a reference beam. Because of the interference signal due to the superposition of the Doppler-shifted scattered laser beam and the reference beam, the axial velocity and thus the displacement of the laser-spotted point could be determined fully time resolved.

6. Microphone measurements

For measuring the acoustic sound, a microphone (*Type BK 2669*) with spherical characteristics was placed 120 mm behind the outlet and 30 mm above the test channel. Furthermore, the microphone was inclined at about 45° to the axis of the test channel.

E. Analysis methods

1. Particle-image velocimetry

To increase the accuracy of the two-dimensional (2D) velocity field of one phase angle, the flow field was calculated by averaging 300 single vector plots obtained by measuring one phase angle over 600 times. Thereafter the single plots were divided manually into those on which the jet was attached to either the upper or the lower vocal fold. This procedure was processed for six time points within an oscillation cycle.

To investigate the positions of maximum deflection within the 2D flow field, the standard deviation of the flow velocity of one time point was calculated. In analogy with the turbulent kinetic energy[27] $k=(1/2)u_i u_i$, here the square of the standard deviation σ is regarded.

$$k = \sigma^2 = \frac{1}{2}\frac{\Sigma_{h=1}^{n}((u_1^h)^2 + (u_2^h)^2)}{n-1}. \tag{1}$$

$u_i^h = \bar{U}_i - U_i^h$ denotes the turbulent fluctuations of the velocity, $\bar{U}_i$ denotes the sample-averaged velocity, U_i^h denotes the instantaneous velocity, $i=1,2$ denotes the index of the spatial directions in the plane, and n denotes the number of measured values of the flow velocity at one position in the 2D plane for one phase angle. Equation (1) determines the squared deflection of the flow velocity including the turbulent fluctuations from cycle to cycle and the fluctuation of the jet position just downstream of the artificial glottis.

2. Synchroneous measurement techniques

For determining the coherence between the acoustic sound, the unsteady pressure, the flow velocity fluctuations, and the mechanical surface velocity of the synthetic vocal fold, all signals were measured synchronously. They were amplified and controlled by our own software program based on LABVIEW. The measurement time was 60 s using a sampling rate of 44 kHz. The calculations of the coherence between time-discrete signals were done in MATLAB. From these calculations, we obtained the coherence of the measured signals. The coherence reveals basic relations between the different signals and therefore deliver important informations about the physics of the fluid-structure-acoustic interactions occurring within human voice phonation.

TABLE II. Outline of the measurements at the two material configurations.

Synthetic vocal folds	
Homogeneous	Inhomogeneous/ multilayer
Visual. of the flow field Q=15–25 l/min E=6.5 kPa	Visual. of vocal fold oscillation Q=23.03 l/min E=6.5 kPa
Structure oscillation (LDV) Q=9.72–33.06 l/min E=6.5 kPa	Supraglottal velocity field (PIV) Q=29.07 l/min E=6.5 kPa
	Synchroneous measurements Unsteady pressure upstream and downstream of the glottis Unsteady flow velocity (CTA) Structure oscillations (LDV) Acoustic field (microphone) Q=23.03 l/min E=6.5 kPa

III. RESULTS

To give a short overview, Table II lists an outline of the processed measurements with the associated flow rate and Young's modulus depending on the material configuration of the synthetic vocal folds.

In a first step, a pair of homogeneous vocal folds with Young's modulus of E=6.5 kPa was used. It showed no recognizable oscillation and produced no tonal acoustic signal in the far field. The glottal width was estimated to be between 0 and 2 mm depending on the flow rate passing the vocal folds.

Due to these facts, the dynamic displacements were investigated applying LDV. The time signal of the velocity of the observed vocal fold was transformed by fast Fourier transform to obtain the spectral character of the structural velocity. Figure 4 shows the amplitude spectra of the velocity of the vocal fold for different flow rates. One can see that for each flow rate, the velocity shows a clear maximum. Furthermore, the frequency of the oscillation at the maximum amplitude decreases with increasing flow rate, whereas the value of the maximum amplitude simultaneously increases nonlinearly. This behavior could be interpreted as a resonant response of the vocal folds to the increasing flow rate. Even

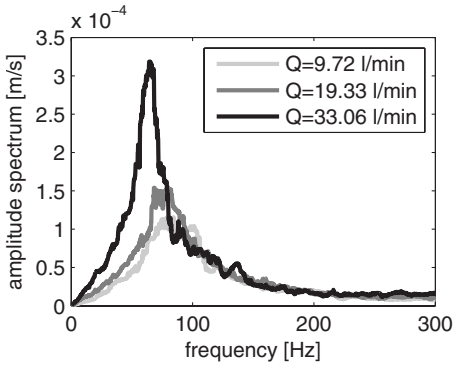


FIG. 4. Amplitude spectra of the homogeneous vocal fold velocity (E =6.5 kPa) for different volume flow rates obtained by LDV.

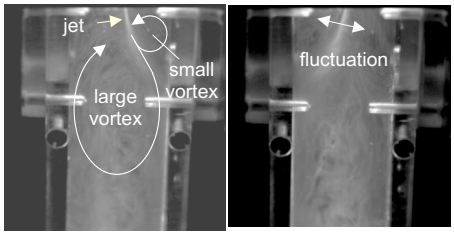


FIG. 5. (Color online) Flow field behind the homogeneous, synthetic vocal folds with the jet attached to either the right or the left vocal fold.

if the fluctuations of the structural velocity were in the range of 10^{-4} m/s for all flow rates, which leads to a displacement in the range between 10 nm and 10 μm. Nevertheless, the amplitude spectra show that the frequency of the vocal folds' motion was well located in the range of frequency occurring during human voice phonation.

To determine the quasisteady flow structures behind the homogeneous vocal folds, the flow field was visualized two dimensionally (Fig. 5). A jet arose from the narrowest gap of the artificial glottis, separated from one vocal fold, and attached to the other. The phenomenon is characterized by the Coanda effect and is well described in literature (see, e.g., Refs. 28, 9, and 10). Furthermore, the jet stochastically changed the side to which it was attached. Under these conditions, two recirculation areas occurred in the wake, which were not stationary but interacted with the surrounding flow in a chaotic manner and therefore produced turbulent structures. The major vortex expanded over the whole width of the channel whereas the smaller one is located between the jet and the downstream-facing surface of the attached vocal fold. The jet fluctuated in a small range although it attached to one vocal fold. A similar behavior was reported by Shinwari *et al.*,[6] who used rigid models of the vocal folds made of Plexiglas to obtain full optical access into the artificial glottis. They referred the flapping jet to small instabilities in the shear layers of the jet. However, with the present experimental setup, the flapping was mainly caused by the weak oscillations of the vocal folds themself.

The reason for the low oscillation amplitudes could be the lower pressure difference over the glottal duct because of the synthetic vocal tract. Other models as that used by Neubauer *et al.*[15] did not have a synthetic vocal tract and produced vocal fold vibrations with higher amplitudes. In those cases the pressure directly behind the glottal exit amounts enviromental pressure. The vocal tract extension causes a higher pressure than enviromental pressure at the glottal exit and therefore a lower pressure difference over the glottal constriction.

To increase the oscillation amplitude, a multilayer model of the vocal folds was used, in which a rigid mass body had been integrated in each tip of the synthetic vocal folds. This mass body increased the inertia of the tip region and roughly simulated the higher inertia of the mucous membrane (*squamous epithelium*) and the lamina propria due to its high portion of fluid. Thereby an oscillation of the vocal folds could be generated with a maximum glottis width of 2 mm. Additionally they produced a clear hum in the acoustic far field. The oscillation frequency was determined as 33 Hz using the

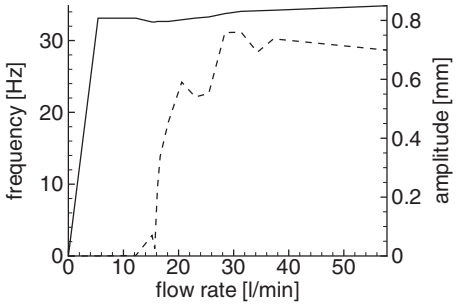


FIG. 6. Oscillation frequency (solid line) and amplitude (dashed line) of the inhomogeneous artificial vocal folds depending on the flow rate.

unsteady pressure signal upstream of the synthetic vocal folds, which is lower than the characteristic frequency in human voice production. It is assumed that the prestress in synthetic models was significantly lower than in human vocal folds. Nevertheless, the production of a clear tonal component in the acoustic signal suggests that the physical process of the sound production is similar to human sound production. Furthermore, the glottal form of the synthetic vocal folds showed a very good consistency with the real folds during the whole oscillation process.

The oscillation frequency remained approximately constant on increasing the flow rate, as Fig. 6 shows. The vibration started at a flow rate of about 5 l/min with a frequency of approximately 33 Hz, where it stayed constant up to 15 l/min. Thereafter, the frequency further increased to 35 Hz at a flow rate of about 56 l/min. In contrast, the amplitude of the oscillating synthetic vocal folds was lower than 0.1 mm for flow rates up to $Q=15$ l/min, increased to approximately 0.6 mm at $Q=20$ l/min, and to about 1 mm at $Q=30$ l/min before the amplitude slowly declined (Fig. 6). Note that the amplitude of the oscillation shown here is oriented in the axial flow direction of the test channel and displays the amplitude of a point in the middle of the downstream-facing surface of one vocal fold. The axial amplitude of a point near the artificial glottis was larger.

A time sequence of the oscillation process is displayed in Fig. 7. It shows the periodic opening and closing of the artificial glottis in selected pictures, recorded with the digital high-speed camera. As synchronized reference value the pressure fluctuation upstream of the glottis was used. The temporal pressure distribution was characterized by a nearly ideal sinusoidal behavior. The cycle started at the phase angle $\alpha=0$ with a fully closed glottis. A certain pressure force was needed to open the glottis. Its value depended on the stiffness of the vocal folds. Considering the different inertias of the fluid and the folds, the pressure still increased after opening until the flow was accelerated to a certain amount and thus the pressure was reduced. The higher inertia of the synthetic folds caused a further increase in the glottal width to its maximum of $w_{G,\max}=2$ mm at a phase angle of $\alpha=\pi$. At this point, the pressure and the shear forces of the fluid on the folds were in balance with the elastic forces of the material of synthetic vocal folds. The pressure further decreased because of the ongoing acceleration of the fluid within the glottis, which initiated the closing process.

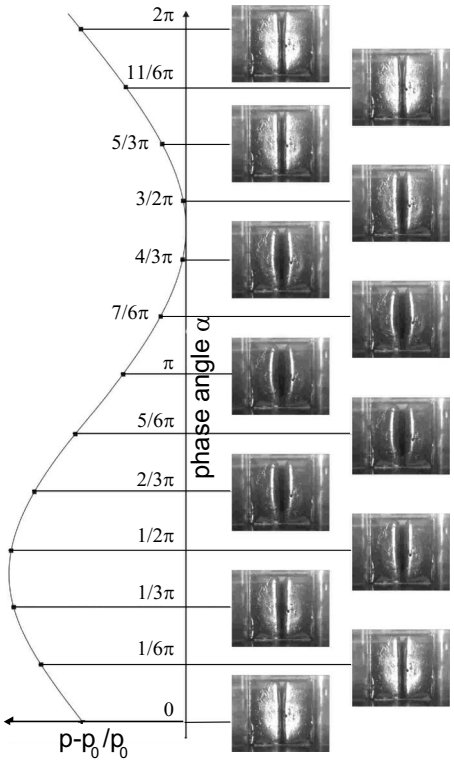


FIG. 7. Form of the glottis during a whole oscillation cycle depending on the pressure signal upstream of the glottis p and reference pressure p_0.

With continuing fluid acceleration, the pressure decreased until the pressure drop was too low and the glottal area too small, so the flow decelerated. After that, the pressure reached its minimum between $\alpha=\frac{4}{3}\pi$ and $\alpha=\frac{3}{2}\pi$. Due to the inertia of the synthetic vocal folds, the glottis completely closed until the pressure achieved the value to open it again and a new cycle started.

At this point, we emphasize that the whole process cannot be explained without the complex fluid-structure interaction. The phase shift between the glottis opening and the pressure distribution downstream of the glottis is the result of the energy flow between the fluid and structure domain. This was also observed by Thomson *et al.*[14] who calculated the energy transfer between the flow and the synthetic vocal folds within a 2D finite element simulation. It shows that reproducing the whole process including the acoustic production is necessary for artificial physical and also numerical models, which are able show the full fluid-structure-acoustic interactive behavior.

The investigation of the time-dependent flow structures during the phonation process was done by using PIV at a volume flow rate of $Q=23.03$ l/min (which represents a characteristic human volume flow rate during phonation). Figure 8 displays two single vector plots at the same phase angle for different oscillation cycles. In both cases the Coanda effect occurs. However, the jet changes the side stochastically to which it is attached. Therefore, the flow shows a bifurcation, which means that the flow can take on two stable states. Depending on small stochastic disturbances in the oncoming flow, the jet is irregularly attached to one of the vocal folds. To increase the accuracy an averaging of the

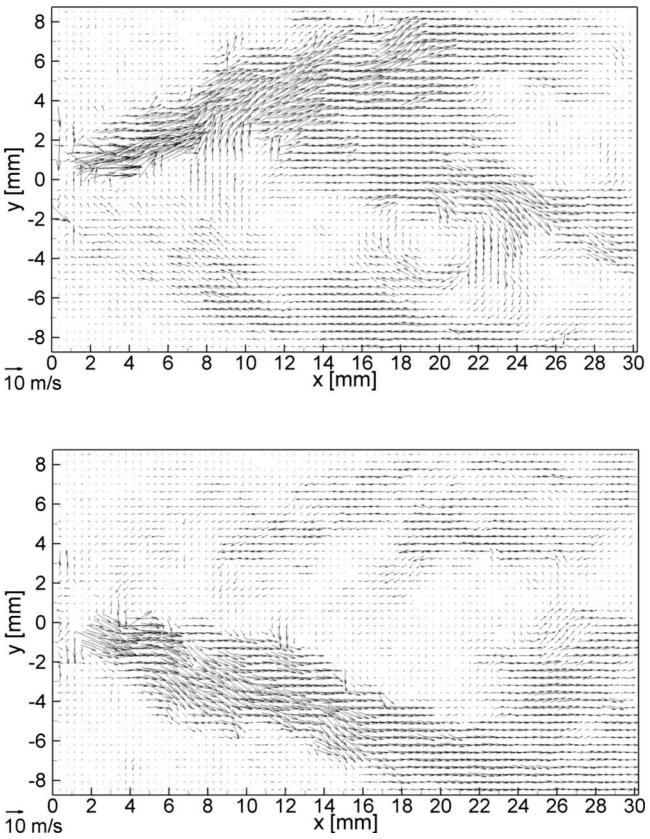


FIG. 8. Velocity vector plots of single PIV measurements with the jet attached either to the upper or lower synthetic vocal fold; the y ordinate displays the transversal and the x ordinate the longitudinal direction of the supraglottal region in human anatomy.

vector plots at a certain phase angle was performed. However, averaging all single vector plots at one phase angle would yield a flow field containing both stable states showing two jets arising from the glottis, which would be symmetrical to the center line. Thus, it was done by using 300 selected pictures on which the flow is attached to the upper vocal fold. For this task, we recorded over 600 single vector plots. Figure 9 shows the phase-averaged flow field for four phase angles. The contour shows the distribution of the turbulent kinetic energy k. The phase averaging was done to obtain a higher accuracy of the PIV data, in contrast to single vector plots (see Fig. 8).

Figure 9 shows the velocity distribution in the center plane orthogonal to the glottis at four different time steps during an oscillation cycle. Furthermore the deflected glottal duct and the direction of the jet are schematically sketched at the same phase angles in Fig. 10. Due to the complete closure of the glottis at $\alpha=0$, the flow through the glottis almost totally stopped. The maximum velocity in the measuring area at this time was about $u=6$ m/s. The flow field of Fig. 9 showed a large recirculation area expanded over the whole height of the channel. In the following phase angle $\alpha=\frac{1}{3}\pi$, the glottis started to open and a jet was formed through the glottis, which was attached to the upper vocal fold. The separation point of the jet was located at the downstream-facing edge of the vocal fold. This resulted in a large angle between the jet axis and the center line of the channel. Due to the movement of the separation point, the angle slightly increased at $\alpha=\frac{1}{2}\pi$. Furthermore the jet velocity increased to its maximum value of $u=22.5$ m/s at $\alpha=\pi$, at which the jet was fully accelerated. The changed geometry of the glottal duct and the higher velocity resulted in a movement of the separation point in the upstream direction (see $\alpha=\pi/2$, $\alpha=\pi$, and $\alpha=\frac{5}{3}\pi$ in Figs. 9 and 10). The jet axis rotated toward the center line because the separation point from the upper vocal fold went further upstream. That is, the angle between the jet axis and the center line became sharper with time and was nearly zero shortly before the full closure of the glottis occurred.

Summarizing, we can state that the separation point of the attached jet moved during the phonation process. Hence, the pulsating jet interacted with the back side of the vocal folds and generated an unsteady oscillating pressure distribution on it. Similar sound generation processes can be observed in trailing-edge noise[29,30] or the aeolian tone production around cylinder-shaped geometries (see Ref. 31). As a result of the Coanda effect, the pressure distributions on the downstream-facing surfaces of the vocal folds were not symmetrical. Furthermore, the pressure difference over the channel height at the orthogonal faces of the vocal folds was much larger during the cycle in comparison with the case of a symmetric jet through the glottis. This could be due to the Coanda effect interacting with the vocal folds as the sound generation mechanism at higher frequencies similar to trailing-edge noise. However, these observations require further detailed investigations.

As already mentioned above, Fig. 9 shows the distribution of the square of the standard deviation of the velocity described as the turbulent kinetic energy k [Eq. (1)] at different time steps. The value k contains several effects. In addition to the uncertainties, there were differences in the flow distribution for the same phase angle in different oscillation cycles. Only selected PIV data, for which the jet attached to the same vocal fold, were used to calculate k. The reasons for the fluctuations of the flow field can be explained on the one hand by the stochastical fluctuations of the flow velocity produced in the shear layers of the jet and on the other hand by the different jet positions from one cycle to another. The highest amount of turbulent kinetic energy developed in the core region of the jet just after entering the supraglottal area (see, e.g., $\alpha=\pi/2$ in Fig. 9). This means that the largest contribution to the energy of the fluctuations results from the differences in the jet angle. The reasons for this behavior are small instabilities of the exact position of the separation point at each phase angle from cycle to cycle.

In the present study, we could not observe a change in the jet from one vocal fold to the other within the same cycle as was found in the work of Neubauer *et al.*[15] The jet is attached to the same vocal fold during the whole cycle. However, the jet changed its position stochastically from one cycle to the following. The oscillations of the vocal folds are nearly symmetrical. This is in agreement with measurements on patients with healthy voice phonation (see Ref. 32).

During the investigations of the oscillation of the synthetic vocal folds and the flow field downstream of the glottis, a clear tonal sound could be detected in the test rig. Superimposed on this aeolian sound, a broadband sound oc-

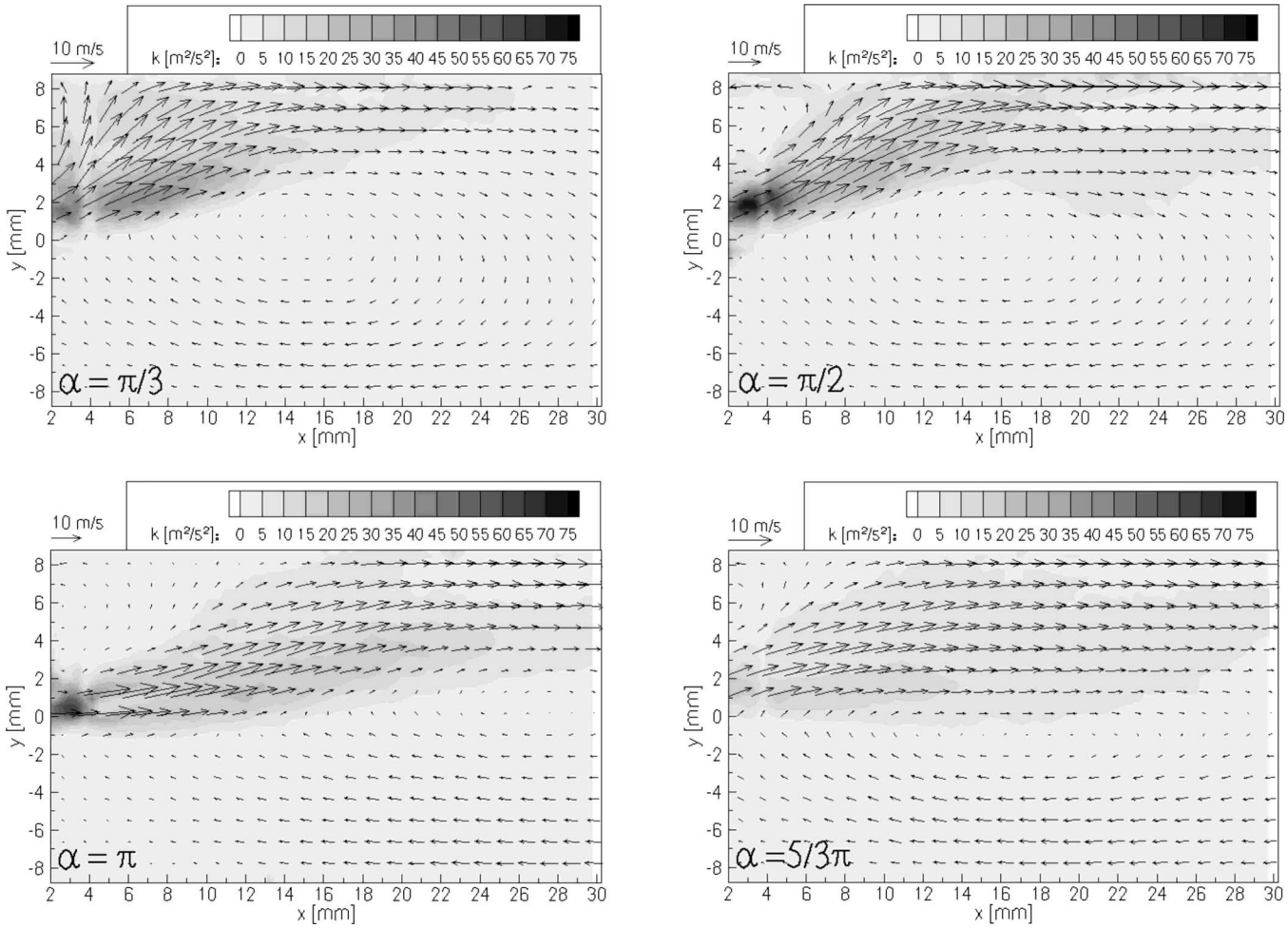


FIG. 9. Vector plot of the phase-averaged flow velocity and contour plot of the turbulent kinetic energy k behind the oscillating vocal folds' models at four instants of an oscillation cycle.

curred. To identify their acoustic source processes, simultaneous measurements for determining were performed. Therefore, the coherence of the UPU and UPD of the synthetic glottis, and the unsteady flow velocity (CTA) at the center point of the channel in the downstream region of the vocal folds, with the AP in the far field was calculated.

Figure 11 displays the amplitude spectra of all measurements. The acoustic signal contained oscillations with high intensities at frequencies lower than 10 Hz. The reason for these high intensity oscillations can be explained by the environmental noise. The major peak in all signals occurred at 33 Hz, which is directly related to the oscillation of the vocal folds, which is confirmed by the amplitude spectra of the vocal folds' velocity (black dashed line in Fig. 11). This signal was measured by LDV. Furthermore, the higher harmonics of the fundamental frequency are present and clearly visible in all spectra. Except for the spectrum of the unsteady flow velocity, the major peaks are equal to the oscillation of highest intensity at the fundamental frequency of $f=33$ Hz. This discrepancy resulted from the location of the hot-wire sensor within the flow field. The probe was located outside of the large recirculation area, which was formed during the opening phase of the cycle just downstream of the glottis. The wire was placed perpendicular to the synthetic glottis. Because of the increase in the volume flow rate after glottis opening, the shedding of the recirculation area, and the limitation of the hot-wire sensor due to the oncoming flow angle,[25] the major peak in the spectrum of the flow velocity occurs at the third harmonic frequency of 99 Hz. Additionally to the harmonic results, some sharp peaks at frequencies between $f=50$ and 700 Hz occurred with a frequency step of 50 Hz. They are caused by the voltage supply. Moreover, a broadband sound contribution at higher frequencies ($f>1000$ Hz) could be observed.

Figure 12 shows the coherence of the three signals measured in the flow region (UPU, UPD, and CTA) with the AP in the far field. Here, a coherence of 0 means no correlation and a coherence of 1 full correlation of the regarded signal with the AP at a certain frequency. The pressure upstream (solid black line) and downstream (solid gray line) of the glottis conform very well with the acoustic sound radiation at the fundamental frequency and their higher harmonics with a coherence of over 0.9. Also, at higher frequencies, there is a high coherence of partially over 0.9 due to the broadband noise for frequencies larger than 1000 Hz. In contrast, the unsteady velocity (dashed black line) only correlates well with the acoustic signal at the oscillation frequency and its higher harmonics. For $f>1000$ Hz there is no coherence. Considering the acoustic analogy of Lighthill[33,34] the turbu-

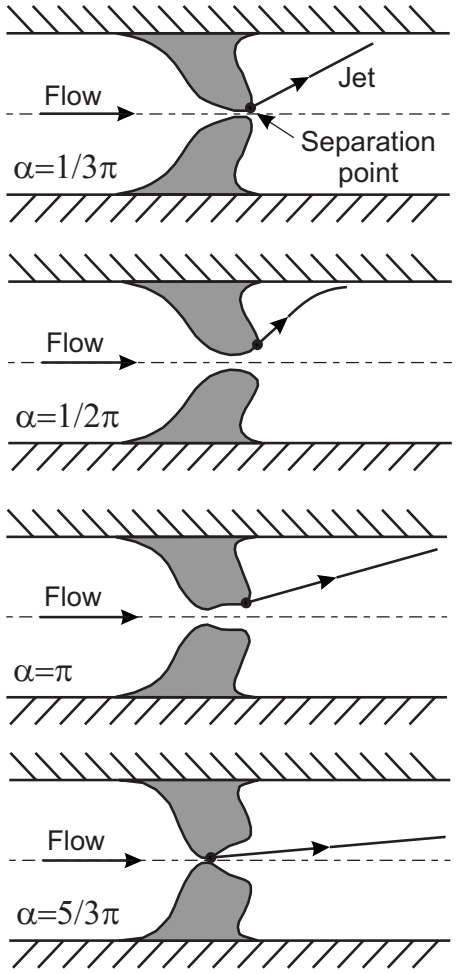


FIG. 10. Schematics of the glottal duct at different phase angles in the oscillation cycle of the vocal folds with the location of the separation point of the jet.

lent sound generation is related to turbulent fluctuations of the flow velocity. The lack of coherence between the flow velocity and the AP indicates that the turbulent structures produced in the shear layer of the jet did not make a dominant contribution to the noise generation in the phonation process. Furthermore the maximum velocity of 22.5 m/s (Ma$<$0.1 within the jet) was to low that it created a determinative acoustic source term in the shear layer.

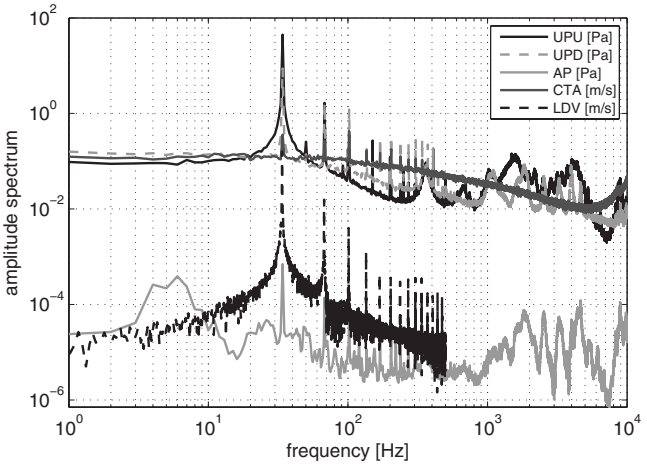


FIG. 11. Amplitude spectra of the UPU and UPD of the glottis, of the unsteady flow velocity downstream of the glottis (CTA), the velocity of the vocal folds' models (LDV), and of the AP in the far field (AP) at a flow rate of Q=23.03 l/min.

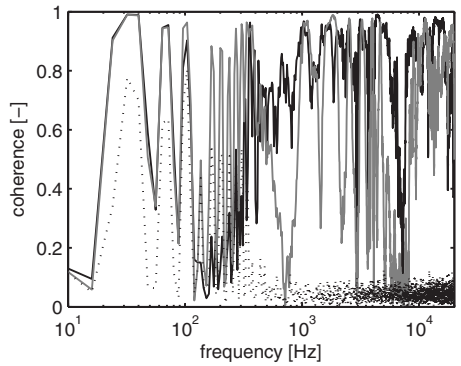


FIG. 12. Coherence between the AP and the UPU (solid black line), the AP and the UPD of the glottis (solid gray line), and between the AP and the unsteady velocity (dashed black line).

IV. CONCLUSION

The paper reports an investigations of flow-induced vibrations of artificial vocal folds. They are integrated in a channel, which partly simulated the vocal tract in the supraglottal region. This extension of the channel results in a lower pressure difference across the vocal folds, which is suggested to be the cause for the small vibration amplitudes. To increase these amplitudes, we included mass body made of lead in our synthetic vocal folds a spherical, which resulted in a lower frequency of the oscillations compared with the real human phonation process. Besides this discrepancy, the present test rig could reproduce the whole flow-induced phonation process in a channel. The complex interaction between the fluid and the structure of the artificial vocal folds turned out to be important. It shows a phase shift between the unsteady pressure distribution upstream of the glottis and the structural oscillations of the vocal folds, which depended on the elasticity forces within the material and the inertia of the fluid and the vocal folds. The jet through the glottis was attached to one vocal fold and therefore separated from the other. Hence, it is bent to one side at the exit of the glottal duct and formed an asymmetric flow distribution in the supraglottal region. The separation point moved during one cycle along the surface of the vocal folds, which resulted in a flapping jet angle between the axis of the jet and the center line of the channel. Moreover, the jet stayed attached to one vocal fold during one cycle but changed the sides in different cycles in a stochastical manner. The sound production is assumed to be caused by the interaction of the separated unsteady jet flow with the downstream-facing surfaces ot the vocal folds. The authors suggested that the oscillating pressure field at the surfaces generate a broadband noise analogically to trailing-edge noise concerning aircraft. It produces an oscillating pressure field. The high-frequency vortex structures in the shear layers of the jet played a minor role in the sound-generating process. The tonal sound was directly related to oscillation frequency of the synthetic vocal folds. It was produced by the oscillating flow rate due to the peri-

odic opening and closing of the glottis. This tonal sound is the major sound-generating process in the test rig.

Future investigations will include detailed studies with different material parameters of multilayer vocal folds and the inclusion of the longitudinal tension within the vocal folds with the aim of imitating the human phonation process in a more realistic way.

[1]S. McCoy, *Your Voice: An Inside View* (View, Princeton, NJ, 2004).
[2]F. Alipour, R. Scherer, and J. Knowles, "Velocity distributions in glottal models," J. Voice **10**, 50–58 (1996).
[3]S. L. Thomson, L. Mongeau, and S. H. Frankel, "Physical and numerical flow-excited vocal fold models," Third International Workshop MAVEBA, 2003.
[4]M. Triep, C. Brücker, and W. Schröder, "High-speed flow measurements of the flow down-stream of a dynamic mechanical model of the human vocal folds," Exp. Fluids **39**, 232–245 (2005).
[5]R. C. Scherer, D. Shinwari, K. J. DeWitt, C. Zhang, B. R. Kucinschi, and A. A. Afjeh, "Intraglottal pressure profile for a symmetric and oplique glottis with a divergence angle of 10 degrees," J. Acoust. Soc. Am. **109**, 1616–1630 (2001).
[6]D. Shinwari, C. Scherer, K. J. DeWitt, and A. A. Afjeh, "Flow visualization and pressure distribution in a model of the glottis with a symmetric and oblique divergent angle of 10 degrees," J. Acoust. Soc. Am. **113**, 487–497 (2003).
[7]H. Coanda, "Device for deflecting a stream of elastic fluid projected into an elastic elastic fluid," U.S. Patent No. 2,052,869 (April 19, 1935).
[8]G. C. J. Hofmans, G. Groot, M. Ranucci, G. Graziani, and A. Hirschberg, "Unsteady flow through in-vitro models of the glottis," J. Acoust. Soc. Am. **113**, 1658–1675 (2003).
[9]B. D. Erath and M. W. Plesniak, "An investigation of bimodal jet trajectory on flow through scaled models of the human vocal tract," Exp. Fluids **40**, 683–696 (2006).
[10]B. D. Erath and M. W. Plesniak, "The occurrence of the Coanda effect in pulsatile flow through static models of the human folds," J. Acoust. Soc. Am. **120**, 1000–1011 (2006).
[11]M. Krane, M. Barry, and T. Wei, "Unsteady behavior of flow in a scaled-up vocal folds model," J. Acoust. Soc. Am. **122**, 3659–3670 (2007).
[12]J. B. Park and L. Mongeau, "Instantaneous orifice discharge coefficient of a physical, driven model of the human larynx," J. Acoust. Soc. Am. **121**, 442–455 (2007).
[13]Z. Zhang, L. Mongeau, and S. H. Frankel, "Experimental verification of the quasisteady approximation for aerodynamics sound generation by pulsating jets in tube," J. Acoust. Soc. Am. **112**, 1652–1663 (2002).
[14]S. L. Thomson, L. Mongeau, and S. H. Frankel, "Aerodynamic transfer of energy to the vocal folds," J. Acoust. Soc. Am. **118**, 1689–1700 (2005).
[15]J. Neubauer, Z. Zhang, R. Miraghaie, and D. A. Berry, "Coherent structures of the near field flow in a self-oscillating physical model of the vocal folds," J. Acoust. Soc. Am. **121**, 1102–1118 (2007).
[16]M. Döllinger, N. Tayama, and D. A. Berry, "Empirical eigenfunctions and medial surface dynamics of a human vocal fold," Methods Inf. Med. **44**, 384–391 (2005).
[17]J. Neubauer, P. Mergell, U. Eysholdt, and H. Herzel, "Spatio-temporal analysis of irregular vocal fold oscillations: Biphonation due to desynchronization of spatial modes," J. Acoust. Soc. Am. **110**, 3179–3192 (2001).
[18]I. R. Titze, *Principles of Voice Production* (Prentice-Hall, Englewood Cliffs, NJ, 2000).
[19]F. Durst, U. Heim, B. Ünsal, and G. Kullik, "Mass flow rate control system for time-dependent laminar and turbulent flow investigations," Meas. Sci. Technol. **14**, 893–902 (2003).
[20]T. D. Rossing, *Handbook of Acoustics* (Springer, New York, 2007).
[21]Kulite Inc. Germany, web: www.kulite.de (Last viewed May, 2008).
[22]Weinberger Visio Inc. Germany, web: www.weinbergervision.com (Last viewed May, 2008).
[23]M. Raffel, C. E. Willert, S. T. Wereley, and J. Kompenhans, *Particle Image Velocimetry: A Practical Guide (Experimental Fluid Mechanics)*, 2nd ed. (Springer, Berlin, 2007).
[24]Dantec Dynamics A/S Denmark, web: www.dantecdynamics.com (Last viewed May, 2008).
[25]H. H. Bruun, *Hot-Wire Anemometry: Principles and Signal Analysis* (Oxford University Press, Oxford, 1995).
[26]Polytec GmbH, web: www.polytec.com (Last viewed May, 2008).
[27]F. Durst, *Fluid Mechanics. An Introduction to the Theory of Fluid Flows* (Springer, Berlin, 2007).
[28]R. C. Scherer, F. Alipour, and J. Knowles, "Velocity distribution in glottal models," J. Voice **10**, 55–58 (1996).
[29]W. K. Blake, *Mechanics of Flow-Induced Sound and Vibration* (Academic, New York, 1986), Vols. 1 and 2.
[30]T. F. Brooks and T. H. Hodgson, "Trailing edge noise prediction from measured surface pressures," J. Sound Vib. **78**, 69–117 (1981).
[31]S. Becker, M. Kaltenbacher, I. Ali, C. Hahn, and M. Escobar, "Aeroacoustic investigation of the flow around cylinder geometries—A benchmark test case," 13th AIAA/CEAS Aeroacoustics Conference, Romam, 2007, AIAA Paper No: 2007-3511.
[32]U. Eysholdt, M. Tigges, T. Wittenberg, and U. Prschel, "Direct evaluation of high-speed recordings of vocal fold vibrations," Folia Phoniatr Logop **48**, 163–70 (1996).
[33]M. J. Lighthill, "On sound generated aerodynamically—I. General theory," Proc. R. Soc. London, Ser. A **211**, 564–587 (1952).
[34]M. J. Lighthill, "On sound generated aerodynamically—II. Turbulence as a source of sound," Proc. R. Soc. London, Ser. A **222**, 1–32 (1954).

Ray-based description of normal modes in a deep ocean acoustic waveguide

A. L. Virovlyansky, A. Yu. Kazarova, and L. Ya. Lyubavin
Institute of Applied Physics, Russian Academy of Science, 46 Ul'yanov Street, 603950 Nizhny Novgorod, Russia

(Received 29 May 2008; revised 31 December 2008; accepted 5 January 2009)

Modal structure of the wave field in a deep ocean environment with sound speed fluctuations induced by random internal waves is considered. An approximate analytical description of the modal structure at megameter ranges is derived by combining two known results: (i) relations expressing mode amplitudes through parameters of ray paths and (ii) stochastic ray theory. For a monochromatic wave field, a simple analytical estimate has been obtained for a coarse-grained distribution of acoustic energy between normal modes. Significant attention has been paid to the investigation of the mode pulses, that is, sound pulses carried by individual modes. Analytical estimates for the spread of mode pulse and bias of its mean travel time in the presence of internal waves are derived. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3075765]

PACS number(s): 43.30.Bp [JAC] Pages: 1362–1373

I. INTRODUCTION

The objective of the present study is the derivation of an approximate analytical description of the modal structure in a deep ocean acoustic waveguide. Traditionally this issue is addressed proceeding from the mode coupling equations.[1–5] These equations are complicated and therefore usually they should be treated numerically. In this paper, an alternative approach is discussed. We argue that a simple (albeit rough) analytical description of the modal structure in a fluctuating ocean can be obtained by combining (i) ray-mode relations expressing the mode amplitude through solutions of ray equations[6–9] and (ii) stochastic ray theory.[10–12]

Neglecting the horizontal refraction, we consider a two-dimensional environmental model with the sound speed field $c(r,z)$ (r is range, and z is depth) presented in the form $c(r,z)=c_0(z)+\delta c(r,z)$, where $c_0(z)$ is a smooth (background) sound speed profile, and $\delta c(r,z)$ is a weak perturbation induced by random internal waves with statistics determined by the empirical Garrett–Munk spectrum.[13] In the presence of perturbation, the ray paths exhibit chaotic behavior.[14–16] They become very sensitive to small variations in starting parameters: the difference between vertical coordinates of two initially close paths grows with range (on the average) as $\exp(\lambda r)$, where $\lambda>0$ is the Lyapunov exponent. In the deep sea $\lambda=O(0.01\ \mathrm{km}^{-1})$ and at megameter ranges, the ray chaos is well developed.[14,16] Manifestation of ray chaos at a finite wavelength is called the wave chaos. The complicated range dependence of the modal structure is an important aspect of this phenomenon.

In recent years it has been demonstrated that, even under condition of ray chaos, the ray-based description of the sound field in a deep ocean can properly predict many important features of the arrival pattern at megameter ranges.[16,17] What is especially important in the context of the present study, the statistics of chaotic rays can be approximately described analytically. In Refs. 10–12 and 18, it is shown that such a description can be obtained by replacing the Hamilton (ray) equations expressed in terms of the action-angle canonical variables with stochastic Langevin equations. Earlier a similar approach was derived in Ref. 19 in terms of the momentum-position variables.

Our idea consists in applying the stochastic ray theory for investigating the modal structure. This is done using the known analytical relations that allow one to express the mode amplitude through parameters of the ray paths. The ray-mode relations are based on the fact that the normal mode is formed by contributions from rays whose action variables at the range of observation are determined by the quantization rule.[6–9] These rays we call the *mode rays*. They play the same role in evaluating the mode amplitude as the eigenrays play in evaluating the wave field at the given observation point.

Under conditions of ray chaos the number of mode rays contributing to the given mode grows with range exponentially and rapidly becomes very large. Assuming that their phases are independent we sum up these rays incoherently. The sum has been estimated using the stochastic ray theory from Refs. 10–12 and 18. As a result an approximate analytical expression for a coarse-grained distribution of acoustic energy between normal modes has been obtained.

When considering a transient wave field, our attention is focused on the *mode pulses*, that is, sound pulses carried by individual modes. It is shown that the temporal structure of the mode pulse to a significant extent is determined by the travel times of mode rays. The use of stochastic ray theory has made it possible to estimate the spread and bias of the mode pulse caused by internal waves. It turns out that for high modes the spread and bias grow with range like $r^{3/2}$ and r^2, respectively. First these range dependencies were found in Ref. 20 on the basis of numerical simulations. Later an analytical estimate for the spread was derived in Refs. 11, 21, and 22.

II. RAY STRUCTURE OF THE WAVE FIELD

A. Parabolic equation

Consider a monochromatic sound field at an angular carrier frequency Ω. In the parabolic equation approximation, its complex amplitude is presented in the form

$$v(r,z,\Omega) = r^{-1/2}u(r,z,\Omega)e^{ikr}, \tag{1}$$

where $k=\Omega/c_r$ is a wave number in a reference medium with the sound speed c_r, and $u(r,z,\Omega)$ is an envelope function. It is assumed that $u(r,z,\Omega)$ obeys the standard parabolic equation[23,24]

$$2ik\frac{\partial u}{\partial r} + \frac{\partial^2 u}{\partial z^2} - 2k^2[U(z)+V(r,z)]u = 0, \tag{2}$$

where

$$U(z) = \frac{1}{2}\left(1 - \frac{c_r^2}{c_0^2(z)}\right), \quad V(r,z) = \frac{\delta c(r,z)}{c_r}. \tag{3}$$

The z-axis is directed upward and the sea surface is located at $z=0$.

A transient wave field excited by a pulse source can be synthesized out of solutions of Eq. (2) at different carrier frequencies as

$$\tilde{v}(r,z,t) = r^{-1/2}\int d\Omega' u(r,z,\Omega')s(\Omega')e^{i\Omega'(r/c_r-t)}, \tag{4}$$

where $s(\Omega)$ is the spectrum of an initially radiated pulse.

B. Geometrical optics

In the geometrical optics approximation, the solution to Eq. (2) is expressed through parameters of ray trajectories.[14,16,19,25] In the unperturbed waveguide ($V=0$), the ray path obeys the Hamilton equations $dz/dr=\partial H/\partial p$ and $dp/dr=-\partial H/\partial z$, where $H=p^2/2+U(z)$ is an unperturbed Hamiltonian and p is the momentum. The latter is related to the ray grazing angle χ through $p=\tan\chi$. In the presence of perturbation, H should be replaced by $H+V$.

The contribution from a single ray to the total wave field is

$$u(r,z,\Omega) = A(r,z)\exp[ikS(r,z)], \tag{5}$$

where A and S are the ray amplitude and eikonal, respectively. The eikonal S is an analog to Hamilton's principal function in classical mechanics and it is given by an integral

$$S = \int [pdz - (H+V)dr] \tag{6}$$

running over the ray path.[14,26]

1. Action-angle variables

Consider an unperturbed waveguide and for simplicity assume that the function $U(z)$ has a single minimum. In this (range-independent) waveguide, each ray is periodic and the Hamiltonian H remains constant along the trajectory. The action variable I is defined as an integral over a cycle of the ray path[25,26]

$$I = \frac{1}{2\pi}\oint pdz = \frac{1}{\pi}\int_{z_{\min}}^{z_{\max}} dz\sqrt{2[H-U(z)]}, \tag{7}$$

where $z_{\min}$ and $z_{\max}$ are the lower and upper turning depths, respectively, satisfying the condition $U(z)=H$. Equation (7) defines function $H(I)$ expressing the Hamiltonian as a function of action. The canonical transformation

$$p = p(I,\theta), \quad z = z(I,\theta) \tag{8}$$

connecting the position-momentum and action-angle variables is defined by equations $p=\partial G/\partial z$ and $\theta=\partial G/\partial I$, where $G(I,z)$ is a generating function. For $p>0$,

$$G(I,z) = \int_{z_{\min}}^{z} dzP(I,z), \quad P(I,z) = \sqrt{2[H(I)-U(z)]}. \tag{9}$$

If $p<0$, $G(I,z)$ should be replaced by $2\pi I - G(I,z)$. The angle variable θ can be interpreted as a phase of the ray path. According to the above definition, it belongs to the interval from 0 to 2π. To make θ continuous, its value should be increased by 2π at the beginning of each new cycle. This is a standard procedure.[26]

The canonical transformation determined for the unperturbed waveguide, formally, can be used in the perturbed waveguide (with $V\neq 0$) as well. In the presence of perturbation, the ray equations take the form

$$\frac{dI}{dr} = -\frac{\partial V}{\partial\theta}, \quad \frac{d\theta}{dr} = \omega + \frac{\partial V}{\partial I}, \tag{10}$$

where $\omega(I)=dH(I)/dI$ is the spatial angular frequency of the ray path oscillation in the unperturbed waveguide. The cycle length of the unperturbed path is $D(I)=2\pi/\omega(I)$.

In what follows, we will use functions $I(r,I_0,\theta_0)$ and $\theta(r,I_0,\theta_0)$ to denote the action and angle of the ray path at range r. Arguments I_0 and θ_0 are initial values of these variables at $r=0$. Sometimes it will be more convenient to use similar functions $I(r,p_0,z_0)$ and $\theta(r,p_0,z_0)$, where p_0 and z_0 are initial momentum and coordinate of the ray path, respectively.

2. Ray travel times

A signal arriving at the observation point through a particular ray path—we call it the *ray pulse*—can be evaluated by substituting Eq. (5) into Eq. (4). Since both A and S do not depend on Ω, it is easy to see that the travel time of a ray pulse is

$$t = (r+S)/c_r. \tag{11}$$

Properties of travel times for both regular and chaotic rays are described in detail in many publications (see, e.g., Refs. 11, 14, 16, 23, and 27). In this section, we present only two formulas that will be used below.

The first of them is a simple estimate of the ray travel time in a range-independent waveguide ($V=0$). It follows from Eqs. (6)–(9) that at 1 cycle of oscillations the eikonal of a ray with action I increases by $2\pi[I-H(I)/\omega(I)]$. Therefore at range r the travel time is approximately equal to

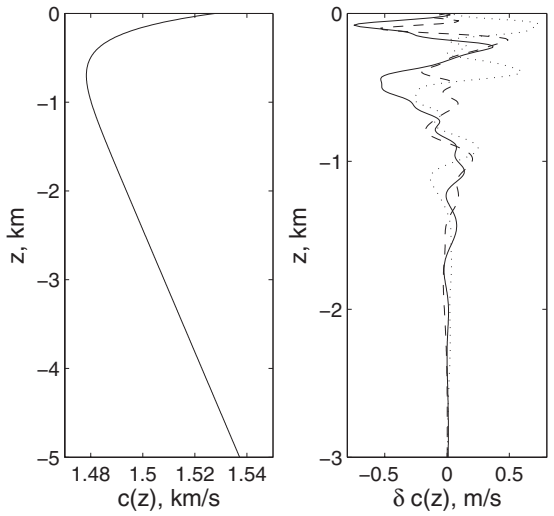


FIG. 1. Unperturbed sound speed profile $c_0(z)$ (left panel) and perturbation δc in vertical sections of the waveguide at three different ranges (right panel).

$$\bar{t}(I,r) = [1 + I\omega(I) - H(I)]r/c_r. \tag{12}$$

In what follows, the argument r of function $\bar{t}(I,r)$ will be omitted.

The second formula gives an estimate for the difference between travel times of two rays escaping a point source with close launch angles. Then the starting actions of both rays are also close and we denote their mean value by I_0. Let N_1 and N_2 be the numbers of cycles of these two rays, and t_1 and t_2 be their travel times at the range of observation. At a long range where $N_{1,2} \gg 1$ and $\Delta N = N_2 - N_1 = O(1)$, the difference $\Delta t = t_2 - t_1$ can be roughly estimated as

$$\Delta t = 2\pi I_0 \Delta N/c_r. \tag{13}$$

Initially, this result was derived for the range-independent waveguide.[27–29] But recently it has been found that Eq. (13) remains valid (as a rough estimate) even under conditions of ray chaos.[10–12,30] This fact is important for our subsequent analysis.

C. Environmental model

In numerical simulations presented below, we use an environmental model with an unperturbed profile $c_0(z)$ (borrowed from Ref. 20) shown in the left panel of Fig. 1. It represents the Munk profile:[13,23]

$$c_0(z) = c_r[1 + \varepsilon(e^{\eta} - \eta - 1)], \quad \eta = 2(z - z_a)/Q, \tag{14}$$

with parameters $c_r = 1.5$ km/s, $\varepsilon = 0.00238$, $Q = 0.485$ km, and $z_a = -0.7$ km.

It is assumed that the weak perturbation $\delta c(r,z)$ is caused by random internal waves with statistics determined by the empirical Garrett–Munk spectrum.[13] To generate realizations of a random field $\delta c(r,z)$, we apply a numerical technique developed by Colosi and Brown.[31] Realizations of $\delta c(r,z)$ have been computed using Eq. (19) from Ref. 31. Components of wave number vectors in the horizontal plane belong to the interval from $2\pi/100$ km^{-1} to $2\pi/4$ km^{-1}. A rms amplitude of the perturbation scales in depth like $\exp(3z/2L)$, where $L = 1$ km, and its surface-extrapolated value is about 0.5 m/s. Depth dependencies of δc at three different ranges are shown in the right panel of Fig. 1.

D. Statistical description of chaotic rays

Consider functions $I(r, I_0, \theta_0)$ and $\theta(r, I_0, \theta_0)$ describing a ray path in the action-angle phase plane. In the presence of perturbation $\delta c(r,z)$, the path exhibits stochastic behavior and both functions become extremely complicated. Therefore, they should be described by means of statistical methods. A traditional approach is associated with ideas of theory of wave propagation in random media. It is based on the notion of statistical ensemble consisting of infinitely many realizations of the waveguide specified by different $\delta c(r,z)$. Statistical characteristics of a ray with given starting parameters are determined by averaging over rays with the same initial parameters in all realizations.

In the present paper, we apply an alternative approach derived in Refs. 11, 12, 18, and 32. It deals with a *single* realization of random perturbation and is based on the theory of ray chaos. The idea of this approach is as follows. At ranges $r \gg \lambda^{-1}$, initially, close ray paths become practically independent and the averaging over their starting parameters can be considered as the statistical averaging. Statistical characteristics of a path escaping point (I_0, θ_0) are determined by considering a bundle of rays starting from a small area of the phase plane $\mathcal{R}$ (it may be one-dimensional) centered at (I_0, θ_0). At a long enough range where the chaos is well developed, the bundle plays the role of a statistical ensemble. The function

$$P_{I\theta}(I, \theta, r | I_0, \theta_0) = \frac{1}{S_{\mathcal{R}}} \int_{\mathcal{R}} dI_0 d\theta_0 \delta(I - I(r, I_0, \theta_0)) \times \delta(\theta - \theta(r, I_0, \theta_0)), \tag{15}$$

where $S_{\mathcal{R}}$ is the square of $\mathcal{R}$, may be interpreted as a probability density function (PDF) of I and θ at range r. This definition makes sense only at long enough ranges where the result of averaging weakly depends on the size and shape of $\mathcal{R}$. Moreover, in Refs. 11, 12, 18, and 32, we argue that the PDF defined by Eq. (15) may weakly depend on a particular realization of the waveguide. At long ranges, initially, close ray paths diverge so significantly that they are spaced apart by intervals exceeding correlation scales of the medium. Then the rays travel through practically independent inhomogeneities and behave as if they propagate in different realizations of the medium. Therefore, it is natural to expect that the averaging over initial conditions may give results that (i) weakly depend on a particular realization and (ii) are close to those obtained by ensemble averaging.

In Refs. 11, 12, 18, and 32, it is shown that an analytical estimate of the above PDF can be obtained by replacing the Hamilton (ray) equations (10) with simple stochastic Langevin equations. In this approach, the action and angle of a ray path are presented in the form

$$I(r,I_0,\theta_0)=I_0+x(r),\quad \theta(r,I_0,\theta_0)=\theta_0+\omega(I_0)r+y(r), \tag{16}$$

where $x(r)$ and $y(r)$ are random functions describing the deviation of the ray path from its unperturbed position. The deviation of action, $x(r)$, can be approximated by a random Wiener process.[33] It is a zero mean Gaussian process whose statistical characteristics are defined by the correlation function $\langle x(r)x(r')\rangle = B\min(r,r')$, where B is a coefficient of diffusion. The value of B was estimated numerically for a few typical deep water waveguides.[11,12,18,22,32] It turned out to be of order 10^{-7} km. For the environmental model described in Sec. II C, $B=1.4\times10^{-7}$ km.

The deviation of the angle variable is given by relations

$$y(r)=\omega'(I_0)\eta(r),\quad \eta(r)=\int_0^r dr'x(r'). \tag{17}$$

Using the standard relations of the probability theory, it is easy to show that the joint PDF of x and η is

$$P_{x\eta}(x,\eta,r)=\frac{\sqrt{3}}{\pi Br^2}\exp\left[-\frac{x^2}{2Br}-\frac{6}{Br^3}\left(\eta-\frac{xr}{2}\right)^2\right]. \tag{18}$$

Substituting $I-I_0$ and $\theta-\theta_0-\omega(I_0)r$ for x and y, respectively, yields an explicit expression for the PDF $P_{I\theta|I_0\theta_0}(I,\theta,r|I_0,\theta_0)$ defined by Eq. (15). As it follows from Eq. (18), the conditional PDF of action I at range r, i.e., the PDF of I given that at $r=0$ this variable equals I_0, is

$$P_I(I,r|I_0)=\frac{1}{\sqrt{2\pi Br}}\exp\left[-\frac{(I-I_0)^2}{2Br}\right]. \tag{19}$$

In the scope of our approach, statistical characteristics of the action variable do not depend on the starting angle θ_0.

Note that Eqs. (18) and (19) are valid only for steep rays, i.e., for rays with large enough I_0. The point is that the action I is non-negative by definition. According to Eq. (19), this condition is met for most rays if I_0 exceeds $\sqrt{Br}$. At megameter ranges, this occurs for rays with grazing angles at the sound channel axis

$$|\chi_a|>5^\circ. \tag{20}$$

For treating flat rays, our approach should be modified as it is discussed in Refs. 11, 12, 18, and 32.

The above formulas can be used for averaging over all rays escaping a source exciting the wave field. To explain this statement, consider a point source set at a depth z_s. Initial momenta, p_0, of rays leaving the source lie within interval $(-p_{\max},p_{\max})$. Consider the quantity

$$\overline{F(I,\theta)}\equiv\int_{-p_{\max}}^{p_{\max}}dp_0F(I(r,p_0,z_s),\theta(r,p_0,z_s)), \tag{21}$$

where $F(I,\theta)$ is a given function. Divide the interval of integration into a set of small subintervals with end points p_j, $j=1,\ldots,N$. Then Eq. (21) can be rewritten as

$$\overline{F(I,\theta)}=\int dI'd\theta'F(I',\theta')\times\sum_{j=1}^{N-1}\int_{p_j}^{p_{j+1}}dp_0\delta(I'-I(r,p_0,z_s))\times\delta(\theta'-\theta(r,p_0,z_s)). \tag{22}$$

But each integral over p_0 in this sum is a particular case of Eq. (15) with an interval (p_j,p_{j+1}) playing the role of $\mathcal{R}$. From Eqs. (15)–(18), we find

$$\frac{1}{p_{j+1}-p_j}\int_{p_j}^{p_{j+1}}dp_0\delta(I'-I(r,p_0,z_s))\delta(\theta'-\theta(r,p_0,z_s))=\frac{1}{\omega'(I_0)}P_{x\eta}\left(I'-I_0,\frac{\theta'-\theta_0-\omega(I_0)r}{\omega'(I_0)},r\right), \tag{23}$$

where I_0 and θ_0 are starting action and angle of a ray with $p_0=(p_j+p_{j+1})/2$. Substituting this into Eq. (22) and replacing the summation over j by integration over p_0 (with $p_{j+1}-p_j$ replaced by dp_0) yields

$$\overline{F(I,\theta)}=\int_{-p_{\max}}^{p_{\max}}dp_0\int dxd\eta F(I_0+x,\theta_0+\omega(I_0)r+\omega'(I_0)\eta)P_{x\eta}(x,\eta,r). \tag{24}$$

If function F does not depend on θ, Eq. (24) reduces to

$$\overline{F(I)}=\int_{-p_{\max}}^{p_{\max}}dp_0\int dIF(I)P_{I|I_0}(I,r|I_0). \tag{25}$$

The above expressions for $\overline{F(I,\theta)}$ and $\overline{F(I)}$, as well as expressions for both PDFs, $P_{x\eta}(x,\eta,r)$ and $P_I(I,r|I_0)$, do not depend on a particular realization of perturbation δc.

III. MODAL STRUCTURE OF A MONOCHROMATIC WAVE FIELD

A. Normal modes in the WKB approximation

The wave field at any range point can be presented in the form of decomposition

$$u(r,z,\Omega)=\sum_m a_m(r,\Omega)\varphi_m(z,\Omega), \tag{26}$$

where $\varphi_m(z,\Omega)$ are eigenfunctions of the Sturm–Liouville problem in the unperturbed waveguide.[23,34] Each term in this sum describes a normal mode. For simplicity, we restrict our attention to modes with turning points within the water bulk. In the WKB approximation, the mth eigenfunction is determined by parameters of an unperturbed ray path whose action variable—we denote it by $I_{m\Omega}$—satisfies the quantization rule[23,34,35]

$$I_{m\Omega}=\frac{m-1/2}{k}=\frac{m-1/2}{\Omega}c_r,\quad m=1,2,\ldots. \tag{27}$$

The eigenfunction of the mth mode between its turning points can be presented as[23,34,35]

$$\varphi_m(z,\Omega)=\varphi_m^+(z,\Omega)+\varphi_m^-(z,\Omega), \tag{28}$$

where

$$\varphi_m^{\pm}(z,\Omega) = [D(I_{m\Omega})P(I_{m\Omega},z)]^{-1/2} \times \exp\{\pm i[kG(I_{m\Omega},z) - \pi/4]\}. \quad (29)$$

B. Ray-based evaluation of mode amplitudes

The ray-based description of mode amplitudes $a_m(r,\Omega)$ can be derived by projecting the ray representation of the wave field onto normal modes and evaluating the corresponding integrals using the stationary phase technique. A detailed discussion of this issue is given in Refs. 6–9 and 11. It turns out that each mode is formed by contributions from rays—we will call them the mode rays—whose actions at the observation range r satisfy the condition

$$I(r,I_0,\theta_0) = I_{m\Omega}. \quad (30)$$

It should be emphasized that at different frequencies, this condition singles out different rays. In this sense, rays contributing to the mode with the given number m are frequency dependent.

If the action of the ray path is expressed by function $I(r,p_0,z_0)$ Eq. (30) translates to

$$I(r,p_0,z_0) = I_{m\Omega}. \quad (31)$$

We will consider two types of sources.

Point source. In this case, the wave field is a solution of Eq. (2) with an initial condition

$$u(0,z,\Omega) = \delta(z - z_s), \quad (32)$$

where z_s is the source depth. At $x=0$ all rays start from $z_0 = z_s$. Then according to Eq. (31) we get the condition

$$I(r,p_0,z_s) = I_{m\Omega}, \quad (33)$$

defining the starting momenta of mode rays. Take one of these rays and denote its coordinate at the range of observation by $Z_{m\Omega}$. Its contribution to the mode amplitude is[7–9,11]

$$a_m(r,\Omega) = Q\exp[i(\Phi + \beta)], \quad (34)$$

where

$$\Phi = k[S(r,Z_{m\Omega}) + \sigma G(I_{m\Omega},Z_{m\Omega})], \quad (35)$$

$\sigma = -\mathrm{sgn}\, p$, p is the momentum of the mode ray,

$$Q = \frac{1}{\sqrt{2\pi|\partial I(r,p_0,z_s)/\partial p_0|}}, \quad (36)$$

where β is a constant that does not depend on frequency. An explicit expression for β is given in Refs. 7–9 and 11. In the present work, it will not be used.

Single-mode source. In this case,

$$u(0,z,\Omega) = \varphi_{m_0}(z,\Omega). \quad (37)$$

Rays escape from depths z_0 between turning points of the m_0th mode. There are two rays escaping each z_0 with starting momenta $p_0 = \pm P(I_{m_0\Omega},z_0)$. All rays have the same initial action $I_{m_0\Omega}$. The condition that singles out mode rays follows from Eq. (30) as

$$I(r,I_{m_0\Omega},\theta_0) = I_{m\Omega}. \quad (38)$$

It determines θ_0 (hence z_0) corresponding to mode rays. A contribution from an individual mode ray to a_m again has the form of Eq. (34) with the same expression for the phase Φ and

$$Q = \frac{1}{\sqrt{2\pi k|\partial I(r,I_{m_0\Omega},\theta_0)/\partial\theta_0|}}\exp[\pm ikG(I_{m_0\Omega},z_0)]. \quad (39)$$

The mode amplitude is evaluated by summing up contributions from all the mode rays.

C. Redistribution of acoustic energy between normal modes under conditions of ray chaos

Let us consider squared mode amplitudes $|a_m(r,\Omega)|^2$ which we call the mode intensities. Our task will be to derive an analytical expression for a smoothed mode intensity

$$J_m(r,\Omega) = \sum_{m'} |a_{m'}(r,\Omega)|^2 e^{-(m-m')^2/2\mu^2} \Big/ \sum_{m'} e^{-(m-m')^2/2\mu^2}, \quad (40)$$

where μ is a smoothing scale. Under conditions of ray chaos, the number of mode rays contributing to a given mode at a long range becomes very large. Then it is natural to expect that a rough estimate of $J_m(r,\Omega)$ may be obtained by incoherent summation of these rays. Numerical results presented in Refs. 8 and 36 support this expectation. Analytical evaluation of an incoherent sum of mode rays expressing $J_m(r,\Omega)$ may be performed on the basis of approach from Sec. II D that allows one to replace the summation of rays by statistical averaging.

For a single mode source determined by Eq. (37), an incoherent sum of mode rays representing the mode intensity, according to Eqs. (34) and (39), can be presented in the form

$$|a_m(r,\Omega)|^2 = \frac{1}{2\pi k}\sum_n \frac{1}{|\partial I(r,I_{m_0\Omega},\theta_0)/\partial\theta_0|_{\theta_0=\theta_{0,n}}} = \frac{1}{2\pi k}\int_0^{2\pi} d\theta_0\,\delta(I_{m\Omega} - I(r,I_{m_0\Omega},\theta_0)), \quad (41)$$

where the index n in the middle expression numbers the contributing mode rays. The integral in the last expression has the form of Eq. (21) with $F=\mathrm{const}$ $\delta(I_{m\Omega}-I)$ and the integration over θ_0 instead of p_0. It can be treated in the same manner as the integral on the right of Eq. (21). Dividing the interval of integration into small subintervals, we note that the integral over each subinterval once again has the form of Eq. (15). For a subinterval $(\theta_0,\theta_0+\delta\theta)$, we have

$$\frac{1}{\delta\theta}\int_{\theta_0}^{\theta_0+\delta\theta} d\theta_0\,\delta(I_{m\Omega} - I(r,I_{m_0\Omega},\theta_0)) = P_I(I_{m\Omega},r|I_{m_0\Omega}). \quad (42)$$

Since the right hand side does not depend on θ_0, we get

$$|a_m(r,\Omega)|^2 = \frac{1}{k} P_I(r, I_{m\Omega}|I_{m_0\Omega}). \tag{43}$$

According to Eq. (27), the summation over m in Eq. (40) may be approximately replaced by an integration over the action variable. Assuming $m \gg \mu$, we can formally extend the integration over m' to the infinite limits. Then

$$\sum_{m'} e^{-(m-m')^2/2\mu^2} = \int_{-\infty}^{\infty} dm' e^{-(m-m')^2/2\mu^2} = k\int_{-\infty}^{\infty} dI' e^{-k^2(I_{m\Omega}-I')^2/2\mu^2} = \sqrt{2\pi}\mu. \tag{44}$$

Substituting this in Eq. (40) and replacing the summation over m by integration over I, as in Eq. (44), we get

$$J_m(r,\Omega) = \frac{1}{\sqrt{2\pi}\mu}\int dI' P_I(r,I'|I_{m_0\Omega}) \times \exp\left[-\frac{k^2(I_{m\Omega}-I')^2}{2\mu^2}\right]. \tag{45}$$

Substituting Eq. (19) yields

$$J_m(r,\Omega) = \frac{1}{\sqrt{2\pi(\mu^2+k^2Br)}}\exp\left[-\frac{(m-m_0)^2}{2(\mu^2+k^2Br)}\right]. \tag{46}$$

This result is valid only for high modes. At a carrier frequency of 75 Hz, the condition (20) in our environmental model is met only for mode rays corresponding to $m>7$. For treating low modes, one should use Eq. (45) with a generalized version of formula for $P_I(r,I|I_0)$ that describes both steep and flat rays.[11,12] Note that according to Eq. (46) at very long ranges, the number of effectively excited modes grows like $r^{1/2}$. In Ref. 5, a similar range dependence was observed in numerical simulations.

To check the validity of Eq. (46), we have solved (numerically) the parabolic equation (2) at a carrier frequency of 75 Hz with an initial condition $u(0,z)=\varphi_{24}(z)$. Mode intensities at range $r=3000$ km are shown in the upper panel of Fig. 2 for two realizations of random perturbation. The lower panel presents results for the smoothed mode intensities evaluated with a smoothing scale $\mu=4$. Thin solid lines graph J_m for four realizations of perturbation (values of $|a_m|^2$ for two of them are presented in the upper panel). The heavy dashed line shows the prediction of Eq. (46). It is seen that the smoothed mode intensity weakly depends on a particular realization of perturbation and our analytical estimate is in a reasonable agreement with simulations.

In Ref. 11, it is shown that the wave field excited by a point source can be treated in a similar way.

IV. TRANSIENT WAVE FIELD

We now turn our attention to studying the modal structure of a narrowband pulse signal radiated by a point source. Our task is to derive a ray-based description of what we call the mode pulses, that is, pulses carried by individual modes. The mode pulse is defined as

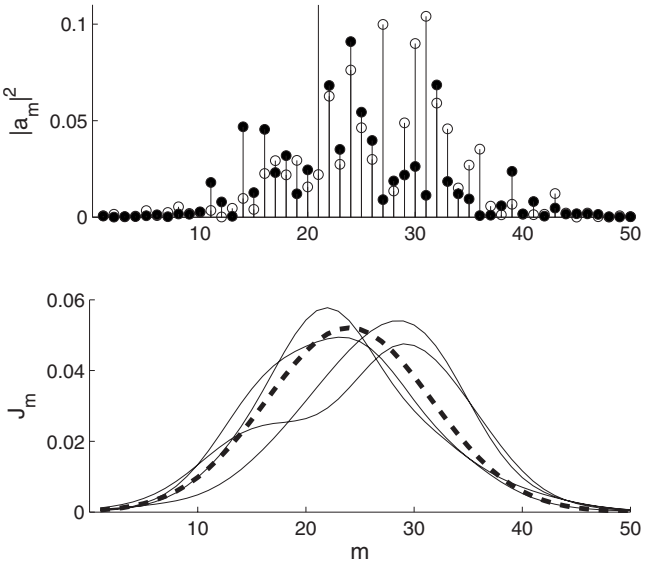


FIG. 2. Mode intensities at range 3000 km for a single mode source exciting the 24th mode $[u(0,z)=\varphi_{24}(z)]$ at a carrier frequency of 75 Hz. (Upper panel) Open and solid circles show mode intensities for two realizations of perturbation. (Lower panel) Smoothed mode intensities computed for four realizations of perturbation (thin solid lines) and prediction of Eq. (46) (heavy dashed line). The smoothing scale $\mu=4$.

$$\psi_m(t,r,\Omega) = \int d\Omega' a_m(r,\Omega') w(\Omega-\Omega') e^{i\Omega'(r/c_r-t)}, \tag{47}$$

where $a_m(r,\Omega')$ are mode amplitudes of the wave field satisfying the parabolic equation (2) with an initial condition (32), and

$$w(\Omega) = \frac{1}{\sqrt{2\pi}\Delta_\Omega}\exp\left(-\frac{\Omega^2}{2\Delta_\Omega^2}\right) \tag{48}$$

is a weighting function. This definition implies that the point source emits sound pulse $g(t)=\exp(-i\Omega t-\Delta_\Omega^2 t^2/2)$, whose bandwidth and duration we estimate as $\delta\Omega=\sqrt{2\pi}\Delta_\Omega$ and $\tau_s=2\pi/\delta\Omega=\sqrt{2\pi}/\Delta_\Omega$, respectively. It is assumed that the radiated wave field is recorded on a long vertical array and at each frequency is decomposed into a sum of normal modes.[37] Then the mode pulses are synthesized in accord with Eq. (47). The argument Ω of function ψ_m indicates the central frequency of an emitted signal. Each mode pulse is a wave packet spreading with range.

A. Arrival time of an instantaneous frequency

As in the case of monochromatic source, we will proceed from the ray-based representation of the mode amplitude described in Sec. III B. Take a mode ray contributing to the mth mode at a central frequency of emitted pulse Ω. Its starting momentum—denote it by p_0^*—satisfies the condition (33) $I(r,p_0^*,z_s)=I_{m\Omega}$. It is natural to assume that there exists a bundle of rays with starting momenta from a small interval centered at p_0^* which are mode rays for the mth mode at frequencies close to Ω: for a ray escaping the source with momentum $p_0^*+\delta p$, there exists frequency Ω' such that $I(r,p_0^*+\delta p,z_s)=I_{m\Omega'}$. The contribution from this bundle to the mode pulse is given by an integral

$$\delta\psi_m(r,t) = \int d\Omega' w(\Omega - \Omega') Q \exp\{i\Phi - i\Omega' t + i\beta\} \quad (49)$$

obtained by insertion of Eq. (34) into Eq. (47). At long ranges, the phase Φ rapidly varies with Ω' and the spreading of the wave packet $\delta\psi_m(r,t)$ can be investigated using the stationary phase technique. The stationary phase point is determined by condition

$$t = \partial\Phi/\partial\Omega'. \quad (50)$$

Frequency Ω' satisfying this equation is interpreted as an instantaneous frequency of the wave packet at time t. For a given Ω', Eq. (50) predicts the arrival time of a constituent of the mode pulse with this instantaneous frequency.

Since β does not depend on frequency, from Eqs. (35) and (27), we find

$$\frac{\partial\Phi}{\partial\Omega'} = \frac{1}{c_r}[r + S(r, Z_{m\Omega'})] + t_1 + t_2, \quad (51)$$

where

$$t_1 = \frac{1}{c_r}\left[\frac{\partial S(r,z)}{\partial z} + \sigma\frac{\partial G(z,I)}{\partial z}\right]_{z=Z_{m\Omega'},\ I=I_{m\Omega'}} \frac{\partial Z_{m\Omega'}}{\partial\Omega'} \quad (52)$$

and

$$t_2 = \frac{\sigma}{c_r}\left[G(Z_{m\Omega'}, I_{m\Omega'}) - I_{m\Omega'} \left.\frac{\partial G(Z_{m\Omega'}, I)}{\partial I}\right|_{I=I_{m\Omega'}}\right]. \quad (53)$$

Term t_1 vanishes because, as it is shown in Refs. 9 and 11, the quantity in square brackets in Eq. (52) is identically zero. According to Eqs. (7), (9), and (27), $G(z,I)$ is a monotonic function of I and $G(Z_{m\Omega'}, I_{m\Omega'})$ is always less than $mT'c_r/2$, where $T' = 2\pi/\Omega'$. Therefore, typically, the magnitude of t_2 is a few times less than $mT'/2$. We will see that this is significantly less than the mode pulse length. Therefore, the arrival time of a fragment of the wave packet $\delta\psi_m(r,t)$ with an instantaneous frequency Ω' can be estimated as

$$t(r,\Omega',m) = \frac{1}{c_r}[r + S(r, Z_{m\Omega'})]. \quad (54)$$

We will call this quantity the arrival time of an instantaneous frequency Ω'. It is the travel time of a sound pulse through the mode ray corresponding to given r, m, and Ω'.

Two comments concerning this result are noteworthy. First, a geometrical ray may be the mode ray for more than one mode. This occurs if its action I at the observation range satisfies the condition $I = I_{m_1\Omega_1} = I_{m_2\Omega_2}$, where $m_1 \neq m_2$ and both Ω_1 and Ω_2 belong to the frequency band of a radiated signal. Second, generally at the observation range, there are several rays with equal travel times and different actions. It means that, typically, the instantaneous frequency of the mode pulse cannot be defined: at any moment t, the pulse is a superposition of signals with different instantaneous frequencies.

B. Mode pulses in the presence of internal waves

At long ranges where the chaos is well developed, the mode pulse is formed by many mode rays. We will sum up their contributions incoherently. Our task will be to derive an approximate analytical description of a smoothed mode pulse defined as

$$Y_m(r,\Omega,t) = \frac{1}{\sqrt{2\pi}\Delta_t}\int dt' |\psi_m(r,\Omega,t')|^2 \times \exp\left[-\frac{(t-t')^2}{2\Delta_t^2}\right], \quad (55)$$

where Δ_t is a smoothing scale. Substituting Eqs. (47) and (48) into Eq. (55), we get

$$Y_m(r,\Omega,t) = \frac{1}{2\pi\Delta_\Omega^2}\int d\Omega_1 d\Omega_2 a_m(r, \Omega_1 + \Omega_2/2) \times a_m^*(r, \Omega_1 - \Omega_2/2)\exp\left[-i\Omega_r t - \frac{(\Omega-\Omega_1)^2}{\Delta_\Omega^2} - \frac{1}{2}\left(\frac{1}{2\Delta_\Omega^2} + \Delta_t^2\right)\Omega_2^2\right]. \quad (56)$$

Once again we present the mode amplitude $a_m(r,\Omega)$ as a sum of contributions from mode rays expressed by Eq. (34). An interval of integration over Ω_2 is of order $\max(\Delta_\Omega, \Delta_t^{-1})$. Assuming that this interval is small enough, we will use an approximation

$$a_m(r,\Omega_1 + \Omega_2/2)a_m^*(r,\Omega_1 - \Omega_2/2) = \frac{1}{2\pi|\partial I(r,p_0,z_s)/\partial p_0|_{p_0=p_{0\Omega_1}}} e^{i\Omega_2\partial\Phi(\Omega_1)/\partial\Omega_1}, \quad (57)$$

where the symbol $p_{0\Omega_1}$ denotes the starting momentum of a mode ray contributing to the mth mode at frequency Ω_1. The action variable of this ray at the range of observation, according to Eq. (27) (for short, we replace $m-1/2$ by m), is

$$I = mc_r/\Omega_1. \quad (58)$$

Equation (57) is our main approximation. It implies that (i) the mode pulse is a superposition of (practically independent) pulses associated with bundles of mode rays described in Sec. IV A and (ii) contributions from these pulses are summed up incoherently.

Substituting Eq. (57) in Eq. (56) and integrating over Ω_2 yield

$$Y_m(r,\Omega,t) = \frac{1}{(2\pi)^{3/2}\Delta_\Omega\sqrt{1/2 + \Delta_\Omega^2\Delta_t^2}}\int \frac{d\Omega_1}{|\partial I/\partial p_0|_{p_0=p_{0\Omega_1}}} \times \exp\left[-\frac{[t - t(r,\Omega_1,m)]^2}{2\Delta_t^2 + \Delta_\Omega^{-2}} - \frac{(\Omega-\Omega_1)^2}{\Delta_\Omega^2}\right]. \quad (59)$$

From Eq. (58), it follows that $d\Omega_1 = -mc_r I^{-2} dI$. Using this relation, we can change the variable of integration in Eq. (59) from Ω_1 to p_0. Then

$$Y_m(r,\Omega,t)=\frac{c_r m}{(2\pi)^{3/2}\Delta_\Omega\sqrt{1/2+\Delta_\Omega^2\Delta_t^2}}\int\frac{dp_0}{I^2(r,p_0,z_s)}\times\exp\left[-\frac{[t-t(r,p_0,z_s)]^2}{2\Delta_t^2+\Delta_\Omega^{-2}}-\frac{[\Omega-mc_r/I(r,p_0,z_s)]^2}{\Delta_\Omega^2}\right]. \quad (60)$$

Equation (60) presents the smoothed mode pulse as an integral over all paths escaping the point source. This integral has the form of Eq. (21) and therefore it can be approximately evaluated by means of Eq. (24). Making use of relation

$$t(r,p_0,z_s)=\bar{t}(I_0)+\eta(r)\gamma(I_0), \quad (61)$$

where

$$\gamma(I)=\frac{I}{c_r}\frac{d\omega(I)}{dI}, \quad (62)$$

that follows from Eq. (13) ($2\pi\Delta N$ is approximated by y), we find

$$Y_m(r,\Omega,t)=\frac{c_r m}{(2\pi)^{3/2}\Delta_\Omega\sqrt{1/2+\Delta_\Omega^2\Delta_t^2}}\times\int\frac{dp_0 dx d\eta}{(I_0+x)^2}P_{x,\eta}(x,\eta,r)\times\exp\left[-\frac{[t-\bar{t}(I_0)-\eta\gamma]^2}{2\Delta_t^2+\Delta_\Omega^{-2}}-\frac{\left(\Omega-\frac{mc_r}{I_0+x}\right)^2}{\Delta_\Omega^2}\right], \quad (63)$$

where I_0 is the starting action of ray escaping the point source with initial momentum p_0.

To simplify this expression, consider function

$$\Psi(I)=\exp\left[-\frac{(\Omega-mc_r/I)^2}{\Delta_\Omega^2}\right] \quad (64)$$

and expand its argument in powers about the point $I=I_{m\Omega}=mc_r/\Omega$. This yields

$$\Psi(I)=\exp\left\{-\frac{m^2c_r^2}{I_{m\Omega}^4\Delta_\Omega^2}\Delta I^2+\frac{2m^2c_r^2}{I_{m\Omega}^5\Delta_\Omega^2}\Delta I^3+O(\Delta I^4)\right\}, \quad (65)$$

where $\Delta I=I-I_{m\Omega}$. If only the first term in the argument of the exponential is retained, then $\Psi(I)$ is of order unity for $|\Delta I|\simeq\Delta I_{cr}$, where $\Delta I_{cr}=I_{m\Omega}^2\Delta_\Omega/(mc_r)=mc_r\Delta_\Omega/\Omega^2$. Substituting ΔI_{cr} for ΔI into the second term in the exponential, we conclude that for our narrowband pulses satisfying the condition $\Delta_\Omega\ll\Omega$ this term is small compared to unity and therefore it can be neglected. The same is true of all terms of higher order. Therefore, we can retain only the first term in the argument of the exponential in Eq. (65).

A further simplification of $\Psi(I)$ is possible if $\Delta I_{cr}\ll\sqrt{Br}$. Then an effective interval of integration over x in Eq. (63) is small compared with the corresponding scale of function $P_{x,\eta}(x,\eta,r)$, and $\Psi(I)$ can be approximated by delta-function

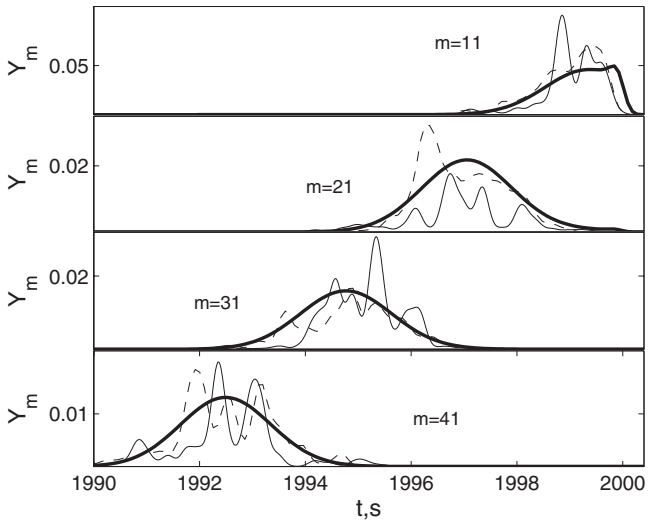


FIG. 3. The smoothed mode pulses at 3000 km for $m=11$, 21, 31, and 41. Parabolic equation based simulations (thin solid lines) are compared to predictions of Eqs. (60) (dashed lines) and (68) (heavy solid lines).

$$\Psi(I)=\frac{\sqrt{\pi}I_{m\Omega}\Delta_\Omega}{\Omega}\delta(I-I_{m\Omega}). \quad (66)$$

In this approximation, Eq. (63) reduces to

$$Y_m(r,\Omega,t)=\frac{1}{2\pi\sqrt{1+2\Delta_\Omega^2\Delta_t^2}}\int dp_0 d\eta P_{x,\eta}(I_{m\Omega}-I_0,\eta,r)\times\exp\left[-\frac{[t-\bar{t}(I_0)-\eta\gamma]^2}{2\Delta_t^2+\Delta_\Omega^{-2}}\right]. \quad (67)$$

Substituting Eq. (18) and evaluating a Gaussian integral over η yield

$$Y_m(r,\Omega,t)=\frac{1}{(2\pi)^{3/2}\Delta_\Omega\sqrt{Br}}\int\frac{dp_0}{\sqrt{2\Delta_t^2+\Delta_\Omega^{-2}+Br^3\gamma^2(I_0)/6}}\times\exp\left[-\frac{(I_{m\Omega}-I_0)^2}{2Br}\right]\times\exp\left\{-\frac{[t-\bar{t}(I_0)-\gamma(I_0)r(I_{m\Omega}-I_0)/2]^2}{2\Delta_t^2+\Delta_\Omega^{-2}+\gamma^2(I_0)Br^3/6}\right\}. \quad (68)$$

Since Eq. (68) is derived using Eq. (18) at frequencies of about 75 Hz it is valid [like Eq. (46)] only for $m>7$.

Thus, we have two approximate formulas for the smoothed mode pulse given by Eqs. (60) and (68). Equation (60) expresses $Y_m(r,\Omega,t)$ through solutions of the ray equations (ray paths) and it depends on a particular realization of perturbation. In contrast, Eq. (68) is an analytical estimate independent of a particular realization of δc. In Fig. 3, predictions of Eqs. (60) and (68) are compared to results of simulations performed by solving (numerically) the parabolic equation (2). A point source has been set at the sound channel axis $z=z_a$. The simulations have a center frequency of 75 Hz, and a smoothing scale $\Delta_t=0.1$ s. The envelope of an emitted signal is determined by Eq. (48) with $\Delta_\Omega/(2\pi)=2$ Hz. The effective bandwidth of the emitted pulse is about 5 Hz. Smoothed mode pulses for $m=11$, 21, 31, and 41

at the observation range 3000 km are shown by thin solid lines. In order to apply Eq. (60) we have traced numerically 50 000 rays leaving the point source with initial momenta p_0 uniformly filling an interval corresponding to launch angles $\pm 14°$. Replacing the integral over p_0 in Eq. (60) with a sum over the computed ray paths yields pulses shown in Fig. 3 by dashed lines. Predictions of Eq. (68) in Fig. 3 are presented by thick solid lines. It is seen that both our approximate formulas are too rough and cannot give a detailed description of the smoothed mode pulse. Nevertheless, they are able to predict the spread of the pulse and its bias caused by the random perturbation. In Sec. IV C, we will consider this issue.

C. Bias and spread of a mode pulse

Introduce symbol $\langle \ldots \rangle_{m\Omega}$ which denotes the averaging over the smoothed mode pulse:

$$\langle g(t) \rangle_{m\Omega} \equiv \int dt g(t) Y_m(r,\Omega,t) \Big/ \int dt Y_m(r,\Omega,t). \quad (69)$$

Our task is to estimate the bias of the mean travel time of the mth mode pulse relative $\bar{t}(I_{m\Omega})$, arrival time of this pulse in the unperturbed waveguide,

$$\delta t_{m\Omega} \equiv \langle t - \bar{t}(I_{m\Omega}) \rangle_{m\Omega} = \langle t \rangle_{m\Omega} - \bar{t}(I_{m\Omega}), \quad (70)$$

and the spread $\tau_{m\Omega}$ defined as

$$\tau^2_{m\Omega} \equiv \langle [t - \bar{t}(I_{m\Omega})]^2 \rangle_{m\Omega} = \langle t^2 \rangle_{m\Omega} - \bar{t}^2(I_{m\Omega}). \quad (71)$$

The quantities $\langle t \rangle_{m\Omega}$ and $\langle t^2 \rangle_{m\Omega}$ can be evaluated in two ways: using Eq. (60) or Eq. (68).

Ray tracing. From Eq. (60), we find

$$\langle t \rangle_{m\Omega} = \int \frac{dp_0}{I^2(r,p_0,z_s)} t(r,p_0,z_s) \times \exp\left[- \frac{[\Omega - mc_r/I(r,p_0,z_s)]^2}{\Delta^2_\Omega} \right] \Big/ \Xi_{m,\Omega} \quad (72)$$

and

$$\langle t^2 \rangle_{m\Omega} = \int \frac{dp_0}{I^2(r,p_0,z_s)} [t^2(r,p_0,z_s) + \Delta_t^2 + \Delta_\Omega^{-2}/2] \times \exp\left[- \frac{(\Omega - mc_r/I((r,p_0,z_s)))^2}{\Delta^2_\Omega} \right] \Big/ \Xi_{m,\Omega}, \quad (73)$$

where

$$\Xi_{m,\Omega} = \int \frac{dp_0}{I^2(r,p_0,z_s)} \exp\left[- \frac{[\Omega - mc_r/I^2(r,p_0,z_s)]^2}{\Delta^2_\Omega} \right]. \quad (74)$$

Equations (72)–(74) allow one to calculate the bias and spread of the mode pulse on the basis of numerical ray tracing.

Analytical estimates. They follow from Eq. (68) as

$$\langle t \rangle_{m\Omega} = \int dp_0 [\bar{t}(I_0) + \gamma(I_0)(I_m - I_0) r/2] \times \exp\left[- \frac{(I_m - I_0)^2}{2Br} \right] \Big/ K_m \quad (75)$$

and

$$\langle t^2 \rangle_{m\Omega} = \int dp_0 \{ [\bar{t}(I_0) + \gamma(I_0)(I_m - I_0) r/2]^2 + \Delta_t^2 + \Delta_\Omega^{-2}/2 + \gamma^2(I_0) B r^3/12 \} \exp\left[- \frac{(I_m - I_0)^2}{2Br} \right] \Big/ K_m, \quad (76)$$

where

$$K_m = \int dp_0 \exp\left[- \frac{(I_m - I_0)^2}{2Br} \right]. \quad (77)$$

Simplified analytical estimates. In order to simplify Eqs. (75) and (76), we first change the integration variable from p_0 to I_0. Then dp_0 is replaced by $q(I_0) dI_0$, where

$$q(I_0) = \frac{dp_0}{dI_0}. \quad (78)$$

Due to the exponential present in Eqs. (75)–(77), the main contributions to these integrals come from I_0 located within a small interval of order $\sqrt{Br}$ centered at I_m. Equations (75)–(77) can be simplified if variations in $q(I_0)$ and $\gamma(I_0)$ at this interval are negligible, that is, if

$$|\sqrt{Br} q'(I_0)/q(I_0)| \ll 1, \quad |\sqrt{Br} \gamma'(I_0)/\gamma(I_0)| \ll 1, \quad (79)$$

then we get a simple expression for K_m,

$$K_m = \int dI_0 q(I_0) \exp\left[- \frac{(I_{m\Omega} - I_0)^2}{2Br} \right] = \sqrt{2\pi Br} q(I_{m\Omega}). \quad (80)$$

From Eqs. (70) and (75), it follows

$$\delta t_{m\Omega} = \int dI_0 q(I_0) [\bar{t}(I_0) + \gamma(I_0)(I_m - I_0) r/2 - \bar{t}(I_{m\Omega})] \times \exp\left[- \frac{(I_{m\Omega} - I_0)^2}{2Br} \right] \Big/ K_{m\Omega}. \quad (81)$$

Using the condition (79), we may expand the pre-exponential term in the above integral in a power series about $I_{m\Omega}$. From expansions

$$\bar{t}(I_0) + \frac{r}{2} \gamma(I_0)(I_m - I_0) = \bar{t}(I_{m\Omega}) + \frac{r}{2} \gamma(I_{m\Omega}) \Delta I + O(\Delta I^3) \quad (82)$$

and

$$q(I_0) = q(I_{m\Omega}) + q'(I_{m\Omega}) \Delta I + O(\Delta I^2), \quad (83)$$

where $\Delta I = I_0 - I_{m\Omega}$, and neglecting terms $O(\Delta I^3)$, we get

$$\delta t_{m\Omega} = \frac{q'(I_m) \gamma(I_m) B}{2q(I_m)} r^2. \quad (84)$$

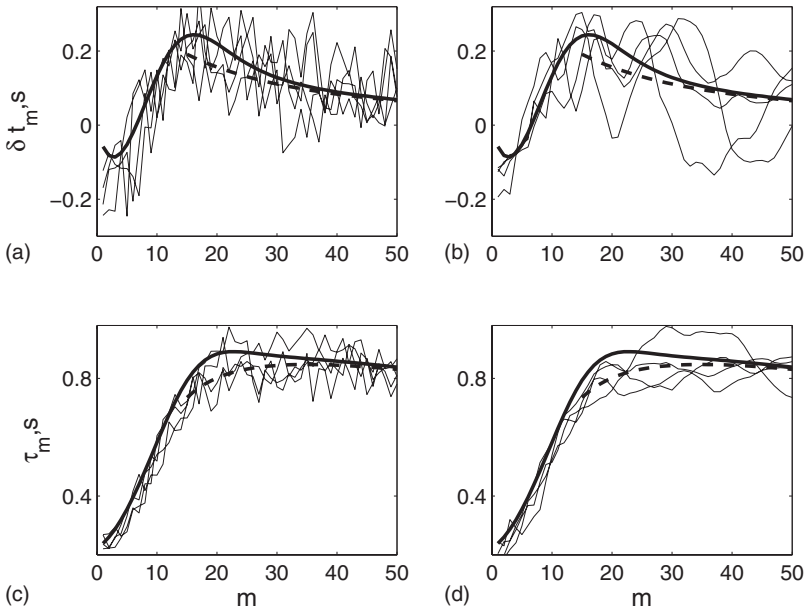


FIG. 4. Biases (upper row of plots) and spreads (lower row) of mode pulses at 3000 km. Heavy solid lines present analytical predictions for the bias [(a) and (b)] and spread [(c) and (d)] at 75 Hz. Heavy dashed line shows simplified analytical estimates for the same quantities. Thin solid lines graph frequency averaged biases, δt_m, and spreads, τ_m, obtained by parabolic equation simulation [(a) and (c)] and predicted on the basis of a numerical ray tracing [(b) and (d)] for four realizations of perturbation.

Now we turn our attention to the pulse spread. We substitute Eq. (82) into Eq. (76), change the integration variable to I_0, and approximately replace $q(I_0)$ and $\gamma(I_0)$ by $q(I_{m\Omega})$ and $\gamma(I_{m\Omega})$, respectively. This yields

$$\langle t^2\rangle_{m\Omega} = \bar{t}^2(I_{m\Omega}) + \Delta_t^2 + \Delta_\Omega^{-2}/2 + \frac{\gamma^2(I_{m\Omega})Br^3}{3}. \tag{85}$$

Correspondingly,

$$\tau_{m\Omega}^2 = \Delta_t^2 + \Delta_\Omega^{-2}/2 + \frac{\gamma^2(I_{m\Omega})Br^3}{3}. \tag{86}$$

Due to restrictions imposed by the condition (79), the estimates (84) and (86) in our environmental model at the frequency of 75 Hz can be used only for modes with $m>14$. Note also that at the range 3000 km, the last term in the right hand side of Eq. (86) strongly dominates. Therefore, this equation can be further simplified to

$$\tau_{m\Omega} = \frac{|\gamma(I_{m\Omega})|B^{1/2}r^{3/2}}{\sqrt{3}}. \tag{87}$$

In a slightly different way, this result was derived in Refs. 11 and 21. An alternative derivation was given in Ref. 22.

Numerical example. To check the applicability of the above results, we have performed the parabolic equation simulation of 100 sound pulses with equal envelope functions $w(\Omega)$ but different central frequencies uniformly filling an interval of 66–84 Hz. The bandwidth of each pulse is determined by the same constant $\Delta_\Omega/(2\pi)=2$ Hz as in Fig. 3. This simulation has been performed for four realizations of perturbation δc. Then we have evaluated the biases, $\delta t_{m\Omega}$, and spreads, $\tau_{m\Omega}$, at the range 3000 km for $m=1,\ldots,50$, and for 100 values of Ω corresponding to the above central frequencies. We will call the values of $\delta t_{m,\Omega}$ and $\tau_{m,\Omega}$ averaged over the 100 frequencies the frequency averaged bias and spread and denote by δt_m and τ_m, respectively.

The estimates of $\delta t_{m\Omega}$ and $\tau_{m\Omega}$ for the same m and Ω have been obtained on the basis of ray tracing. In each of the four realizations of our random waveguide, 50 000 rays with the same starting parameters as indicated in Sec. IV B have been traced numerically. Then estimates of $\delta t_{m\Omega}$, $\tau_{m\Omega}$, δt_m, and τ_m have been calculated using Eqs. (70)–(74).

Thin solid lines in Figs. 4(a) and 4(b) show the frequency averaged bias δt_m obtained using the parabolic equation simulation and numerical ray tracing, respectively, for four realizations of δc. A heavy solid line, the same in both plots, depicts an analytical estimates of $\delta t_{m\Omega}$ at a frequency of 75 Hz predicted by Eqs. (70), (75), and (77). The simplified analytical estimate (84) at 75 Hz is shown by a heavy dashed line. Figures 4(c) and 4(d) present similar results for the spread. A heavy solid line is an analytical estimate of $\tau_{m\Omega}$ at 75 Hz given by Eqs. (71) and (75)–(77). A heavy dashed line is a simplified analytical estimate given by Eq. (87).

In Fig. 5, we see the same dependencies as in Fig. 4 but the thin solid lines present not only frequency averaged results but also results obtained for a single frequency of 75 Hz.

The agreement between simulations and theory based estimates is seen to be good for the spread. It is somewhat surprising that in spite of the restriction indicated after Eq. (68) the analytical estimate based on Eqs. (70), (75), and (77) are valid not only for the high modes. The predictions of bias are less accurate. They can be considered only as order-of-magnitude estimates.

V. CONCLUSION

In the present paper, we discuss a simple analytical approach that allows one to derive rough estimates characterizing the modal structure of the sound field. In the scope of our approach, the mode amplitude is given by a sum of terms (34) representing contributions from individual mode rays. Thus, we have an analog of the geometrical optics for modes with mode rays playing the role of eigenrays.[6–9]

From the viewpoint of stochastic ray theory derived in Refs. 10–12 and 18, and shortly described in Sec. II D, an incoherent summation of chaotic mode rays can be treated as statistical averaging. Moreover, it may be expected that at

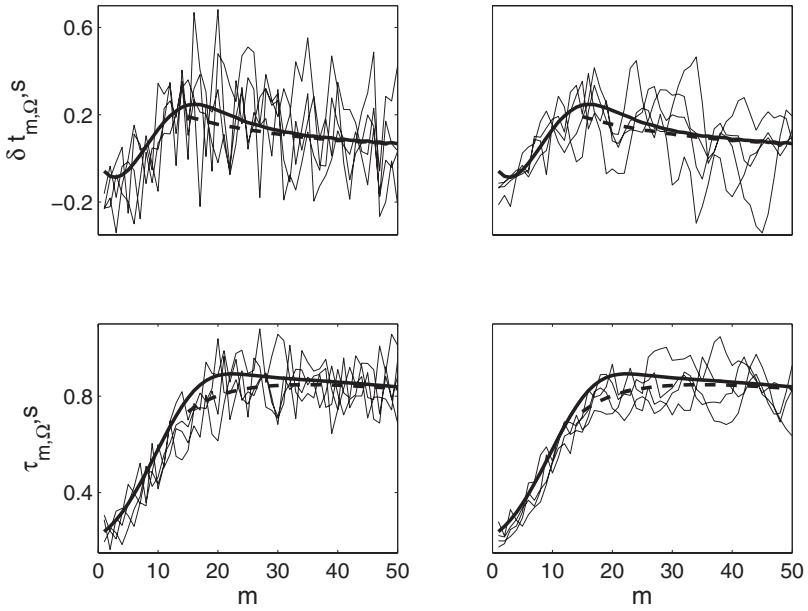


FIG. 5. The same as in Fig. 4 but thin solid lines in all plots present not frequency averaged biases, δt_m, and spreads, τ_m, but values of $\delta t_{m\Omega}$ and $\tau_{m\Omega}$ computed for $\Omega/(2\pi)=75$ Hz.

long ranges the result of this averaging weakly depends on a particular realization of the random perturbation. This expectation is suggested by the law of large numbers: contributions from mode rays are considered as independent random variables with more or less similar PDFs. In Sec. III C, we have derived an approximate analytical expression for a smoothed mode intensity. The smoothing over mode number m increases the number of mode rays to be summed up. The larger the smoothing scale μ, the better are satisfied the applicability conditions of the law of large numbers. This conjecture is supported by results presented in Fig. 2. The smoothed mode intensity depends much less on a particular realization of random perturbation than intensities of individual modes.

The travel time of a mode ray corresponding to the given pair (m,Ω) is interpreted as an arrival time of a fragment of the mth mode pulse with an instantaneous frequency Ω. At a megameter range in a fluctuating ocean, there may be a lot of rays with equal travel times corresponding to pairs (m,Ω) with the same m and different Ω. Therefore, the mode pulse at each moment is formed by contributions from a large number of mode rays. As in the case of a monochromatic source, we sum up their contributions incoherently and, once again, replace summation by statistical averaging.

We have derived formulas of two types for the spread and bias of the mode pulse caused by internal waves. Equations (70)–(74) express these quantities through integrals over ray paths escaping the source. In order to exploit these results, one should perform numerical ray tracing. In contrast, Eqs. (75)–(77) give analytical estimates that do not depend on a particular realization of perturbation. For high modes, the analytical estimates may be further simplified yielding Eqs. (84) and (87). These equations show the same range dependencies of bias and spread as those observed in numerical simulations presented in Ref. 20. Note that while the ray-based predictions of spread are in a reasonable agreement with results of the parabolic equation simulations, the predictions of bias give only order-of-magnitude estimates.

A major limitation of our approach is an insufficient understanding of its applicability conditions. This is a common problem for practically all ray-based approaches. But in our case an additional difficulty arises due to the conjecture that the averaging over ray starting parameters may be replaced by the statistical averaging. Obviously, this may be done only when evaluating smoothed characteristics of the wave field. However, the question of selecting proper smoothing scales remains open and requires a further investigation. The smoothing scales used in our calculations (μ, Δ_t, and Δ_Ω) have been selected empirically.

ACKNOWLEDGMENT

The work was supported by the Grant Nos. 07-02-00255 and 08-05-00596 from the Russian Foundation for Basic Research, the Program "Coherent acoustic fields and signals" of Physical Sciences Division of Russian Academy of Sciences, and the Leading Scientific Schools Grant No. 1055.2008.2.

[1]L. B. Dozier and F. D. Tappert, "Statistics of normal mode amplitudes in a random ocean. I. Theory," J. Acoust. Soc. Am. **63**, 353–365 (1978).
[2]W. Kohler and G. C. Papanicolaou, *Wave Propagation and Underwater Acoustics* (Springer, New York, 1977).
[3]E. Yu. Gorodetskaya, A. I. Malekhanov, A. G. Sazontov, and N. K. Vdovicheva, "Deep water acoustic coherence at long ranges: Theoretical prediction and effects on large array signal processing," IEEE J. Ocean. Eng. **24**, 156–171 (1999).
[4]A. G. Voronovich and V. E. Ostashev, "Low-frequency sound scattering by internal waves in the ocean," J. Acoust. Soc. Am. **119**, 1406–1419 (2006).
[5]A. K. Morozov and J. A. Colosi, "Stochastic differential equation analysis for sound scattering by random internal waves in the ocean," Acoust. Phys. **53**, 335–347 (2007).
[6]G. P. Berman and G. M. Zaslavsky, "Condition of stochasticity in quantum nonlinear systems," Physica A **97**, 367–382 (1979).
[7]A. L. Virovlyansky and G. M. Zaslavsky, "Wave chaos in terms of normal modes," Phys. Rev. E **59**, 1656–1668 (1999).
[8]A. L. Virovlyansky, "Manifestation of ray stochastic behavior in a modal structure of the wave field," J. Acoust. Soc. Am. **108**, 84–95 (2000).
[9]A. L. Virovlyansky, A. Yu. Kazarova, and L. Ya. Lyubavin, "Ray-based description of normal mode amplitudes in a range-dependent waveguide," Wave Motion **42**, 317–334 (2005).

[10]A. L. Virovlyansky, "Ray travel times in range-dependent acoustic waveguides," http://arxiv.org/abs/nlin/0012015 (Last viewed May, 2008).
[11]A. L. Virovlyansky, *Ray Theory of Long-Range Sound Propagation in the Ocean* (Institute of Applied Physics, Nizhny Novgorod, 2006) (in Russian).
[12]A. L. Virovlyansky, A. Yu. Kazarova, and L. Ya. Lyubavin, "Statistical description of chaotic rays in a deep water acoustic waveguide," J. Acoust. Soc. Am. **121**, 2542–2552 (2007).
[13]S. M. Flatté, R. Dashen, W. M. Munk, K. M. Watson, and F. Zakhariasen, *Sound Transmission Through a Fluctuating Ocean* (Cambridge University Press, London, 1979).
[14]J. Simmen, S. M. Flatté, and G.-Y. Wan, "Wavefront folding, chaos, and diffraction for sound propagation through ocean internal waves," J. Acoust. Soc. Am. **102**, 239–255 (1997).
[15]M. G. Brown, J. A. Colosi, S. Tomsovic, A. L. Virovlyansky, M. A. Wolfson, and G. M. Zaslavsky, "Ray dynamics in long-range deep ocean sound propagation," J. Acoust. Soc. Am. **113**, 2533–2547 (2003).
[16]F. J. Beron-Vera, M. G. Brown, J. A. Colosi, S. Tomsovic, A. L. Virovlyansky, M. A. Wolfson, and G. M. Zaslavsky, "Ray dynamics in a long-range acoustic propagation experiment," J. Acoust. Soc. Am. **114**, 1226–1242 (2003).
[17]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).
[18]A. L. Virovlyansky, "Statistical description of ray chaos in an underwater acoustic waveguide," Acoust. Phys. **51**, 71–80 (2005).
[19]M. G. Brown and J. Viechnicki, "Stochastic ray theory for long-range sound propagation in deep ocean environment," J. Acoust. Soc. Am. **104**, 2090–2104 (1998).
[20]J. A. Colosi and S. M. Flatté, "Mode coupling by internal waves for multimegameter acoustic propagation in the ocean," J. Acoust. Soc. Am. **100**, 3607–3620 (1996).
[21]A. L. Virovlyansky, A. Yu. Kazarova, and L. Ya. Lyubavin, "Modal structure of the field under conditions of wave chaos," *Ocean Acoustics, Proceedings of the 11th L. M. Brekhovskikh's Conference* (GEOS, Moscow, 2006), pp. 40–43 (in Russian).
[22]I. A. Udovydchenkov and M. G. Brown, "Modal group time spreads in weakly range-dependent deep ocean environments," J. Acoust. Soc. Am. **123**, 41–50 (2008).
[23]L. M. Brekhovskikh and Yu. P. Lysanov, *Fundamentals of Ocean Acoustics* (Springer-Verlag, Berlin, 1991).
[24]F. D. Tappert, *Wave Propagation and Underwater Acoustics* (Springer, New York, 1977).
[25]S. S. Abdullaev and G. M. Zaslavsky, "Classical nonlinear dynamics and chaos of rays in wave propagation problems in inhomogeneous media," Sov. Phys. Usp. **34**, 645–687 (1991).
[26]L. D. Landau and E. M. Lifshitz, *Mechanics*, 3rd ed. (Pergamon, Oxford, 1976).
[27]W. Munk and C. Wunsch, "Ocean acoustic tomography: Rays and modes," Rev. Geophys. Space Phys. **21**, 777–793 (1983).
[28]A. L. Virovlyanskii, "Travel times of acoustic pulses in the ocean," Sov. Phys. Acoust. **31**, 399–401 (1985).
[29]A. L. Virovlyansky, "On general properties of ray arrival sequences in oceanic acoustic waveguides," J. Acoust. Soc. Am. **97**, 3180–3183 (1995).
[30]A. L. Virovlyansky, "Ray travel times at long ranges in acoustic waveguides," J. Acoust. Soc. Am. **113**, 2523–2532 (2003).
[31]J. A. Colosi and M. G. Brown, "Efficient numerical simulation of stochastic internal-wave-induced sound-speed perturbation field," J. Acoust. Soc. Am. **103**, 2232–2235 (1998).
[32]A. L. Virovlyanskii and G. Z. Zaslavskii, "Ray and wave chaos in problems of sound propagation in the ocean," Acoust. Phys. **53**, 282–297 (2007).
[33]C. W. Gardiner, *Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences* (Springer-Verlag, New York, 1985).
[34]L. M. Brekhovskikh and O. A. Godin, *Acoustics of Layered Media II: Point Sources and Bounded Beams* (Springer-Verlag, Berlin, 1999).
[35]L. D. Landau and E. M. Lifshitz, *Quantum Mechanics* (Pergamon, Oxford, 1977).
[36]A. L. Virovlyansky and G. M. Zaslavsky, "Evaluation of the smoothed interference pattern under conditions of ray chaos," Chaos **10**, 211–223 (2000).
[37]K. E. Wage, A. B. Baggeroer, and J. C. Preisig, "Modal analysis of broadband acoustic receptions at 3515-km range in the north pacific using short-time Fourier techniques," J. Acoust. Soc. Am. **113**, 801–817 (2003).

Prediction of acoustic radiation from axisymmetric surfaces with arbitrary boundary conditions using the boundary element method on a distributed computing system

Louise Wright[a)]
Mathematics and Scientific Computing Group, National Physical Laboratory,[b)] Teddington, Middlesex TW11 0LW, United Kingdom

Stephen P. Robinson
Acoustics Group, National Physical Laboratory, Teddington, Middlesex TW11 0LW, United Kingdom

Victor F. Humphrey
Institute of Sound and Vibration Research, University of Southampton, Southampton SO17 1BJ, United Kingdom

(Received 12 February 2007; revised 28 November 2008; accepted 1 December 2008)

This paper presents a computational technique using the boundary element method for prediction of radiated acoustic waves from axisymmetric surfaces with nonaxisymmetric boundary conditions. The aim is to predict the far-field behavior of underwater acoustic transducers based on their measured behavior in the near-field. The technique is valid for all wavenumbers and uses a volume integral method to calculate the singular integrals required by the boundary element formulation. The technique has been implemented on a distributed computing system to take advantage of its parallel nature, which has led to significant reductions in the time required to generate results. Measurement data generated by a pair of free-flooding underwater acoustic transducers encapsulated in a polyurethane polymer have been used to validate the technique against experiment. The dimensions of the outer surface of the transducers (including the polymer coating) were an outer diameter of 98 mm with an 18 mm wall thickness and a length of 92 mm. The transducers were mounted coaxially, giving an overall length of 185 mm. The cylinders had resonance frequencies at 13.9 and 27.5 kHz, and the data were gathered at these frequencies.
[DOI: 10.1121/1.3056467]

PACS number(s): 43.30.Jx, 43.58.Ta, 43.58.Vb, 43.30.Xm [KGF] Pages: 1374–1383

I. INTRODUCTION

The work described in this paper uses boundary element methods to predict the far-field response from discrete samples of the complex near-field acoustic pressure measured on cylindrical surfaces. The aim of the work has been to develop a methodology that can be applied to real measurement data gathered from underwater acoustic transducers with a minimum of preprocessing. This paper focuses on the numerical technique used to evaluate the integrals that result from reformulating the problem as a two-dimensional axisymmetric problem, explains the implementation of the methodology on a distributed computing system, and presents some results that have been generated using measured data. Many of the details of the measurement procedure and the numerical approach are not described in detail in this paper. More extensive descriptions of these aspects can be found in Refs. 1–4.

The generation and detection of acoustic fields in water are typically undertaken using electroacoustic transducers, either deployed individually or in arrays. Such devices are most often characterized in terms of absolute sensitivity levels (which is important since source level and receive sensitivity are often used to determine range or detection limits) and device directivity (which is important for beam forming and for ambient noise rejection).[5]

Measurement of the transmit and receive sensitivity of transducers and arrays may be carried out in a variety of types of facility, for example, laboratory tanks or open-water sites. At NPL, such facilities include laboratory tanks, the largest of which is a cylindrical wooden tank of 5.5 m diameter and 5 m depth; an open-water facility located on a reservoir with 20 m water depth; and an acoustic pressure vessel, which can simulate ocean conditions for depths of up to 700 m and water temperatures from 2 to 35 °C.

The sensitivities and directional response are usually required to be determined in the acoustic far-field. However, if the array or transducer is physically large when measured in acoustic wavelengths (i.e., the value of ka is high, where k is the wavenumber and a is the largest physical dimension), it may be impossible for far-field conditions to be achieved in a facility of finite size while maintaining a free-field environment and preserving steady state conditions. An open-water facility will, in general, enable a greater source-receiver separation to be used, but such facilities have the disadvantage that there is little or no control of environmental conditions, and, in any case, even an open-water facility will place

[a)]Electronic mail: louise.wright@npl.co.uk
[b)]URL: www.npl.co.uk/ssfm

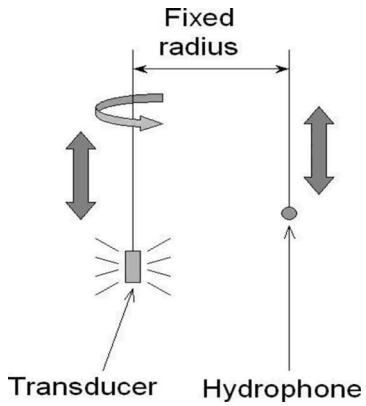


FIG. 1. Sketch showing the arrangement and relative motion of transducer (sound source) and hydrophone (receiver) during the measurement process.

some limit on the maximum separation achievable. Making measurements at sea is extremely expensive (often an order of magnitude more expensive than any tank or open-water measurements), provides no environmental control, and is usually reserved for evaluation of full system performance. There is consequently a strong motivation to maximize the range of acoustic testing possible in laboratory tanks.[6]

An approach to overcome the restrictions posed by finite-sized laboratory tanks is to undertake measurements in the acoustic near-field and then predict the acoustic far-field response from the near-field data.[7,8] Several general approaches to calibration in the near-field have been reported in the literature, including the use of Fourier acoustic methods,[9] the use of near-field calibration arrays,[10,11] and an approach that solves the integral form of the Helmholtz equation by the use of boundary element methods.[12,13] The latter approach has been adopted in this paper.

Section II describes the measurement process used to collect the data for validation of the numerical technique. Section III outlines the initial problem formulation and identifies some difficulties associated with the solution of the problem in its simplest form. Section IV explains a reformulation of the problem that avoids some of the difficulties and describes how the application of distributed computing can improve the efficiency of the solution of the reformulated problem. Section V gives details of methods of evaluation for some of the singular integrals required for the solution of the problem. Section VI describes the validation procedure for the computation, shows some calculated results, and compares them with measured data. Finally, Sec. VII summarizes the paper and presents the conclusions drawn from the work.

II. DETAILS OF THE MEASUREMENT PROCESS

Measurements of acoustic pressure amplitude and phase are gathered in water on open-ended cylindrical surfaces within NPL's largest open tank. The axes of the source transducer and the receiving hydrophone are fixed parallel a set distance apart, as shown in Fig. 1. This distance is the radius of the cylinder. For each scan line, the vertical distance between the transducer and the hydrophone is fixed, and data are gathered as the transducer is slowly rotated incrementally, so that the hydrophone measures the pressure field at set angles around a circle at a height relative to the transducer. Measurements are made along arcs at different heights, with the transducer both above and below the hydrophone, and a cylindrical surface scan similar to that shown in Fig. 2 is assembled from the scan lines.

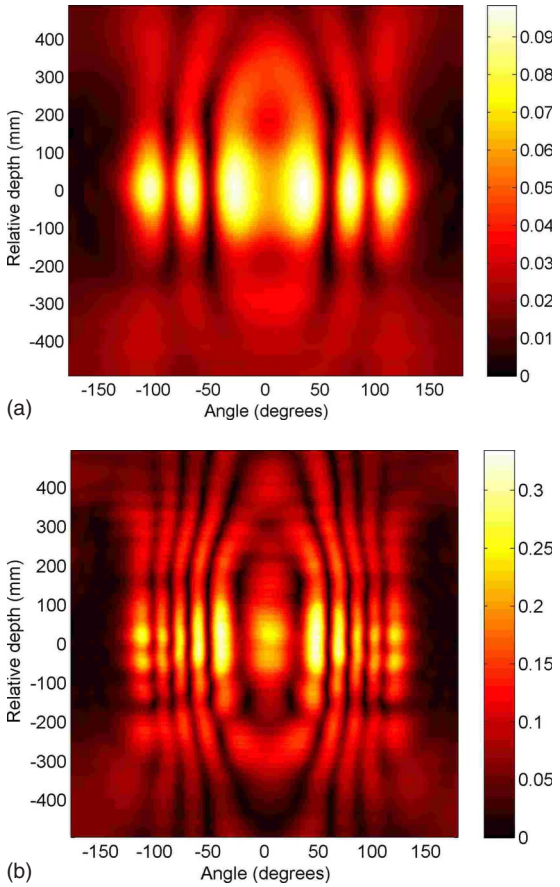


FIG. 2. (Color online) Two typical sets of pressure amplitude data from a transducer configuration with a reflecting plate, as described in Sec. VI A. (a) was measured at a frequency of 13.9 kHz, and (b) was measured at 27.5 kHz. The data were gathered on cylinders that are shown unwrapped in this figure. Both cylinder radii were 0.338 m, and scan lines were made between −0.49 and 0.49 m, where the transducer is at depth 0 m.

Table I shows the separations between near-field measurements in the vertical and azimuthal directions for the two frequencies considered in this paper. Sound speed throughout is taken to be 1477 ms^{-1}. The measurements made on these surfaces are used as input data for the computational tech-

TABLE I. Summary of the measurement surfaces for the near-field measurements.

Frequency (kHz)	13.9	27.5
Wavelength (mm)	106	53.7
Cylinder radius (m)	0.338	0.338
Total cylinder height (m)	0.98	0.98
Vertical separation between scan lines (mm)	10	5
Angular separation between scan points (deg)	2	1
Angular separation between scan points (mm)	12	6

nique for calculating far-field behavior. The table shows that measurements were taken approximately every $\lambda/10$. The same hydrophone and measurement system were used for all measurements, and gain settings were not altered between near-field and far-field scans. However, all amplitude data presented in this paper are in units of volts rather than pascals (described in the plots as "scaled pressure amplitude"). Therefore, the data shown represent a scaled version of the acoustic pressure amplitudes, the scaling factor being the sensitivity of the hydrophone and preamplifier. Notwithstanding the common scaling factor, the results allow comparison to be made of the absolute amplitudes of the measured and computed far-field acoustic pressures.

III. INITIAL PROBLEM FORMULATION

Small amplitude time-harmonic acoustic waves of constant frequency f (and corresponding wavenumber k) in homogeneous media can be described using the boundary integral formulation of the Helmholtz wave equation[1]

$$\int_S \varphi(\mathbf{q})\frac{\partial G_k(\mathbf{p};\mathbf{q})}{\partial n_q} - G_k(\mathbf{p};\mathbf{q})\frac{\partial \varphi(\mathbf{q})}{\partial n_q} dS_q(\mathbf{q}) = \varepsilon(\mathbf{p})\varphi(\mathbf{p}), \tag{1}$$

where S is a closed surface, $\varepsilon(\mathbf{p})=1$ if $\mathbf{p}$ lies inside S in the internal domain D, $\varepsilon(\mathbf{p})=\frac{1}{2}$ if $\mathbf{p}$ lies on S, and $\varepsilon(\mathbf{p})=0$ if $\mathbf{p}$ is outside of S, in the external domain E, and

$$G_k(\mathbf{p};\mathbf{q}) = \frac{e^{ikr}}{4\pi r}, \tag{2}$$

where

$$r = |\mathbf{p}-\mathbf{q}| = \sqrt{(\rho_p^2+\rho_q^2-2\rho_p\rho_q\cos(\theta_p-\theta_q))+(z_p-z_q)^2}, \tag{3}$$

and the coordinates of point $\mathbf{p}$ in a cylindrical polar system are $\{\rho_p,\theta_p,z_p\}$. Since the value of φ on the closed surface S is known from measurement, in theory $\partial\varphi/\partial n_q$ can be determined for $\mathbf{p}\in S$ from Eq. (1) and can be used to calculate φ for points in D and E. For an arbitrary surface S, this equation must be solved numerically. The formulation of Eq. (1) does not have a unique solution for $\mathbf{p}\in S$ for all values of k. If k is such that there is a nontrivial solution to the related homogeneous problem,

$$\nabla^2\varphi + k^2\varphi = 0, \quad \mathbf{p}\in D, \tag{4}$$

$$\varphi(\mathbf{p}) = 0, \quad \mathbf{p}\in S, \tag{5}$$

then the solution to Eq. (1) is nonunique. These critical values of k are not known analytically for an arbitrary surface. In addition, the numerical solution to Eq. (1) becomes ill conditioned for values of k close to the critical values. As frequency increases, the critical values become more dense amongst the real numbers, independent of the shape of S, and solving Eq. (1) becomes more difficult.

A number of methods have been used to get around the problems described above, for example, the so-called combined Helmholtz integral equation formulation method,[14–16] but for the work described here we have used the Burton–Miller method.[1] This method reformulates the problem as a linear combination of Eq. (1) for $\mathbf{p}\in S$ and its derivative with respect to the normal at $\mathbf{p}$, giving

$$\left[\left(-\frac{1}{2}I + M_k + \mu N_k\right)\varphi\right](\mathbf{p}) = \left[\left(L_k + \mu\left(\frac{1}{2}I + M_k^T\right)\right)\frac{\partial\varphi}{\partial n_q}\right](\mathbf{p}), \quad \mathbf{p}\in S, \tag{6}$$

where

$$[(L_k)f](\mathbf{p}) = \int_S f(\mathbf{q})G_k(\mathbf{p};\mathbf{q})dS_q(\mathbf{q}), \tag{7}$$

$$[(M_k)f](\mathbf{p}) = \int_S f(\mathbf{q})\frac{\partial G_k(\mathbf{p};\mathbf{q})}{\partial n_q}dS_q(\mathbf{q}), \tag{8}$$

$$[(M_k^T)f](\mathbf{p}) = \int_S f(\mathbf{q})\frac{\partial G_k(\mathbf{p};\mathbf{q})}{\partial n_p}dS_q(\mathbf{q}), \tag{9}$$

$$[(N_k)f](\mathbf{p}) = \int_S f(\mathbf{q})\frac{\partial^2 G_k(\mathbf{p};\mathbf{q})}{\partial n_p \partial n_q}dS_q(\mathbf{q}), \tag{10}$$

and μ is a constant chosen to give good conditioning [generally $\mu\sim i/(1+k)$]. Equation (6) can be solved numerically using the boundary element method. This method approximates the surface S with a set of disjoint elements S_j, $j=1,2,\ldots,n$, called the mesh, such that the elements form a closed surface approximation to the surface S, and approximates φ and its derivative as

$$\varphi(\mathbf{q}) = p_j, \quad \frac{\partial\varphi(\mathbf{q})}{\partial n_q} = v_j, \quad \mathbf{q}\in S_j, \tag{11}$$

where the values of p_j and v_j are constant within each element. More complicated formulations for the dependence of $\varphi(\mathbf{q})$ and $\partial\varphi(\mathbf{q})/\partial n_q$ on position within the element S_j are possible, but constant values within each element have been chosen in order to simplify the calculations as much as possible. The elements are designed such that the points at which measurements were made, the points $\mathbf{p}_i$, lie at the center of the elements. Either p_i, v_i, or some linear combination of the two is known (from measurement, assumption, or other expert knowledge) for each value of i. Then Eq. (6) can be written as

$$\sum_{j=1}^{n} p_j\left(-\frac{1}{2}I + M_{k,j} + \mu N_{k,j}\right)(\mathbf{p}_i) = \sum_{j=1}^{n} v_j\left(L_{k,j} + \mu\left(\frac{1}{2}I + M_{k,j}^T\right)\right)(\mathbf{p}_i), \quad \mathbf{p}_i\in S_i,\ i=1,2,\ldots,n, \tag{12}$$

where

$$(L_{k,j})(\mathbf{p}) = \int_{S_j} G_k(\mathbf{p};\mathbf{q})dS_q(\mathbf{q}), \tag{13}$$

$$(M_{k,j})(\mathbf{p}) = \int_{S_j} \frac{\partial G_k(\mathbf{p};\mathbf{q})}{\partial n_q} dS_q(\mathbf{q}), \tag{14}$$

$$(M^T_{k,j})(\mathbf{p}) = \int_{S_j} \frac{\partial G_k(\mathbf{p};\mathbf{q})}{\partial n_p} dS_q(\mathbf{q}), \tag{15}$$

$$(N_{k,j})(\mathbf{p}) = \int_{S_j} \frac{\partial^2 G_k(\mathbf{p};\mathbf{q})}{\partial n_p \partial n_q} dS_q(\mathbf{q}). \tag{16}$$

This formulation of the problem leads to an $n \times n$ matrix of equations of the form $A_{ij}v_j = B_{ij}p_j$ $(i=1,2,\ldots,n,\ j=1,2,\ldots,n)$, where the p_j are known from measurement, and the A_{ij} and B_{ij} are calculated from $L_{k,j}$, $M_{k,j}$, $M^T_{k,j}$, and $N_{k,j}$. Unlike the finite element formulation of the problem, the matrices A and B are not sparse, so for a general surface the algorithm used for the solution of the problem cannot take advantage of a special matrix structure to speed-up calculation.

The boundary element methodology can be applied to the external problem to obtain

$$\sum_{j=1}^{n} p_j(M_{k,j})(\mathbf{p}) - v_j(L_{k,j})(\mathbf{p}) = \varphi(\mathbf{p}), \quad \mathbf{p} \in E, \tag{17}$$

which can be used to calculate acoustic pressures in the far-field.

In general it is necessary to choose the S_j such that the maximum dimension of each element is less than $1/6$ of the wavelength of the acoustic wave. This choice ensures that the pressure distribution is described adequately by the boundary conditions. For a fully three-dimensional problem and a high frequency, this requirement leads to very large values of n and hence very large matrices for the numerical computation. Problems of this type rapidly become very expensive in terms of computation time and memory storage, making calculations on a desktop PC impractical. An alternative formulation is described in Sec. IV that avoids this issue by rewriting equation Eq. (6) as a set of independent two-dimensional problems.

Another difficulty with the numerical calculation is that if $\mathbf{p}=\mathbf{q}$, then the integrals used in the calculation of $L_{k,j}$ and $N_{k,j}$ become singular, and so numerical calculations become problematic as $\mathbf{p} \to \mathbf{q}$. This is true of the reformulation described in Sec. IV as well as the form in Eq. (12). Section V describes the approach taken to avoid this singularity and make the calculation possible.

IV. REFORMULATION

As has been shown by other authors,[2,12,17–21] significant savings in computation time and complexity can be achieved by reducing the full three-dimensional problem to an axisymmetric two-dimensional problem. Since the measurement data used in this work are gathered on a cylinder, such an approach is possible here. The reduction is achieved by writing the measurement data as a Fourier series in terms of the azimuthal cylindrical polar coordinate,

$$p(\rho,\theta,z) = \sum_{m=0}^{\infty} (a_m(\rho,z)\sin(m\theta) + b_m(\rho,z)\cos(m\theta)), \tag{18}$$

where the cylindrical polar coordinates of a point are (ρ,θ,z).

It can be shown[2,12,17–21] that if c_m is one of the $a_m(\rho,z)$ or $b_m(\rho,z)$, then c_m satisfies

$$\int_{\zeta} c_m(\mathbf{q}) \frac{\partial G^m_k(\mathbf{p};\mathbf{q})}{\partial n_q} - G^m_k(\mathbf{p};\mathbf{q}) \frac{\partial c_m(\mathbf{q})}{\partial n_q} d\zeta(\mathbf{q}) = \varepsilon(\mathbf{p}) c_m(\mathbf{p}), \tag{19}$$

where the axisymmetric surface S is generated by rotating the curve ζ about the z axis, and

$$G^m_k(\rho_p,z_p;\rho_q,z_q) = \int_{-\pi}^{\pi} \cos(m\alpha) G_k(\rho_p,0,z_p;\rho_q,\alpha,z_q) d\alpha. \tag{20}$$

In practice the sum of Eq. (18) is truncated after N terms. N is chosen such that the Fourier series is a sufficiently good representation of the input data and is ultimately limited by the number of discrete azimuthal points at which the measurements were made (although in general N can be significantly smaller than this limit and still be a good approximation to the input data). The similarity of Eq. (19) to a two-dimensional form of Eq. (1) means that the boundary element method with the Burton–Miller approach can be adopted. Each boundary element is now an axisymmetric surface, flat in the (ρ,z) plane, rather than a general flat surface in three-dimensional space, so that the element now has a center line $\mathbf{p}_j=(\rho_j, 0 \leqslant \theta \leqslant 2\pi, z_j)$ rather than a unique center point. Writing

$$p(\rho_j,\theta,z_j) = \sum_{m=0}^{\infty} (a_{m,j}\sin(m\theta) + b_{m,j}\cos(m\theta)), \tag{21}$$

$$v(\rho_j,\theta,z_j) = \sum_{m=0}^{\infty} (d_{m,j}\sin(m\theta) + e_{m,j}\cos(m\theta)) \tag{22}$$

and letting

$$c_{m,j} = \begin{cases} a_{m,j}, & m = 1,\ldots,N-1 \\ b_{m-N,j}, & m = N,\ldots,2N-1, \end{cases} \tag{23}$$

$$f_{m,j} = \begin{cases} d_{m,j}, & m = 1,\ldots,N-1 \\ e_{m-N,j}, & m = N,\ldots,2N-1, \end{cases} \tag{24}$$

for all values of j, the equivalents to Eqs. (12)–(16) are

$$\sum_{j=1}^{n} c_{m,j}\left(-\frac{1}{2}I + M^m_{k,j} + \mu N^m_{k,j}\right)(\mathbf{p}_i) = \sum_{j=1}^{n} f_{m,j}\left(L^m_{k,j} + \mu\left(\frac{1}{2}I + M^{mT}_{k,j}\right)\right)(\mathbf{p}_i), \quad \mathbf{p}_i \in S_i,$$
$$i = 1,2,\ldots,n, \quad m = 1,2,\ldots,2N-1, \tag{25}$$

$$(L^m_{k,j})(\mathbf{p}) = \int_{S_j} G^m_k(\mathbf{p};\mathbf{q}) dS_q(\mathbf{q}), \tag{26}$$

$$(M^m_{k,j})(\mathbf{p}) = \int_{S_j} \frac{\partial G^m_k(\mathbf{p};\mathbf{q})}{\partial n_q} dS_q(\mathbf{q}), \tag{27}$$

$$(M^{mT}_{k,j})(\mathbf{p}) = \int_{S_j} \frac{\partial G^m_k(\mathbf{p};\mathbf{q})}{\partial n_p} dS_q(\mathbf{q}), \tag{28}$$

$$(N^m_{k,j})(\mathbf{p}) = \int_{S_j} \frac{\partial^2 G^m_k(\mathbf{p};\mathbf{q})}{\partial n_p \partial n_q} dS_q(\mathbf{q}). \tag{29}$$

A similarly formulated version of Eq. (17) can be used to calculate each of the components of the Fourier series in the external domain, and the pressure value can be obtained by summing as in Eq. (18). Note that the equations above are a set of $2N-1$ independent two-dimensional equations, and so the numerical solution of these equations will require the solution of $2N-1$ matrix problems of the type outlined in Sec. III. Since each problem is two dimensional, the number of elements needed is smaller, and hence the memory and storage problems mentioned in Sec. III are less likely to arise.

The remarks made in Sec. III regarding the singularity of the integrals $L_{k,j}$ and $N_{k,j}$ are also true of the integrals $L^m_{k,j}$ and $N^m_{k,j}$. The numerical evaluation of these integrals will be addressed in Sec. V.

A. Distributed computation

As the sophistication of mathematical models increases, so does their computational complexity. This increase in complexity leads to a requirement for increased computing power. In recent years the concept of distributed (or parallel) computing, where multiple processors carry out different parts of a computational task, has become increasingly popular as a way of getting the necessary computing power. Since many organizations have a large number of desktop PCs whose processors are not used to their maximum capacity at all times (for instance, most PCs are not used overnight), interest has grown in exploiting this spare processor capacity for distributed computing applications.

The speed at which a distributed computation can be carried out depends on two main factors: the speed of the individual processors and the amount of intercommunication required between the processors. Intercommunication slows the computation down since transferring large amounts of data takes time, and sometimes a fast computer will have to wait for the results from a slower one. Intercommunication can cause usability problems if the network used is not solely dedicated to the communication. Many common modeling tasks, in particular the solution of very large sets of linear equations, require intercommunication between different processors. Another class of tasks, called "ludicrously parallel" processes, does not require any intercommunication: each part of the computation is independent of the other parts. Common examples of applications that can be implemented as ludicrously parallel processes include Monte Carlo simulation, image analysis, database and text searching, and data mining.

The formulation of the problem given in Eq. (25) is a ludicrously parallel formulation because each of the components in Eq. (18) is propagated independently of the others, and so the calculations using each component can be carried out in parallel. This parallel computation can result in a reduction in computation time by a factor of $2N-1$.

A distributed computing system has been installed at NPL that uses the spare capacity of 200 desktop PCs to carry out computationally expensive calculations.[22] In order to avoid network problems for other users, attention has largely been restricted to ludicrously parallel processes. The system has been successfully used to solve problems involving electron transport in quantum dots, adiabatic energy transfer in cesium atoms, sensitivity analysis of finite element models, and Monte Carlo calculations for ionizing radiation applications.

The software implementation of the method outlined in this paper has been designed to run on the NPL distributed computing system, and the results presented in Sec. VI were all generated on the system. The most extensive calculation, using $N=179$, took 7 h 20 min to carry out 359 calculations, each of which would normally take around 2.75 h to complete. This reduced the overall computational time from nearly 6 weeks by a factor of about 140, making calculations with measurement data tractable.

V. NUMERICAL INTEGRATION

This section addresses the numerical evaluation of the integrals $L^m_{k,j}$ and $N^m_{k,j}$ as $\mathbf{p}\to\mathbf{q}$ for two-dimensional axisymmetric elements. It adopts the approach outlined by Kirkup[3] for axisymmetric problems but adapts his methodology to account for the new form of the Green's function G^m_k.

From Eqs. (2) and (3), G_k is a function of $\cos(\theta_p-\theta_q)$, and since the normal to an axisymmetric surface has no θ component, the derivatives of G_k with respect to n_p and n_q are also functions of $\cos(\theta_p-\theta_q)$. In the evaluation of G^m_k, the limits of integration with respect to θ_q are $-\pi$ and π. Hence for a fixed value of θ_p, a periodic function of $\theta_p-\theta_q$ is integrated over one full cycle, and so the value of θ_p is arbitrary and can be set to zero [as has been done in Eq. (20)]. As shown in Ref. 3, the quantities

$$G_k(\mathbf{p};\mathbf{q}) - G_0(\mathbf{p};\mathbf{q}) \tag{30}$$

and

$$\frac{\partial^2 G_k}{\partial n_p \partial n_q}(\mathbf{p};\mathbf{q}) - \frac{\partial^2 G_0}{\partial n_p \partial n_q}(\mathbf{p};\mathbf{q}) - \frac{k^2}{2} G_0(\mathbf{p};\mathbf{q}) \tag{31}$$

are bounded as $\mathbf{p}\to\mathbf{q}$. Hence, from Eqs. (26) and (29), the integrals can be written as

$$L^m_{k,j}(\mathbf{p}) = L^m_{0,j}(\mathbf{p}) + \int_{S_j} T_1(\mathbf{p};\mathbf{q}) dS_j(\mathbf{q}), \tag{32}$$

where

$$T_1 = (G_k(\mathbf{p};\mathbf{q}) - G_0(\mathbf{p};\mathbf{q}))\cos(m\theta_q), \tag{33}$$

and

$$N^m_{k,j}(\mathbf{p}) = N^m_{0,j}(\mathbf{p}) + \frac{k^2}{2}L_{0,m}(\mathbf{p}) + \int_{S_j} T_2(\mathbf{p};\mathbf{q})dS_j(\mathbf{q}), \tag{34}$$

where

$$T_2 = \cos(m\theta_q)\left(\frac{\partial^2 G_k}{\partial n_p \partial n_q}(\mathbf{p};\mathbf{q}) - \frac{\partial^2 G_0}{\partial n_p \partial n_q}(\mathbf{p};\mathbf{q}) - \frac{k^2}{2}G_0(\mathbf{p};\mathbf{q})\right), \tag{35}$$

and the terms T_1 and T_2 are regular as $\mathbf{p}\to\mathbf{q}$ and can be evaluated using standard surface quadrature techniques. Hence the remaining challenge is to evaluate $L^m_{0,j}$ and $N^m_{0,j}$.

$L^m_{0,j} \sim r^{-1}$ as $\mathbf{p}\to\mathbf{q}$. This is a sufficiently mild singularity that the integral can be evaluated using standard quadrature techniques, provided the situation $\mathbf{p}=\mathbf{q}$ is avoided. The implementation at NPL splits the element S_j along the line $\rho=\rho_p$, $0\leqslant\theta\leqslant 2\pi$, $z=z_p$ and evaluates the two halves of the integral separately using quadrature rules that avoid the dividing line.

$N^m_{0,j} \sim r^{-3}$ as $\mathbf{p}\to\mathbf{q}$, which is too severe a singularity for regular quadrature. Instead, the surface integral is reformulated as the sum of two regular surface integrals and a regular volume integral.[3] Consider a closed surface S_{jc} that encloses a volume V_{jc},

$$\int_{S_{jc}} \frac{\partial G_0}{\partial n_q}(\mathbf{p};\mathbf{q})\cos(m\theta_q)dS_{jc}(\mathbf{q}) = \int_{S_{jc}} (\cos(m\theta_q)\,\nabla\, G_0(\mathbf{p};\mathbf{q}))\cdot\mathbf{n}_q dS_{jc}(\mathbf{q}) \tag{36}$$

$$= \int_{V_{jc}} \nabla\cdot(\cos(m\theta_q)\,\nabla\, G_0(\mathbf{p};\mathbf{q}))dV_{jc}(\mathbf{q}), \tag{37}$$

by the divergence theorem. Furthermore,

$$\int_{V_{jc}} \nabla\cdot(\cos(m\theta_q)\,\nabla\, G_0(\mathbf{p};\mathbf{q}))dV_{jc}(\mathbf{q}) = \int_{V_{jc}} \nabla(\cos(m\theta_q))\cdot\nabla G_0(\mathbf{p};\mathbf{q})dV_{jc}(\mathbf{q}) + \int_{V_{jc}} (\cos(m\theta_q)\nabla^2 G_0(\mathbf{p};\mathbf{q}))dV_{jc}(\mathbf{q}). \tag{38}$$

G_0 is defined as the solution to

$$\nabla^2 G_0(\mathbf{p};\mathbf{q}) = -\delta(\|\mathbf{p}-\mathbf{q}\|), \tag{39}$$

and so

$$\int_{V_{jc}} (\cos(m\theta_q)\nabla^2 G_0(\mathbf{p};\mathbf{q}))dV_{jc}(\mathbf{q}) = -\cos(m\theta_p) = -1 \tag{40}$$

because θ_p has been set to 0. Considering the first term on the right-hand side of Eq. (38),

$$\nabla(\cos(m\theta_q)) = \left\{0, \frac{m}{\rho_q}\sin(m\theta_q), 0\right\}, \tag{41}$$

$$\nabla G_0(\mathbf{p};\mathbf{q}) = \frac{-1}{4\pi r^2}\left\{\frac{\partial r}{\partial \rho_q}, \frac{1}{\rho_q}\frac{\partial r}{\partial \theta_q}, \frac{\partial r}{\partial z_q}\right\}, \tag{42}$$

and so

$$\nabla(\cos(m\theta_q))\cdot\nabla G_0(\mathbf{p};\mathbf{q}) = \frac{-m\sin(m\theta_q)}{4\pi r^2\rho_q^2}\frac{\partial r}{\partial \theta_q}. \tag{43}$$

Now consider the derivative of Eq. (38) with respect to n_p. Substituting Eqs. (40) and (43) into Eq. (38) gives

$$\frac{\partial}{\partial n_p}\int_{S_{jc}} \frac{\partial G_0}{\partial n_q}(\mathbf{p};\mathbf{q})\cos(m\theta_q)dS_{jc}(\mathbf{q}) = \frac{\partial}{\partial n_p}\left(-1 + \int_{V_{jc}} \frac{-m\sin(m\theta_q)}{4\pi r^2\rho_q^2}\frac{\partial r}{\partial \theta_q}dV_{jc}\right) = \frac{\partial}{\partial n_p}\int_{V_{jc}} \frac{-m\sin(m\theta_q)}{4\pi r^2\rho_q^2}\frac{\partial r}{\partial \theta_q}dV_{jc}. \tag{44}$$

This integrand is regular strictly inside the volume V_{jc} for $\rho_q>0$ and is sufficiently well behaved as $\mathbf{p}\to\mathbf{q}$ that it can be evaluated for points close to (but not lying on) the bounding surface S_{jc}. Since $\partial r/\partial\theta_q \sim \rho_q/r$, the singularity at $\rho_q=0$ is sufficiently weak that it can be avoided with a suitable choice of quadrature rule. If S_{jc} is made up of the boundary element S_j, on which $N^m_{0,j}$ is singular, and S_{j1} and S_{j2}, which are two surfaces chosen to close the volume V_{jc} on which the integral Eq. (38) is regular, then

$$N^m_{0,j} = \frac{\partial}{\partial n_p}\int_{S_j} \frac{\partial G_0}{\partial n_q}(\mathbf{p};\mathbf{q})\cos(m\theta_q)dS_j(\mathbf{q}) = \frac{\partial}{\partial n_p}\int_{V_{jc}} \frac{-m\sin(m\theta_q)}{4\pi r^2\rho_q^2}\frac{\partial r}{\partial \theta_q}dV_{jc} - \frac{\partial}{\partial n_p}\int_{S_{j1}} \frac{\partial G_0}{\partial n_q}(\mathbf{p};\mathbf{q})\cos(m\theta_q)dS_{j1}(\mathbf{q}) - \frac{\partial}{\partial n_p}\int_{S_{j2}} \frac{\partial G_0}{\partial n_q}(\mathbf{p};\mathbf{q})\cos(m\theta_q)dS_{j2}(\mathbf{q}). \tag{45}$$

The implementation of this calculation at NPL uses constant collocation elements, so that all elements are flat. S_{j1} and S_{j2} are chosen to be right-angled cones that close the volume partially defined by the element since the surface integrals are easily evaluated on such surfaces and because cones of this type create a closed volume with any flat element, no matter how it is oriented. An example of this closure is shown in Fig. 3.

All the terms on the right-hand side of Eq. (45) can be calculated using standard quadrature methods. In our implementation, all linear integrals used the standard Gaussian eight-point rule, and all integrals over a triangle used a seven-point Gaussian quadrature rule over a standard triangle.[23] This expression makes calculations using Eq. (25) numerically viable, and, as was stated previously, the paral-

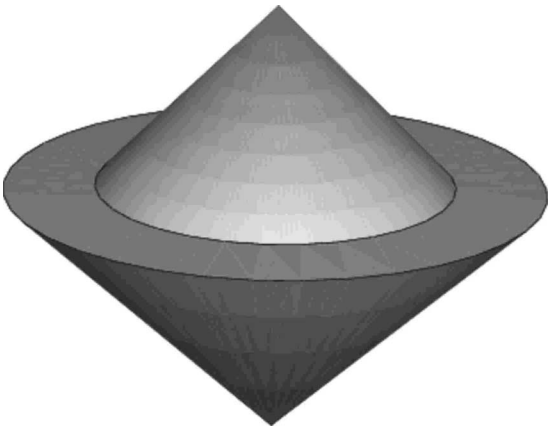


FIG. 3. Sketch showing a typical boundary element (flat annulus) and the two right-angled cones used to generate a closed volume in conjunction with the element in order to evaluate the singular integrals.

lelization of the calculation makes high-frequency calculations with a large number of elements possible within a reasonable amount of time.

VI. RESULTS

This section presents some of the results that have been obtained using the software. For all data sets, the maximum possible number of Fourier coefficients has been used based on the number of circumferential measurement points since earlier computational problems led to attempts to minimize all possible sources of error.

As has been described, the measurement data were gathered on open-ended cylindrical surfaces. The boundary element method requires a closed surface for calculations. All the results presented here were generated using disks to close the measurement surface, so that the closed surface was a cylinder in all cases. The extra disks were populated with data by decreasing the Fourier coefficient on the perimeter to zero at the center of the disk linearly. Trials run using problems with known solutions showed that this method gave good agreement.

Alternatives to this closure method were investigated, including hemispherical closures and conical closures. The investigations suggested that, provided the closure surfaces were populated with data in a way that reflected the likely behavior of the transducer, the closure surfaces had little effect on accuracy. Flat closures were chosen because they are the closing surfaces with the smallest area and hence require fewest elements to mesh them. Use of flat closure surfaces minimizes the overall size of the matrix. Various methods of populating the closure surfaces were tested. Besides tests using a linear decrease in amplitude, tests were run where all pressure values on the closure surfaces were assumed to be zero, and some tests were run using a Tukey window to gradually decrease the pressure to zero on the line $r=0$. The tests showed that the linear decrease gave the best results both for model problems with known solutions and for measurement data.

As an initial simple check, the software was tested by simulating monopole and dipole sources at a range of frequencies between 1 and 10 kHz. Agreement between the numerical solution and the analytic solution was excellent for all tests. Further details of the results can be found in Ref. 4. Once the software had passed the initial tests, it was applied to measurement data.

A. Experimental configuration

The measurement data were generated from a pair of free-flooding hollow cylindrical piezoelectric projectors. The transducers used were a pair of EDO Western Corporation model 249 free-flooding ring transducers. Each transducer was 98 mm in outer diameter with an 18 mm wall thickness and was 92 mm long. The transducers were mounted coaxially, giving an overall length of 185 mm, and were encapsulated in a polyurethane polymer. All dimensions are of the outer surface including the coating, rather than the dimensions of the active elements. The cylinders had resonance frequencies at both 13.9 and 27.5 kHz (where the transmitting voltage response reached a local maximum).

The transducers were mounted in two different configurations to provide two different sources and were driven together, in parallel, with a voltage of approximately 10 V. Each configuration was operated at the two resonance frequencies of the cylindrical projectors. For both configurations and both frequencies, cylindrical measurement scans were made over tall cylinders in the near-field for use as input data, and at various radii and depths for validation data.

The first configuration mounted the transducers coaxially. With both transducers driven in phase, this provided some directionality in the vertical orientation with a maximum output in the far-field produced in an orthogonal plane passing through the midpoint of the source (the plane bisecting the two transducers). In general this configuration required few Fourier components in the sum of Eq. (21) and hence was not computationally challenging. Results for this configuration are shown in Ref. 4, but in all cases the measurements and calculated values showed good agreement, and excellent agreement in the far-field.

To provide some directionality in the azimuthal direction, a reflector was positioned close to the pair of coaxially mounted transducers to create sufficient interference between the reflected and directly radiated fields to produce lobes in the azimuth direction in the far-field response. The reflector was made from a thin steel plate coated on the inner surface with a 3 mm thick layer of closed-cell neoprene rubber. The air-filled cavities within the neoprene ensured that a strong reflection was obtained, producing significant interference and well-defined lobes. Typical scans obtained from this configuration are shown in Fig. 2. Use of the steel plate alone did not provide a strong reflection for sound at 13.9 kHz (where the wavelength is 53 mm). Only results created using the second configuration are shown here because they are computationally more challenging and hence of more interest. Note that the measured values and calculated results using this configuration are not the same as those in Ref. 4 since concern about the integrity of the original data and results led to measurement and calculations being repeated. Measurements were made at frequencies of 13.9 and 27.5 kHz, and at each frequency cylindrical scans with radii

TABLE II. Details of the frequencies, wavelengths, and boundary element meshes used to calculate results and of the computational time and elapsed time required to generated each set of results.

Frequency (kHz)	13.9	13.9	27.5	27.5
Wavelength (mm)	106	106	53.7	53.7
Elements per wavelength	6	12	6	12
Element length (mm)	17.7	8.9	9.0	4.5
Value of N used	89	89	179	179
Total number of elements	139	189	275	374
Approximate total computational time (h)	42	74	561	1002
Approximate elapsed time for calculation (h)	2	3	6	7
Average computational time per component (min)	14	24	93	167

of 0.338 and 3.5 m were made, with the scan depth varying between −0.49 and 0.49 m in all cases. The transducers were driven with long tone bursts, and the amplitude and phase of the received signal were extracted from the steady state part of the signal under effectively free-field conditions.

It is thought that a slight change in the rotational alignment occurred between the near and far-field measurements at each frequency. The near-field and far-field scans were gathered at different times, and the experimental setup was dismantled between these times as the measurement tank was required for other tasks. When the equipment was reassembled, the only method available for the alignment of the cylindrical array was manual adjustment based on visual alignment. The uncertainty associated with this alignment method during hydrophone calibration is usually ±3°. The hydrophone calibration procedure uses a marked pole and custom-built mounts, whereas the array used in this work has used an unmarked pole and a less snugly fitting mount, so it is possible that the alignment procedure is more susceptible to error in this case. The misalignment has not been corrected for in the plots in this paper as there was no independent way of determining the appropriate correction.

B. Typical computed results

As was stated above, the transducers were driven at frequencies of 13.9 and 27.5 kHz with corresponding wavelengths of 106 and 54 mm, respectively. The mesh sizes, number of frequencies, and number of elements are shown in Table II. Two different mesh sizes were used for each frequency. The mesh sizes correspond to $\lambda/12$ and $\lambda/6$, enabling the results to be checked for convergence. A generally applied rule of thumb (see, for example, Foote and Francis[24]) suggests that $\lambda/6$ should give reasonably accurate results. The results with $\lambda/6$ were virtually indistinguishable from those with $\lambda/12$, and so $\lambda/6$ will be used in any subsequent work.

Table II also shows the elapsed time and computational time taken to obtain results for each of the jobs. The total computational time for the 27.5 kHz data was approximately 23 days for the $\lambda/6$ mesh and 42 days (6 weeks) for the $\lambda/12$ mesh, but the corresponding elapsed times were approximately 6 and 7 h, respectively. This shows the importance of distributed computing to this work. The speed-up factors of the lower frequency jobs are less impressive because the two jobs were started simultaneously and thus had to compete for computational resources. Despite this competition, the elapsed time of both jobs decreased by a factor of more than 17.

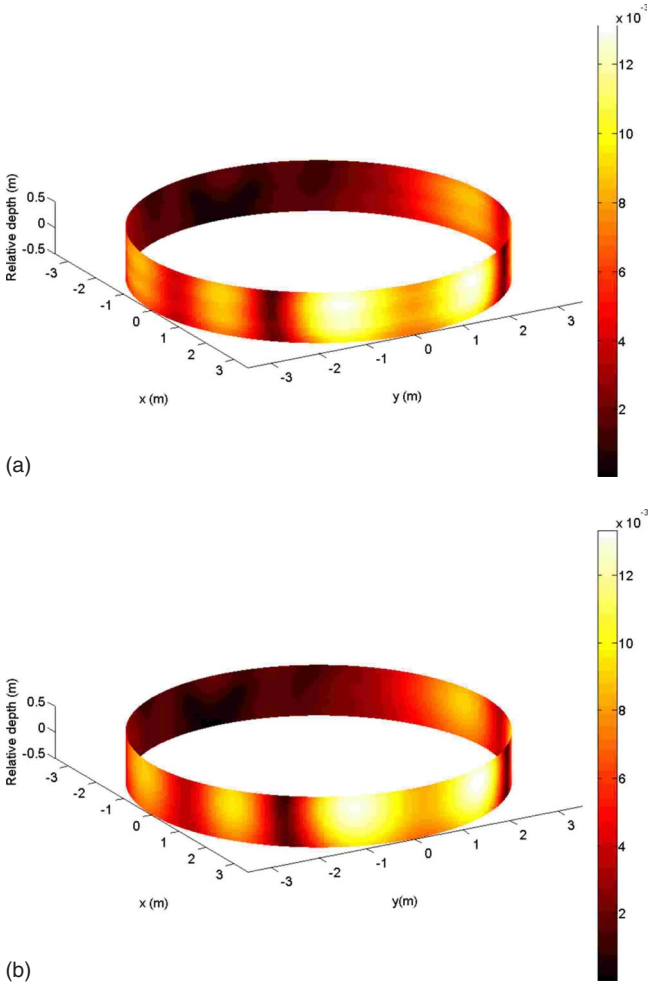


FIG. 4. (Color online) Comparison of (a) measured and (b) calculated acoustic pressure amplitudes at 13.9 kHz. Cylinder has a radius of 3.5 m and depths between −0.49 and 0.49 m, where the transducer is at depth 0 m.

In Figs. 4–9, the results plotted represent a comparison of the absolute amplitudes of the acoustic fields. The ordinate axes in Figs. 5, 6, 8, and 9 are labeled scaled pressure amplitude since the data have not been scaled for the sensitivity of the hydrophone and preamplifier, the data shown being equivalent to the received voltage amplitudes.

Figure 4 shows calculated and measured far-field acoustic pressure amplitudes at 13.9 kHz. The radius of the cylinders is 3.5 m, the depths ranged from −0.49 to 0.49 m in steps of 0.01 m, and data were gathered every 2° in the angular direction. Figures 5 and 6 show a comparison between two sets of measured and calculated pressure amplitude scan lines. Calculated results using both mesh densities are shown, but the lines overlay one another in both cases, showing that convergent results have been obtained. Lines at

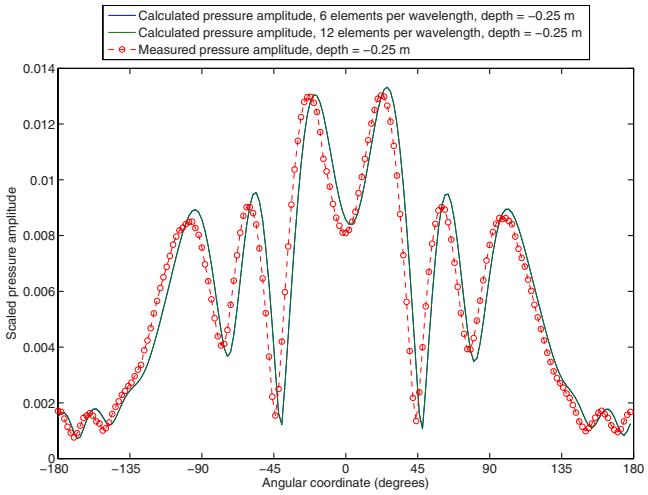


FIG. 5. (Color online) Line plot of measured and calculated acoustic pressure amplitudes vs azimuthal coordinate at a frequency of 13.9 kHz. Results are shown for a radius of 3.5 m and a depth of −0.25 m relative to the transducer's midpoint.

depths of −0.25 and 0.25 m were chosen, but agreement between measurement and calculation is about the same at all depths.

These figures indicate that, in general, the agreement between the measured and calculated pressure amplitudes is good at this frequency. The main and side lobes are in the correct positions, and the sizes of the lobes agree well, although there is a slight angular misalignment between measured and calculated values. The probable cause of this misalignment was described in the previous section. Visual assessment of the results suggests that the misalignment is approximately 5°.

Calculated and measured far-field acoustic pressure amplitudes at 27.5 kHz, under identical conditions of measurement, are shown similarly in Figs. 7–9. Again, good agreement is achieved, with the positions and sizes of the main and side lobes closely matched. As at 13.9 kHz, there is an angular shift of approximately 5° between predicted values and measurements. The shifts between measured and calculated results at the two frequencies are in good agreement, suggesting that the shifts are caused by a systematic error, such as that caused by the misalignment between near-field and far-field measurements as described in the previous section.

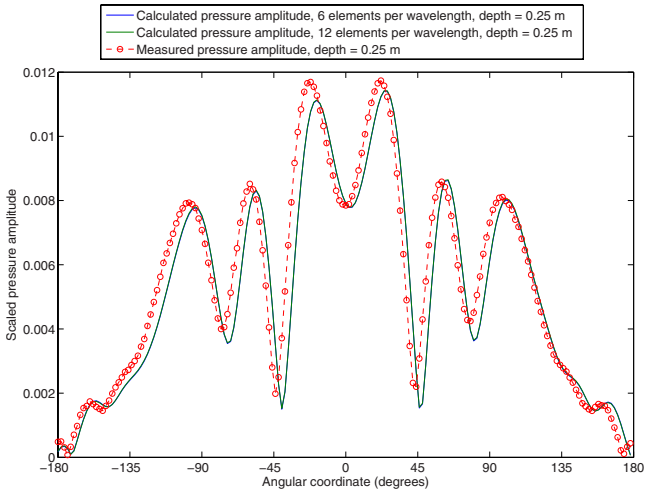


FIG. 6. (Color online) Same as in Fig. 5, but at a depth of 0.25 m relative to the transducer's midpoint.

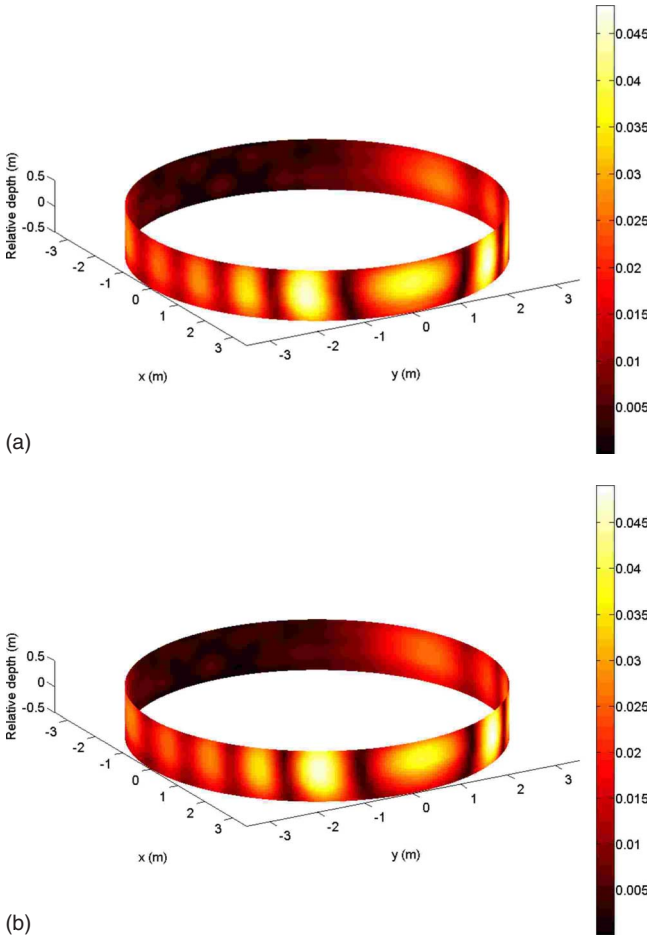


FIG. 7. (Color online) Comparison of (a) measured and (b) calculated acoustic pressure amplitudes at 27.5 kHz. Cylinder has a radius of 3.5 m and depths between −0.49 and 0.49 m, where the transducer is at depth 0 m.

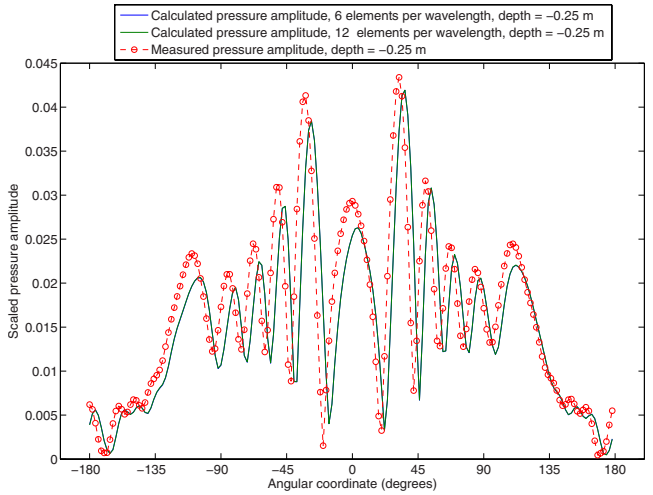


FIG. 8. (Color online) Line plot of measured and calculated acoustic pressure amplitudes versus azimuthal coordinate at a frequency of 27.5 kHz. Results are shown for a radius of 3.5 m and a depth of −0.25 m relative to the transducer's midpoint.

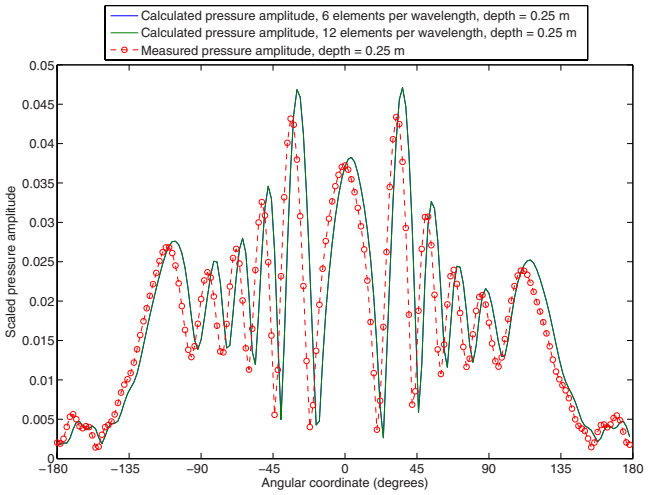


FIG. 9. (Color online) Same as in Fig. 8, but at a depth of 0.25 m relative to the transducer's midpoint.

VII. CONCLUSIONS

The Fourier series formulation of the boundary integral form of the Helmholtz equation on an axisymmetric surface has been used to calculate the far-field response of underwater acoustic transducers from the measurement of their behavior in the near-field. The use of a Fourier series formulation has meant that the calculation could be run on a distributed computing system, thus significantly reducing the run time and rendering the problem computationally tractable. The reformulation has ensured numerical stability of the calculation for all wavenumbers and has used volume integrals to avoid the need for numerical evaluation of singular surface integrals. The results of calculations have been validated against measured data, and the agreement has been good at frequencies of 13.9 and 27.5 kHz. As distributed computation becomes more widespread, it is hoped that similar methodologies will be applied to other problems in computational acoustics.

ACKNOWLEDGMENTS

The authors would like to acknowledge the support of the National Measurement System Policy Unit of the UK Department of Innovation, Universities and Skills, which funded this work as part of the Acoustical Metrology and Software Support for Metrology programs. The authors would also like to express their thanks to Dr. Peter M. Harris of NPL for his initial work on the formulation of the problem and to Mr. Gary Hayman of NPL for his assistance with the measurements on the transducers. Finally, the authors would like to thank the reviewers for their detailed and extremely helpful comments that led to the revised version of this paper and an improved version of the software used to generate the results.

[1]A. J. Burton and G. F. Miller, "The application of integral equation methods to the numerical solution of some exterior boundary value problems," Proc. R. Soc. London, Ser. A **323**, 201–210 (1971).
[2]P. C. Macey, "Oblique incidence diffraction by axisymmetric structures," Proceedings of the Institute of Acoustics **16**, 67–74 (1994).
[3]S. Kirkup, *The Boundary Element Method in Acoustics* (Integrated Sound Software, Hebden Bridge, UK, 1998).
[4]L. Wright, S. Robinson, V. Humphrey, P. Harris, and G. Hayman, "The application of boundary element methods to near-field acoustic measurements on cylindrical surfaces at NPL," Technical Report No. DQL-AC 011, NPL, Teddington, U.K., 2005.
[5]R. J. Urick, *Principles of Underwater Sound* (McGraw-Hill, New York, 1983).
[6]R. Bobber, *Underwater Electroacoustic Measurements* (Peninsula, Los Altos, 1988).
[7]B. B. Baker and E. T. Copson, *The Mathematical Theory of Huygens' Principle* (Clarendon, Oxford, UK, 1950).
[8]D. D. Baker, "Determination of far-field characteristics of large underwater sound transducers from near-field measurements," J. Acoust. Soc. Am. **34**, 1737–1744 (1962).
[9]E. G. Williams, *Fourier Acoustics: Sound Radiation and Nearfield Holography* (Academic, New York, 1999).
[10]A. Van Buren, "Near-field transmitting and receiving properties of planar near-field calibration arrays," J. Acoust. Soc. Am. **89**, 1423–1427 (1991).
[11]W. J. Trott, "Underwater sound transducer calibration from near-field data," J. Acoust. Soc. Am. **53**, 192–199 (1964).
[12]W. L. Meyer, W. A. Bell, M. P. Stallybrass, and B. T. Zinn, "Prediction of the sound field radiated from axisymmetric surfaces," J. Acoust. Soc. Am. **65**, 631–638 (1979).
[13]D. T. I. Francis, "A gradient formulation of the Helmholtz equation for acoustic radiation and scattering," J. Acoust. Soc. Am. **93**, 1700–1709 (1993).
[14]H. A. Schenck, "Improved integral formulation for acoustic radiation problems," J. Acoust. Soc. Am. **44**, 41–58 (1968).
[15]L. G. Copley, "Fundamental results concerning integral representations in acoustic radiation," J. Acoust. Soc. Am. **44**, 28–32 (1968).
[16]S. Forsythe, "A Matlab version of CHIEF (Combined Helmholtz Integral Equation Formulation) for solving acoustic radiation and scattering problems," J. Acoust. Soc. Am. **106**, 2193 (1999).
[17]A. F. Seybert, B. Soenarko, F. J. Rizzo, and D. J. Shippy, "A special integral equation formulation for acoustic radiation and scattering for axisymmetric bodies and boundary conditions," J. Acoust. Soc. Am. **80**, 1241–1247 (1986).
[18]B. Soenarko, "A boundary element formulation for radiation of acoustic waves from axisymmetric bodies with arbitrary boundary conditions," J. Acoust. Soc. Am. **93**, 631–639 (1993).
[19]W. Wang, N. Atalla, and J. Nicolas, "A boundary integral approach for acoustic radiation of axisymmetric bodies with arbitrary boundary conditions valid for all wave numbers," J. Acoust. Soc. Am. **101**, 1468–1478 (1997).
[20]S. V. Tsinopoulos, J. P. Agnantiaris, and D. Polyzos, "An advanced boundary element/fast Fourier transform axisymmetric formulation for acoustic radiation and wave scattering problems," J. Acoust. Soc. Am. **105**, 1517–1526 (1999).
[21]P. Juhl, "An axisymmetric integral equation formulation for free space nonaxisymmetric radiation and scattering of a known incident wave," J. Sound Vib. **163**, 397–406 (1993).
[22]T. J. Esward, N. J. McCormick, K. M. Lawrence, and M. J. Stevens, "Software support for metrology good practice guide No. 17: Distributed computing for metrology applications," Technical Report No. DEM-ES 006, NPL, Teddington, U.K., 2006.
[23]M. E. Laursen and M. Gellert, "Some criteria for numerically integrated matrices and quadrature formulas for triangles," Int. J. Numer. Methods Eng. **12**, 67–76 (1978).
[24]K. G. Foote and D. T. I. Francis, "Comparing Kirchoff-approximation and boundary element models for computing gadoid target strengths," J. Acoust. Soc. Am. **111**, 1644–1654 (2002).

Use of acoustic navigation signals for simultaneous localization and sound-speed estimation

E. K. Skarsoulis and G. S. Piperakis
Institute of Applied and Computational Mathematics, Foundation for Research and Technology Hellas, P.O. Box 1385, 711 10 Heraklion, Crete, Greece

(Received 24 October 2008; accepted 8 January 2009)

The possibility of exploiting low-resolution acoustic signals used for the navigation of Lagrangian floats to simultaneously estimate the speed of sound is studied. Acoustic navigation of Lagrangian floats is regularly carried out by measuring travel times from three fixed stations assuming a known value for the speed of sound. The sound speed is considered here as a variable of the problem to be estimated from the travel-time data simultaneously with the horizontal location of the float. The estimation problem is linearized and solved analytically, and closed-form expressions for the sound-speed estimation errors are derived. Typical acoustic navigation (RAFOS) signals are characterized by limited time resolution (0.2 s) challenging the accuracy of sound-speed estimation, depending on the location of the float with respect to the fixed stations. By exploiting travel-time data from multiple floats, the sound-speed estimation accuracy can be increased, which reflects in higher localization accuracy as well. In the case of a single float improved sound-speed estimates and localization results can be obtained by combining travel-time data from different float locations. Numerical results verify the theoretical error estimates and demonstrate the efficiency of the method. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3076202]

PACS number(s): 43.30.Pc, 43.60.Jn [AIT] Pages: 1384–1393

I. INTRODUCTION

Lagrangian floats play a significant role for the study of ocean currents and have been successfully used in the world's oceans and seas.[1–5] They are neutrally buoyant free drifting instruments which settle at a predetermined depth and follow the trajectory of the corresponding water parcel.[6,7] Long-range navigation of Lagrangian floats is commonly performed using acoustics: pulsed acoustic signals are emitted from moored sources and received on the floats (RAFOS floats[8,9]), or alternatively emitted from the floats and received at moored hydrophones (original tracking concept[10]), and acoustic travel times are measured. Assuming a value for the speed of sound the measured travel times are converted into ranges and the horizontal location of the floats is calculated by triangulation.

The advent of the ARGO system[11] has established a worldwide network of Lagrangian floats at large depths (2000 m), which every 10 days profile the water column by ascending to the sea surface from where they transmit temperature/salinity and position data through satellite and then return back to their parking depth to start another cycle. The deployment of ARGO floats at high latitudes is seriously hindered because of the presence of sea ice. In this connection, the enhancement of ARGO floats with a RAFOS receiving capability has been addressed.[12,13] This additional functionality will allow tracking of the floats at arbitrarily short intervals, e.g., two to three times a day (instead of every 10 days), and optimization of the surfacing process.

The signals used for acoustic navigation of Lagrangian floats are usually narrowband signals resulting in simple source design, at the cost of reduced time resolution, whereas the temporal sampling is also performed with low resolution—sampling period of the order of 0.2 s. This resolution can provide accuracy of ~300 m in range estimation assuming that the sound speed is precisely known. In reality, however, there are uncertainties in the sound speed, which result in additional localization errors. The aim of the present work is to study the feasibility of simultaneously estimating the sound speed and solving the localization problem from the low-resolution travel times. The anticipation is that besides the sound-speed estimation this approach will also increase localization accuracy.

To estimate the horizontal location of a RAFOS float, three sources (fixed stations)—all synchronized—are commonly used. The first two sources are used to estimate the location of the float subject to left-right ambiguity, with respect to the interconnecting line of the sources. The third source is used to resolve this ambiguity by providing additional travel-time (range) information.[8] This additional information from the third source is used here not only to remove the left-right ambiguity but also to make an estimation of the sound speed, simultaneously with solving the localization problem. The present study takes place in the framework of the Damocles Arctic research project,[14] and, in this connection, the main interest is in high-latitude areas. The Damocles project has recently deployed a group of RAFOS floats in the Arctic ocean. For their tracking pulsed acoustic signals emitted from a number of ice-tethered platforms are used. The frequency of the emitted pulses is 780 Hz and the pulse duration is ~0.1 s, whereas the sampling period on the floats is 0.18 s.

Localization problems relying on measured travel times and simultaneously improving the environmental information have been addressed by Dosso *et al.*, with application in array element localization,[15–17] including optimization of the

 0001-4966/2009/125(3)/1384/10/$25.00

source geometry,[18] as well as in acoustic tracking of sonobuoys.[19] These are short-range applications with time-measurement accuracy of the order of 0.001 s and normally distributed errors, whereas the solution of the non-linear localization problem relies on optimization algorithms, Monte Carlo methods, or iterative approaches combined with linearization. In long-range ocean acoustic tomography combined localization and environmental inversion problems arise in the case of untracked moving sources and/or receivers; such problems have been addressed in actual experiments[20] and simulation studies[21–23] in which the measurement of travel times has been sufficiently accurate to resolve and exploit multipath structure.

In the case of RAFOS transmissions the challenging factor is the low temporal resolution, of the order of 0.2 s. With such a resolution it is impossible to analyze the fine structure of the received signals and exploit multipath for the retrieval of the depth structure of the sound-speed distribution. Exploitation of multipath would require a higher temporal resolution, by at least one order of magnitude, as, e.g., in tomography experiments.[24,25] Besides, at high latitudes the variability of the sound speed with depth is small (10 m/s or less over the entire water column), resulting in temporal congestion of arrivals at the receiver for ranges up to a few hundred kilometers.[26] At lower latitudes the variability of sound speed with depth is larger resulting in larger time spread; this does not necessarily lead to an increase in travel-time uncertainty if robust observables are used, such as the arrival finale.[24]

As for the horizontal variability, range independence is a common assumption in ocean acoustic tomography experiments for areas away from ocean fronts. The information content of RAFOS data is very limited to retrieve the horizontal variability of the sound speed in addition to the float location. In this connection, a zero-order approach is adopted here focusing on the estimation of variable but constant sound speed. Although simple, this is an improvement compared to conventional RAFOS localization using *a priori* fixed sound-speed values. Based on the assumption of uniformity the estimation problem can be treated analytically in a linear framework. This approach leads to closed-form expressions for the sound-speed estimates and the associated estimation errors.

The main source of travel-time error considered here is the low-resolution discretization process, with typical step size of the order of 0.2 s. Further sources of error include the modeling approach, in particular, the assumption of uniform sound-speed distribution neglecting multipath along a section or differentiation of the sound speed between the different sections, and the measurement process including signal characteristics and signal processing.[24,27] For RAFOS signals the large size of the temporal discretization steps is considered to be the dominant factor behind travel-time errors. In this connection, the remaining error sources are not accounted for in the present analysis.

The contents of the work are organized as follows. Section II addresses the estimation of sound speed from travel times between floats and fixed stations in a linear analytic framework, and provides closed-form expressions for the sound-speed errors in the cases of one or several floats. In Sec. III some numerical examples are presented for sound-speed estimation and float localization and the validity of the theoretical predictions is assessed. In Sec. IV the results of the work are discussed and conclusions are drawn.

II. SOUND-SPEED ESTIMATION

In this section the case of one or several floats of unknown location communicating with three fixed stations in a medium of unknown uniform sound speed is considered.

A. Single float

For a float at (unknown) distances r_i, $i=1,2,3$, from three fixed stations, the corresponding exact travel times t_i will be

$$t_i = \frac{r_i}{c}, \tag{1}$$

where c is the true sound speed. Assuming that the float and the fixed stations are synchronized, the measured travel times are considered as truncated versions of t_i with discretization step $\delta\tau$ equal to the temporal resolution (typically 0.2 s in RAFOS transmissions)

$$\tau_i = [t_i]_{\delta\tau}, \tag{2}$$

where the brackets $[\cdot]_{\delta\tau}$ denote truncation to the nearest discrete value, with discretization step $\delta\tau$.

Even though the distances r_i are unknown, an approximate reference location (F_0) of the float can be estimated by making a guess of the sound speed and using the measured travel times, see Fig. 1. This reference location may in general not comply with all measured travel times, still it can be assumed to be close to the true location (F) of the float. Displacements about the reference location are described in terms of a rectangular coordinate system (x,y) with origin at the reference location F_0. If the distance between the float and the fixed stations is much larger than the float displacement (x,y) then the directions between the float and the fixed stations remain approximately the same as the float moves from the reference to the true location, such that the resulting change in the corresponding ranges can be approximated by the projection of the displacement vector on the reference directions $\varphi_{i,0}$,[20]

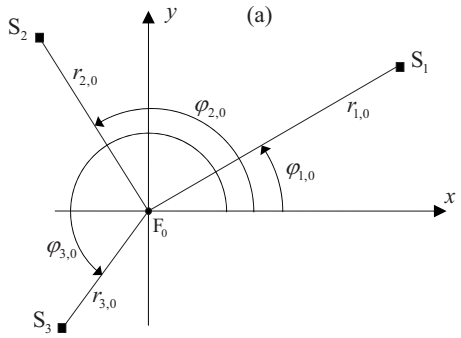
$$r_i = r_{i,0} - x\cos\varphi_{i,0} - y\sin\varphi_{i,0}, \quad i=1,2,3, \tag{3}$$

where $r_{i,0}$, $i=1,2,3$, are the reference ranges (the ranges from the fixed stations to the reference float location). This is, in fact, a linear approximation of the range r_i as a function of the displacement vector (x,y).

An estimate of the sound speed $(\hat{c})$ and the float displacement $(\hat{x},\hat{y})$ can be obtained by substituting Eq. (1) into Eq. (3) and replacing the exact with the discrete travel times. The resulting equations form a system of three linear equations with three unknowns

$$\hat{c}\tau_i + \hat{x}\cos\varphi_{i,0} + \hat{y}\sin\varphi_{i,0} = r_{i,0}, \quad i=1,2,3. \tag{4}$$

The solution of this system for the sound speed is



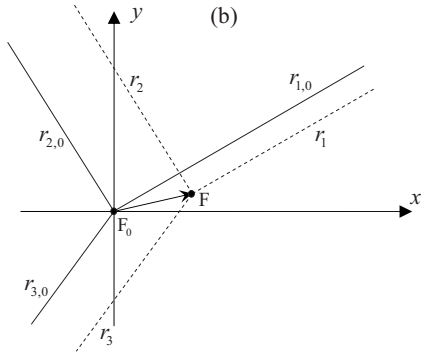


FIG. 1. (a) Geometrical configuration of fixed stations (S_1, S_2, and S_3) and float (F). The origin of the Cartesian coordinate system is taken at the reference position (F_0) of the float. (b) Detail of float displacement from the reference (F_0) to the true position (F) preserving directions to the fixed stations.

$$\hat{c} = \frac{r_{1,0}A_{32} + r_{2,0}A_{13} + r_{3,0}A_{21}}{\tau_1 A_{32} + \tau_2 A_{13} + \tau_3 A_{21}}, \tag{5}$$

where $A_{ij}=\sin(\varphi_{i,0}-\varphi_{j,0})$, $i,j=1,2,3$. The estimate $\hat{c}$ will in general deviate from the true sound speed c due to the deviation of the discrete travel times from the exact ones. Introducing the travel-time errors

$$\varepsilon_i = \tau_i - t_i, \quad \text{where } |\varepsilon_i| < \frac{\delta\tau}{2}, \quad i=1,2,3, \tag{6}$$

the measured travel time τ_i can be written as

$$\tau_i = \frac{r_i}{c} + \varepsilon_i. \tag{7}$$

Substituting this relation into the expression (5) for $\hat{c}$, the latter can be written in terms of the true sound speed and the travel-time errors as follows:

$$\hat{c} = \frac{r_{1,0}A_{32} + r_{2,0}A_{13} + r_{3,0}A_{21}}{\varepsilon_1 A_{32} + \varepsilon_2 A_{13} + \varepsilon_3 A_{21} + (r_1 A_{32} + r_2 A_{13} + r_3 A_{21})/c}. \tag{8}$$

In the Appendix it is shown that, as long as Eq. (3) is valid, the following identity holds:

$$r_1 A_{32} + r_2 A_{13} + r_3 A_{21} = r_{1,0}A_{32} + r_{2,0}A_{13} + r_{3,0}A_{21}. \tag{9}$$

This means that the expression $r_1A_{32}+r_2A_{13}+r_3A_{21}$ is invariant for small displacements about the reference location. Thus, the expression in parentheses in the denominator in Eq. (8) is identical to the numerator. Assuming that the travel-time errors are small and expanding the right-hand side of Eq. (8) to the first order with respect to ε_i the sound-speed estimate $\hat{c}$ can be written as

$$\hat{c} = c - \frac{c^2(\varepsilon_1 A_{32} + \varepsilon_2 A_{13} + \varepsilon_3 A_{21})}{r_{1,0}A_{32} + r_{2,0}A_{13} + r_{3,0}A_{21}} + O(\varepsilon^2). \tag{10}$$

For modeling purposes the travel-time errors ε_i are considered in the following as random variables uniformly distributed over the interval $(-\delta\tau/2, \delta\tau/2)$. In this case the expectation and variance of ε_i will be $\langle\varepsilon_i\rangle=0$ and $\langle\varepsilon_i^2\rangle=\delta\tau^2/12$, respectively. Hence, the expectation of $\hat{c}$ becomes

$$\langle\hat{c}\rangle = c, \tag{11}$$

which means that $\hat{c}$ is an unbiased estimator of c to the first order. This should be expected from a first-order expansion and zero-mean errors ε_i. Interestingly, second-order calculations also result in negligible bias, much smaller than the corresponding rms-error. Assuming that the errors ε_i are pairwise uncorrelated the variance of the sound-speed estimate $\hat{c}$ to the first order will be

$$\begin{aligned}\langle(\hat{c}-\langle\hat{c}\rangle)^2\rangle &= \frac{c^4(\langle\varepsilon_1^2\rangle A_{32}^2 + \langle\varepsilon_2^2\rangle A_{13}^2 + \langle\varepsilon_3^2\rangle A_{21}^2)}{(r_{1,0}A_{32} + r_{2,0}A_{13} + r_{3,0}A_{21})^2} \\ &= \frac{\delta\tau^2}{12}\frac{c^4(A_{32}^2 + A_{13}^2 + A_{21}^2)}{(r_{1,0}A_{32} + r_{2,0}A_{13} + r_{3,0}A_{21})^2},\end{aligned} \tag{12}$$

and the corresponding rms-error will be

$$\delta\hat{c}_{\text{rms}} = \frac{c^2\delta\tau}{2\sqrt{3}}\frac{\sqrt{A_{32}^2 + A_{13}^2 + A_{21}^2}}{r_{1,0}A_{32} + r_{2,0}A_{13} + r_{3,0}A_{21}}. \tag{13}$$

In this expression the quantities A_{ij} have to do with the gross geometry of the problem whereas the denominator, although involving the reference ranges, is invariant to perturbations of the reference position, Eq. (9), as shown in the Appendix. This means that the above estimate, Eq. (13), for the rms-error is independent from the reference position used. The error is proportional to the time discretization step, inversely proportional to the distances between the float and the fixed stations, and depends on the gross geometrical configuration.

B. Multiple floats

In the following the case of multiple floats $n=1,\ldots,N$ is considered, each communicating with three fixed stations. The fixed stations communicating with each float may be different, i.e., the fixed stations may be more than 3 in general. The previous analysis can be applied to the individual floats resulting in sound-speed estimates

$$\hat{c}_n = \frac{r_{n1,0}A_{n,32} + r_{n2,0}A_{n,13} + r_{n3,0}A_{n,21}}{\tau_{n1}A_{n,32} + \tau_{n2}A_{n,13} + \tau_{n3}A_{n,21}}, \tag{14}$$

where $r_{ni,0}$, $i=1,2,3$, are the reference ranges, $\varphi_{ni,0}$ the reference angles, $A_{n,ij}=\sin(\varphi_{ni,0}-\varphi_{nj,0})$, $i,j=1,2,3$, and τ_{ni} the measured travel times for the nth float. While the index n describes a particular float, the index i does not describe a particular fixed station on its own but in association with n. The corresponding sound-speed error is given by

$$\delta\hat{c}_{n,\mathrm{rms}} = \frac{c^2\delta\tau}{2\sqrt{3}} \frac{\sqrt{A_{n,32}^2 + A_{n,13}^2 + A_{n,21}^2}}{r_{n1,0}A_{n,32} + r_{n2,0}A_{n,13} + r_{n3,0}A_{n,21}}, \tag{15}$$

and will be different for each different float, depending on the particular geometry.

The minimum-variance unbiased linear estimator of the sound speed c can be shown to be the weighted average of the individual sound-speed estimates $\hat{c}_n$ with inverse variance weighting

$$c_{\mathrm{MV}} = \sum_{n=1}^{N} w_n \hat{c}_n, \quad w_n = \frac{1/\delta\hat{c}_{n,\mathrm{rms}}^2}{\Sigma_{m=1}^{N} 1/\delta\hat{c}_{m,\mathrm{rms}}^2}. \tag{16}$$

Assuming that the errors $\varepsilon_{ni} = \tau_{ni} - t_{ni}$, $n = 1, \ldots, N$, $i = 1, 2, 3$, are uniformly distributed in $(-\delta\tau/2, \delta\tau/2)$ the expectation of c_{MV} is clearly equal to c to the first order. Further, assuming that the errors ε_{ni} are pairwise uncorrelated the rms-error of the above sound-speed estimate takes the form

$$\delta c_{\mathrm{MV,rms}} = \frac{1}{\sqrt{N}} \frac{c^2\delta\tau}{2\sqrt{3}\Sigma_{n=1}^{N} 1/\delta\hat{c}_{n,\mathrm{rms}}^2} \times \sqrt{N\sum_{n=1}^{N} \frac{A_{n,32}^2 + A_{n,13}^2 + A_{n,21}^2}{(r_{n1,0}A_{n,32} + r_{n2,0}A_{n,13} + r_{n3,0}A_{n,21})^2 \delta\hat{c}_{n,\mathrm{rms}}^4}}. \tag{17}$$

Assuming that all floats communicate with the same three fixed stations and that they are close to each other, such that the same reference position can be used for all of them $(r_{ni,0} = r_{i,0})$, the rms-error in sound-speed estimation becomes

$$\delta c_{\mathrm{MV,rms}} = \frac{1}{\sqrt{N}} \frac{c^2\delta\tau}{2\sqrt{3}} \sqrt{\frac{A_{32}^2 + A_{13}^2 + A_{21}^2}{(r_{1,0}A_{32} + r_{2,0}A_{13} + r_{3,0}A_{21})^2}} = \frac{1}{\sqrt{N}} \delta\hat{c}_{\mathrm{rms}}, \tag{18}$$

i.e., using a set of N moving floats reduces the rms-error by a factor $1/\sqrt{N}$. This is compatible with the behavior of the rms-error of a number of independent observations.

III. NUMERICAL RESULTS

In this section some numerical results are presented for sound-speed estimation and float localization, assuming one or several floats communicating with three fixed stations. In all examples the temporal resolution is considered 0.2 s and the true sound-speed value is 1506.2 m/s, whereas the floats and the fixed stations are considered synchronized. From the exact ranges and the true sound-speed value the exact travel times are calculated, and then they are discretized to the nearest 0.2 s increment. The resulting discrete travel times are considered as the measured data, from which the sound-speed and float locations are estimated.

A. Effects of float location

In the first example the fixed stations form an orthogonal triangle of side 100 km in the two orthogonal directions. Figure 2 shows the theoretically predicted errors in sound-speed estimation (upper panel) and the actual sound-speed estimates (lower panel) as functions of the float location in the horizontal plane. In particular, the upper panel shows the theoretical error bound, given by the rms-error estimate, Eq. (13), multiplied by $\sqrt{3}$. This corresponds to a uniform distribution and represents the theoretical limit for the sound-speed error, which to the first order follows uniform distribution, cf. Eq. (10). From this figure it is seen that the predicted errors are small in the interior of the triangle defined by the three fixed stations and become large for float locations close to the circumcircle of the triangle (black line). Outside the circumcircle the errors drop with distance; nevertheless the decay is slower across the hypotenuse than across the other two sides.

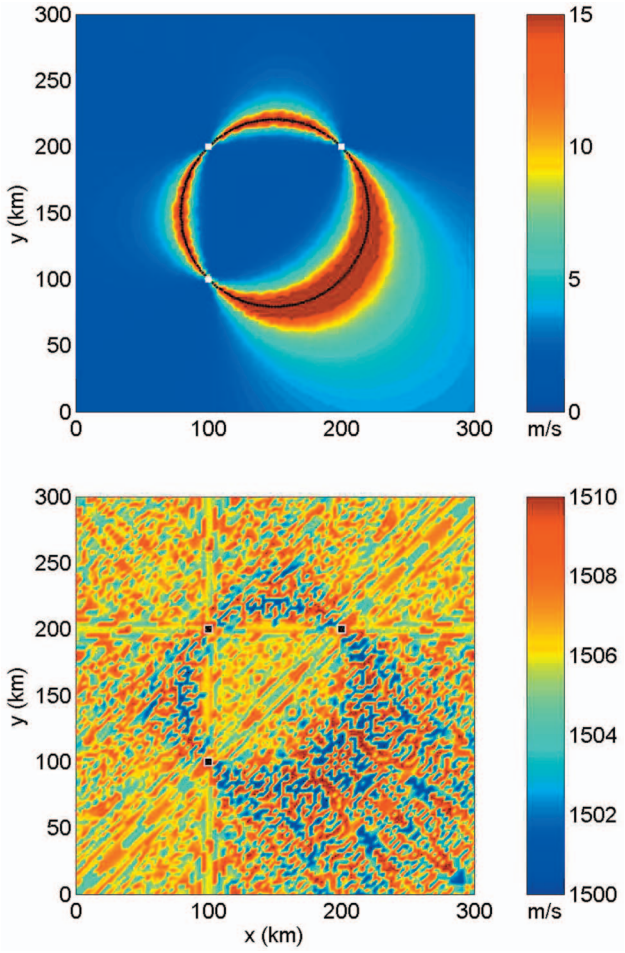


FIG. 2. Top: Theoretical error bounds as function of the location of the float communicating with three fixed stations (white squares) in orthogonal configuration. The black line denotes the circle through the fixed stations. Bottom: Actual sound-speed estimates for various float locations. The fixed stations are denoted by black squares.

The lower panel in Fig. 2 shows the actual sound-speed estimates from the discrete travel times, Eq. (5), at the various float locations. The yellow color in this figure represents the true sound-speed value (1506.2 m/s). The deviations from this value are compatible with the theoretical error estimates and behavior shown in the upper panel. In general the errors appear to be smaller inside the circumcircle than outside, and largest in the vicinity of the circle. The three straight lines defined by the fixed stations can be discrimi-

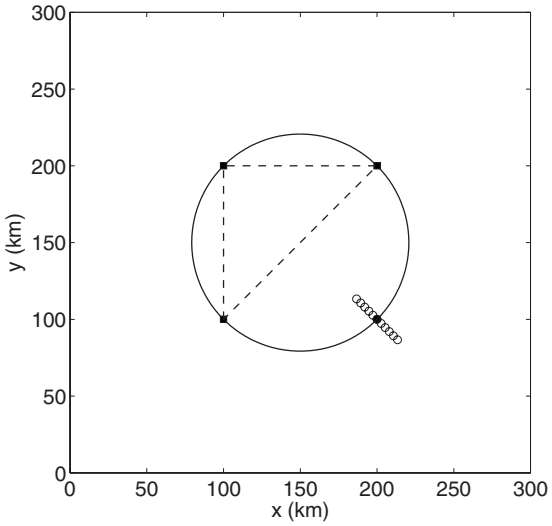


FIG. 3. Indefinitiveness of sound-speed estimation–localization problem: Estimated float locations (○) for 11 different sound-speed values, from 1300 to 1700 m/s, when the true float location (●) lies on the circumcircle.

nated in this figure (one at $x=100$ km, one at $y=200$ km, and one corresponding to the hypotenuse). When the float lies on one of those lines the problem of sound-speed estimation becomes one-dimensional. The reason is that the two in-line stations are sufficient in this case for estimating the sound-speed, and this is what Eq. (5) does neglecting the travel-time information from the third station.

In Fig. 2 the sound-speed errors increase as the float location approaches the circumcircle defined by the three fixed stations. The equation of that circle is

$$r_1 A_{32} + r_2 A_{13} + r_3 A_{21} = 0. \tag{19}$$

This explains the above-mentioned behavior since by virtue of Eq. (19) the predicted error, Eq. (13), becomes infinite for float locations lying on the circle. From another point of view, the denominator in Eq. (5), i.e., the determinant of the linear system (4), becomes very close to zero for float locations in the vicinity of the circumcircle, which renders the three equations (4) linearly dependent. This means that only two equations are left for the three unknowns $(\hat{c},\hat{x},\hat{y})$, such that they accept a solution (an acceptable float location) for any arbitrary value of the sound speed $\hat{c}$. This indefinitiveness is shown in Fig. 3 for the case of a float whose true location (heavy dot) is on the circumcircle. The figure shows 11 float locations (open circles) resulting from Eq. (4) for 11 different sound-speed values, from 1300 to 1700 m/s. The indefinitiveness for float locations on the circumcircle does not mean singularity of the localization problem. In fact, if the true sound-speed value is known then the true float location can be correctly estimated, even for locations on the circumcircle. The singularity refers to the simultaneous localization and sound-speed estimation problem.

In the following examples, a configuration of three fixed stations forming an equilateral triangle of side 100 km is considered. Figure 4 presents the theoretical error bounds in sound-speed estimation (upper panel) and the actual sound-speed estimates (lower panel) as functions of the float location. The triangle symmetry in this case reflects in the predicted errors and also in the deviations of the estimated sound speed from the true value (1506.2 m/s). The straight lines defined by the fixed stations can be distinguished in the lower panel as in the previous case. The sound-speed error becomes larger as the float location comes closer to the circumcircle. On the other hand the error for float locations inside the circumcircle appears to be lower than for locations outside. To look into this more clearly, a section along the line AA parallel to the x-axis for $y=175$ km is considered.

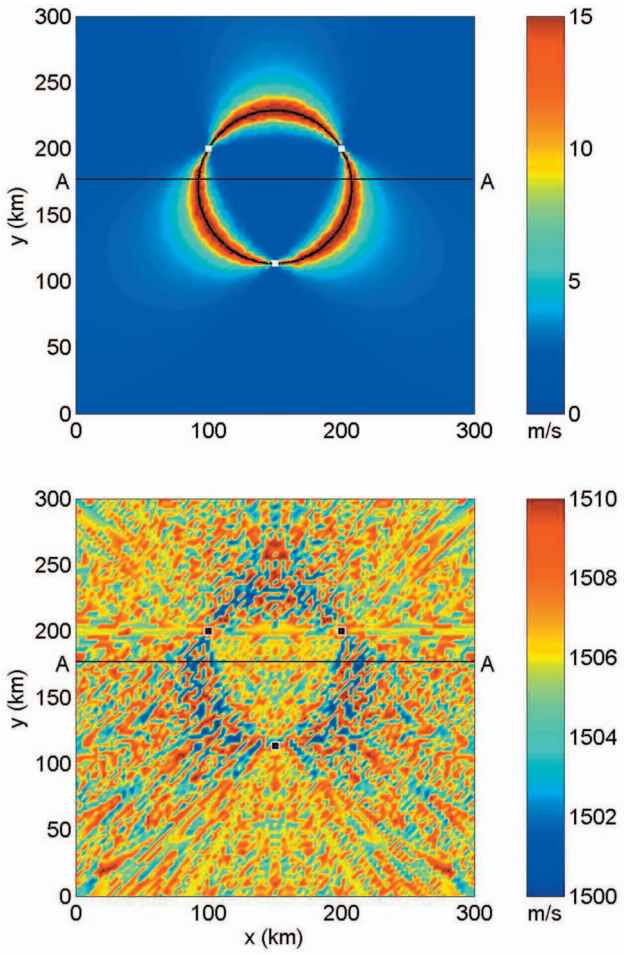


FIG. 4. Top: Theoretical error bounds as function of the location of the float communicating with three fixed stations (white squares) in equilateral configuration. The circle through the fixed stations is denoted in black. Bottom: Actual sound-speed estimates for various float locations. The fixed stations are denoted by black squares.

Figure 5 shows the predicted error bounds (solid lines) and the actual errors (resulting by subtracting the true sound-speed value 1506.2 m/s from the actual sound-speed estimates), along the line AA of Fig. 4. Both the theoretical and the actual errors are seen to be smaller inside the circumcircle than outside, whereas they become unbounded on the trace of the circumcircle on the line AA. The actual errors (dots) lie in most cases within the predicted error bounds, and the agreement between the two is remarkable. The points that lie outside the predicted bounds correspond to cases where the second-order terms $O(\varepsilon^2)$ in the sound-speed expansion, Eq. (10), become significant. In that case the distribution of the sound speed deviates from the uniform distri-

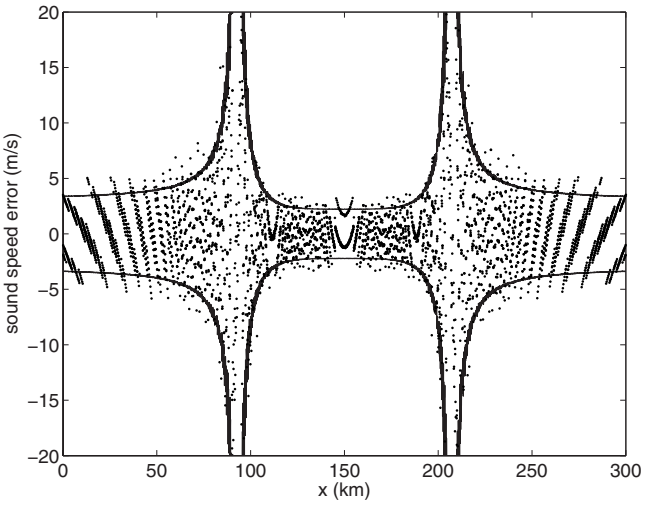


FIG. 5. Section of Fig. 4 along the line AA. The solid lines show the predicted error bounds and the dots represent the actual sound-speed errors (deviations between sound-speed estimates and true value). The resolution along the *x*-axis (measuring the float location) is 50 times higher than that of Fig. 4.

bution and begins to form "tails." A sign of these tails is the existence of sound-speed values outside the error bounds predicted by linear theory. Nevertheless, it is remarkable how well the first-order theoretical prediction performs in describing the actual error variance. Further, it is seen that the errors of the actual sound-speed estimates are evenly distributed about zero, pointing to a negligible bias. The *x*-resolution used in Fig. 5 is 60 m, 50 times higher than that used in Fig. 4 (3000 m), and reveals interesting structures, such as the striation patterns, particularly toward the left and right ends as well as in the middle; these patterns are associated with jumps between different discrete travel-time values. Further, the actual errors close to the traces of the straight lines connecting the fixed stations exhibit increased coherence, which is associated with the fact that the estimation problem becomes one-dimensional for float locations on these lines.

A way to avoid the infinite sound-speed errors for float locations on the circumcircle is by constraining the inversion using *a priori* information. To check this, a statistical inversion approach[28,18] is applied to the localization and sound-speed estimation problem. The *a priori* information for both the sound speed and the float locations is assumed to have the form of Gaussian distributions. In particular, the deviation from the reference location is assumed to follow a zero-mean Gaussian distribution with standard deviation 1000 m, whereas the sound speed is assumed to be normally distributed about a "guess" value of 1500 m/s. Three alternative values are considered for the standard deviation of the latter distribution: 20, 10, and 5 m/s. Finally, the exact travel times are considered to follow uniform distribution about the observed travel-time values, as in the analytic approach.

The *a posteriori* probability density function is sampled using a Monte Carlo method[29] for float locations along the same line AA shown in Fig. 4. The resulting sound-speed populations (not shown here) are distributed about the correct value (1506.2 m/s). The standard deviations of these populations about 1506.2 m/s are shown in Fig. 6 for the

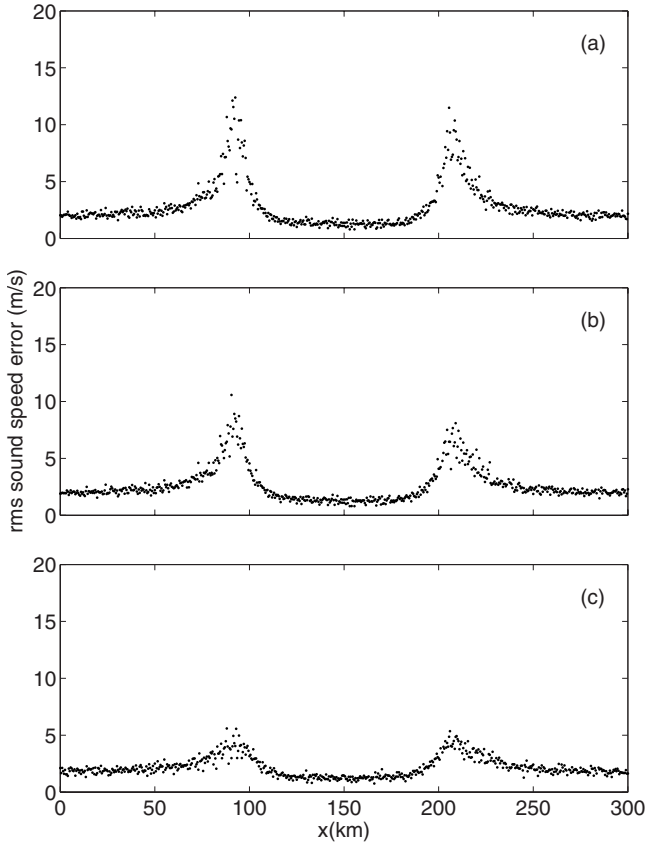


FIG. 6. Standard deviations of the statistical inversion results from the true sound-speed value corresponding to three values of *a priori* standard deviation for the sound speed: (a) 20 m/s, (b) 10 m/s, and (c) 5 m/s.

three cases of the standard deviation for the sound speed: 20, 10, and 5 m/s. It is seen in Fig. 6 that the reduction in the standard deviation for the sound speed leads to gradual suppression of the errors for float locations close to the circumcircle, as expected. Apart from that, the statistical inversion scheme results in smaller sound-speed errors inside the circle than outside, i.e., in a similar behavior as the analytic approach. In comparing Figs. 5 and 6 it should be noted that Fig. 5 shows the actual sound-sound estimates and the predicted error bounds, whereas Fig. 6 shows the rms-errors.

B. Combination of travel-time data from different floats

The next numerical example addresses the effect that a combination of travel-time data from a group of floats has on sound-speed estimation and localization. In this connection a random distribution of 37 floats around the three fixed stations (same as in the previous example) is considered, as shown in Fig. 7. Floats A, B, and C are floats for which localization results will be shown. Figure 8 shows the sound-speed estimates obtained by combining the 37 floats in three different ways. In the upper two panels the floats are sorted by their distance from the center of the circle; in the upper panel they are taken in order of decreasing distance, i.e., from the outmost float (C) inward, whereas in the middle panel they are taken in order of increasing distance, i.e., from the inmost float (A) outward. In the lower panel the floats are taken in order of increasing error. The floats outside the cir-

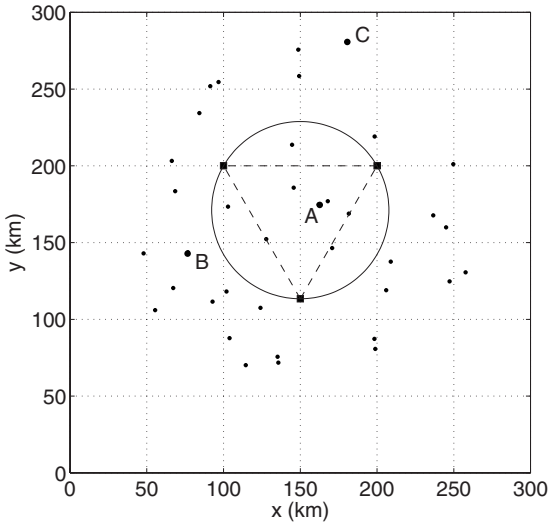


FIG. 7. Random distribution of floats (●) around a set of three fixed stations (■) in equilateral configuration. Floats A, B, and C are floats for which localization results will be shown.

cumcircle are characterized in general by larger sound-speed errors than the floats inside, whereas the floats close to the circumcircle are characterized by the largest errors, cf. Fig. 4.

From Fig. 8 it is seen that in all cases the addition of more floats lowers the predicted sound-speed error bounds

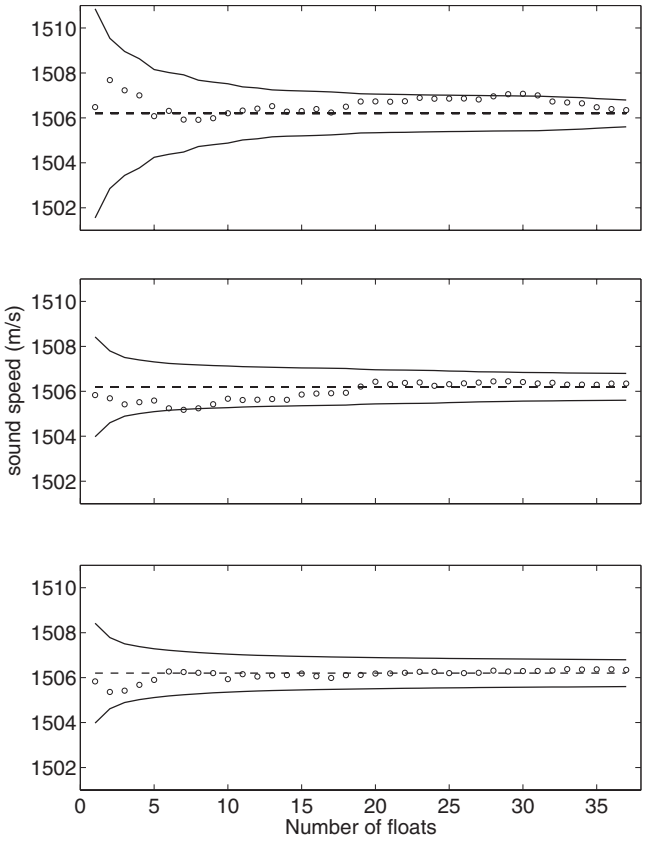


FIG. 8. Sound-speed estimates (○) and theoretical error bounds (solid lines) about the true sound-speed value (dashed line) resulting from combining the floats of Fig. 7 in three different ways. Top: In order of decreasing distance from the center. Middle: In order of increasing distance from the center. Bottom: In order of increasing error.

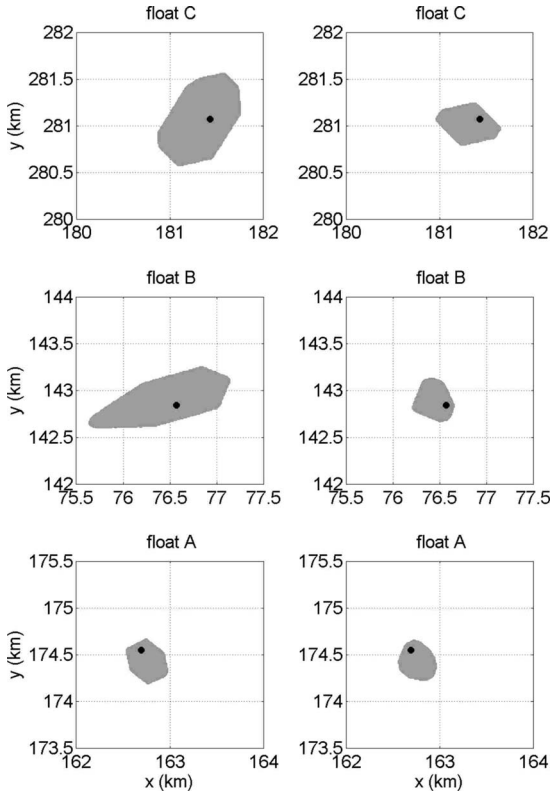


FIG. 9. Localization results for floats A, B, and C, shown in Fig. 7, based on individual sound-speed estimates (single float location—left) and combined sound-speed estimates (from all 37 floats—right). The estimated locations are denoted by the shaded areas and the true float locations by the heavy dots.

(solid lines) monotonically. Further, the theoretical predictions are compatible with the actual sound-speed estimates (open circles). In the first case (upper panel) the initial error—the one corresponding to float C—is significantly larger than in the second case (middle panel)—the one corresponding to float A. This is due to the less favorable location of float C, outside the circumcircle, compared to the location of float A, close to the center of the circumcircle. On the other hand, when all 37 floats are taken into account the same error is obtained: ± 0.6 m/s for the theoretically predicted variability interval (solid lines) and 0.15 m/s for the actual error (open circle). The same holds when the floats are sorted in order of increasing error (lower panel), in which case, however, the above limits appear to be approximated more efficiently.

Figure 9 shows the results of the localization for floats A, B, and C located at characteristic locations with respect to the circumcircle: float A lies closest to the center, float B lies close to the circumference, and float C is the outmost one, cf. Fig. 7. The left panels of Fig. 9 show the localization results based on the individual sound-speed estimates from the single floats, whereas the results in the right panels are based on the best sound-speed estimates (combination of all 37 floats). The localization improvement is marginal for float A but significant for the outmost float C and largest for float B. This is associated with the uncertainties in sound-speed estimation resulting from the travel times of the single floats: Floats close to the center are characterized by smaller uncer-

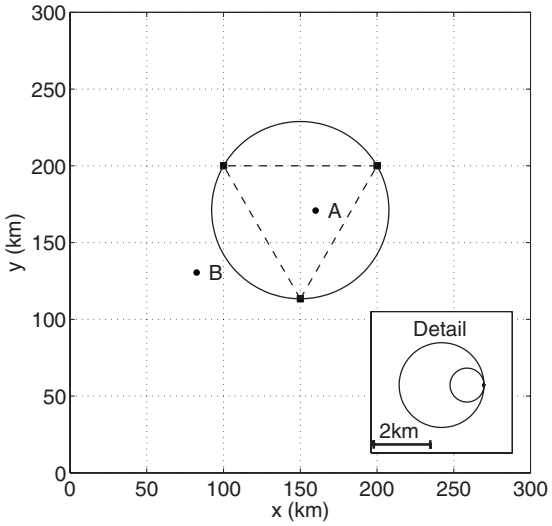


FIG. 10. Configuration for sound-speed estimation by combining data from a single float as it moves along a circular trajectory. Two float locations (A and B) are considered whereas the fixed stations form an equilateral triangle. The circular trajectories are shown in the embedded figure at the lower-right corner.

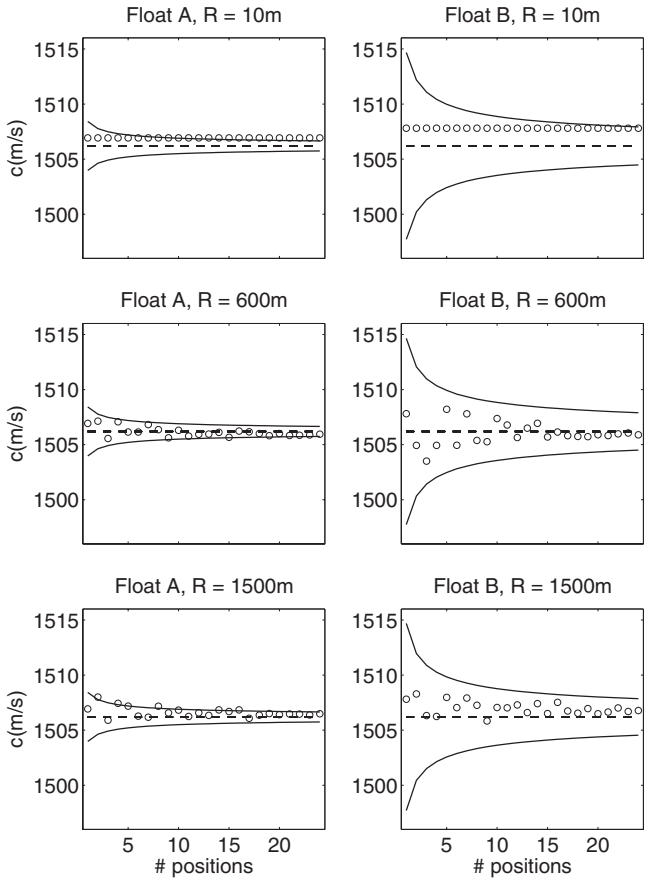


FIG. 11. Sound-speed estimates (○) for float A (left panels) and float B (right panels) and for three different circular radii: 10 m (top), 600 m (middle), and 1500 m (bottom) as functions of the number of locations uniformly distributed on the circular trajectory. The theoretical estimates for the error bounds assuming uncorrelated travel-time errors are shown through solid lines. The true sound speed is denoted by the dashed lines.

tainties than exterior floats, whereas the floats close to the circumference have the largest uncertainties. The resulting location uncertainty based on the individual sound-speed estimates spans about 1 km for float C and 1.5 km for float B whereas it is less than 500 m for float A; the highest range accuracy that can be obtained with a travel-time discretization step of 0.2 s is 300 m. The difference between the individual and the average sound-speed estimate is small for float A, larger for float C (cf. Fig. 8) and largest for float B lying close to the circumcircle. The obtained localization uncertainties using the average sound-speed estimate are of the order of 500 m or lower.

C. Combination of travel-time data from a single float at different locations

In the examples considered so far the combination of travel-time data from multiple floats leads in general to improvements both in sound-speed estimation and localization accuracy, compared to the case of a single float. The improvement is due to the random location of the floats relative to the fixed stations, reflecting into random travel-time errors averaging to zero. In the light of this result one could think of improving the sound-speed estimate obtained from a single float by combining travel-time data from different float locations, i.e., at different times, assuming that the sound speed remains unchanged. If the locations are far enough from each other the effect on the travel times should be similar to that of multiple floats. In this connection, the last example involves exploitation of travel-time data from a single float at different locations. For convenience the float is assumed to follow an inertial current, i.e., a circular trajectory. Radii of 600 and 1500 m are considered corresponding to current velocities of 0.04 and 0.1 m/s, respectively, assuming polar latitudes. A third case with radius as small as 10 m is also considered for comparison purposes. Two float locations are examined, one inside the circumcircle (A) and one outside (B), as shown in Fig. 10. The float locations and their trajectories are shown in Fig. 10. The fixed stations are considered in the same equilateral configuration as before.

Figure 11 shows the sound-speed estimates for floats A and B for each of the three trajectories as a function of the number of positions combined (assumed uniformly distributed on the circular trajectory). The solid lines in this figure represent the theoretical error bounds assuming the travel-time errors to be statistically independent. When the radius is only 10 m the various float locations on the circle are characterized by identical discrete travel times. This means identical sound-speed estimates from each float position, in which case the average makes no difference; this is seen in the top panels in Fig. 11. In this case the assumption of statistical independence is clearly not fulfilled. As the radius becomes larger there are differences in the discrete travel times for the different float locations and this in turn leads to differences in the sound-speed estimates; this is seen in the middle and lower panels of Fig. 11. In general the errors are significantly larger for the outside float B than for the inside float A, still in both cases better sound-speed estimates are obtained when the number of positions on the trajectories increases. The agreement between the predicted error bounds and the spread of the actual sound-speed estimates in the last two cases is remarkable.

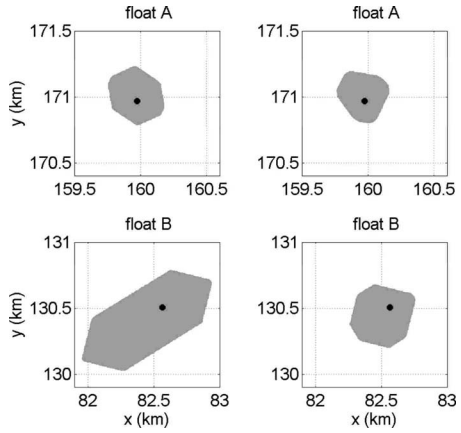


FIG. 12. Localization results for floats A and B of Fig. 10 based on individual sound-speed estimates (single float location—left) and combined sound-speed estimates (24 locations along the circular trajectory of radius 1500 m—right). The estimated locations are denoted by the shaded areas and the true float locations by the heavy dots.

Figure 12 shows the localization results for floats A and B based on the initial sound-speed estimates—travel-time data from a single float location—(panels on the left) and the results based on the combined travel-time data from all 24 locations along the circular trajectory of radius 1500 m (panels on the right). The improvement is marginal for the case of float A, whose favorable location close to the center results in a small sound-speed error (and thus localization error) even using travel-time data from a single location of the float. In the case of float B the increase in the localization accuracy is much clearer to see. The reason is that float B is close to the circumcircle and thus the initial sound-speed error is large contributing to localization uncertainty of the order of 1 km; note that the locations A and B in Fig. 10 are different from those in Fig. 7. With the improvement in the sound-speed accuracy obtained by the combination of travel-time data at 24 different float locations the localization uncertainty drops to less than 500 m.

IV. DISCUSSION AND CONCLUSIONS

In this work a method was proposed for simultaneous estimation of sound speed and localization of RAFOS floats from low-resolution travel-time data. Closed-form expressions for the sound-speed estimates and the corresponding errors were obtained based on two linearization assumptions; (i) linearization of range variations between floats and fixed stations with respect to float displacements and (ii) linearization of sound-speed estimates with respect to travel-time errors. The first linearization can be justified if the distances between floats and fixed stations are large, larger than a few kilometers, taking into account that the initial uncertainty associated with the float location is of the order of 1 km. The second linearization is supported by the numerical results presented here, in particular, by the agreement between the actual sound-speed errors and the error bounds predicted from the first-order approach.

The resulting closed-form expressions allow for a parametric study of the sound-speed and localization errors. The present approach does not account for *a priori* information constraining the sound speed, and in this connection, it leads to infinite sound-speed errors for float locations on the circumcircle through the fixed stations. The singularity arises because of linear dependence of the associated equations, which means that the sound-speed estimation problem becomes indefinite (any value of the sound speed is accepted). This indefinitiveness has to do with the combined sound-speed estimation and localization problem, not with the localization problem alone, i.e., if the true sound speed is known then Eq. (4) leads to the true float positions, even on the circumcircle.

Further, the predicted sound-speed errors are in general smaller for float locations inside the circumcircle than outside, and largest for locations in the neighborhood of the circumcircle. These results are in agreement with the behavior of actual sound-speed estimates. As a further check a statistical inversion approach was applied to the non-linear estimation problem imposing *a priori* constraints to sound-speed and position uncertainties in the form of Gaussian probability density functions. Apart from the suppression of the sound-speed singularities for float locations close to the circumcircle, the statistical approach results in a similar behavior of the sound-speed error, with smaller values inside the circle, larger outside, and largest (yet finite) on the circle, thus confirming the analytic results to a large extent.

Improved sound-speed estimates can be obtained from a group of floats by taking weighted averages over individual sound-speed estimates (based on travel times from single floats). In this case the best, in terms of minimum variance, linear estimate of the true sound speed is the weighted mean of the individual sound-speed estimates with weights proportional to the inverse variances. The thus obtained sound-speed estimate exhibits a monotonic improvement (decrease in estimation error) as additional floats are taken into account irrespective of the location of the floats. Nevertheless, the location of each new float affects the degree of improvement since, e.g., floats near the circumcircle are characterized by larger variances, and they are given smaller weights, than floats close to the center.

Improved knowledge of the sound speed reflects in improved localization results. For floats close to the center of the circumcircle the errors of the individual sound-speed estimates are small enough such that the localization uncertainty is mainly governed by the temporal resolution ($\sim$300 m for $\delta\tau$=0.2 s). For floats close to or outside the circumcircle the individual sound-speed estimates are subject to larger uncertainties which contribute to increased localization errors (1000–1500 m in the examples considered). These errors can be reduced by improving the sound-speed estimates. This can be achieved by combining data from different floats or from the same float at different locations. The latter is a useful alternative under the assumption that the sound speed remains unchanged, and provided that the different locations are sufficiently apart from each other, such that the resulting changes in travel times are larger than the discretization step, and the corresponding travel-time errors can be considered as pairwise uncorrelated.

ACKNOWLEDGMENTS

This work was supported by the EU Sixth Framework Program as part of the Damocles Integrated Project (Contract No. 018509).

APPENDIX

Using the linearized relation, Eq. (3), for the actual ranges r_1, r_2, and r_3 the expression $r_1A_{32}+r_2A_{13}+r_3A_{21}$ can be written as

$$
\begin{aligned}
r_1A_{32}+r_2A_{13}+r_3A_{21} &= (r_{1,0}-x\cos\varphi_{1,0}-y\sin\varphi_{1,0})\sin(\varphi_{3,0}-\varphi_{2,0})+(r_{2,0}-x\cos\varphi_{2,0}-y\sin\varphi_{2,0})\sin(\varphi_{1,0}-\varphi_{3,0})\\
&\quad+(r_{3,0}-x\cos\varphi_{3,0}-y\sin\varphi_{3,0})\sin(\varphi_{2,0}-\varphi_{1,0})\\
&= r_{1,0}A_{32}+r_{2,0}A_{13}+r_{3,0}A_{21}-x[\cos\varphi_{1,0}\sin(\varphi_{3,0}-\varphi_{2,0})+\cos\varphi_{2,0}\sin(\varphi_{1,0}-\varphi_{3,0})\\
&\quad+\cos\varphi_{3,0}\sin(\varphi_{2,0}-\varphi_{1,0})]-y[\sin\varphi_{1,0}\sin(\varphi_{3,0}-\varphi_{2,0})+\sin\varphi_{2,0}\sin(\varphi_{1,0}-\varphi_{3,0})\\
&\quad+\sin\varphi_{3,0}\sin(\varphi_{2,0}-\varphi_{1,0})]\\
&= r_{1,0}A_{32}+r_{2,0}A_{13}+r_{3,0}A_{21}-x[\cos\varphi_{1,0}\sin\varphi_{3,0}\cos\varphi_{2,0}-\cos\varphi_{1,0}\cos\varphi_{3,0}\sin\varphi_{2,0}\\
&\quad+\cos\varphi_{2,0}\sin\varphi_{1,0}\cos\varphi_{3,0}-\cos\varphi_{2,0}\cos\varphi_{1,0}\sin\varphi_{3,0}+\cos\varphi_{3,0}\sin\varphi_{2,0}\cos\varphi_{1,0}\\
&\quad-\cos\varphi_{3,0}\cos\varphi_{2,0}\sin\varphi_{1,0}]-y[\sin\varphi_{1,0}\sin\varphi_{3,0}\cos\varphi_{2,0}-\sin\varphi_{1,0}\cos\varphi_{3,0}\sin\varphi_{2,0}\\
&\quad+\sin\varphi_{2,0}\sin\varphi_{1,0}\cos\varphi_{3,0}-\sin\varphi_{2,0}\cos\varphi_{1,0}\sin\varphi_{3,0}+\sin\varphi_{3,0}\sin\varphi_{2,0}\cos\varphi_{1,0}\\
&\quad-\sin\varphi_{3,0}\cos\varphi_{2,0}\sin\varphi_{1,0}].
\end{aligned}
$$

The six summands in each bracket cancel each other pairwise, such that finally

$$r_1A_{32}+r_2A_{13}+r_3A_{21}=r_{1,0}A_{32}+r_{2,0}A_{13}+r_{3,0}A_{21}.$$

This means that the expression $r_1A_{32}+r_2A_{13}+r_3A_{21}$ is invariant for small displacements about the reference location.

[1]T. Rossby and D. Webb, "Observing abyssal motion by tracking Swallow floats in the SOFAR channel," Deep-Sea Res. **17**, 359–365 (1970).
[2]R. E. Davis, "Preliminary results from directly measuring mid-depth circulation in the Tropical and South Pacific," J. Geophys. Res. **103**, 24619–24639 (1998).
[3]H. M. Zhang, M. D. Prater, and T. Rossby, "Isopycnal Lagrangian statistics from the North Atlantic Current RAFOS float observations," J. Geophys. Res. **106**, 13817–13836 (2001).
[4]P. Testor and J.-C. Gascard, "Large-scale spreading of deep waters in the Western Mediterranean Sea by submesoscale coherent eddies," J. Phys. Oceanogr. **33**, 75–87 (2003).
[5]K. L. Lavender, W. B. Owens, and R. E. Davis, "The mid-depth circulation of the subpolar North Atlantic Ocean as measured by subsurface floats," Deep-Sea Res., Part I **52**, 767–785 (2005).
[6]H. T. Rossby, E. R. Levine, and D. N. Connors, "The isopycnal Swallow float—A simple device for tracking water parcels in the ocean," Prog. Oceanogr. **14**, 511–525 (1985).
[7]D. D. Swift and S. C. Riser, "RAFOS floats: Defining and targeting surfaces of neutral buoyancy," J. Atmos. Ocean. Technol. **11**, 1079–1092 (1994).
[8]T. Rossby, D. Dorson, and J. Fontaine, "The RAFOS system," J. Atmos. Ocean. Technol. **3**, 672–679 (1986). RAFOS is the inverted acronym for Sound Fixing and Ranging—SOFAR.
[9]C. M. Wooding, H. H. Furey, and M. A. Pacheco, "RAFOS floats processing at the Woods Hole Oceanographic Institution," (Technical Report No. WHOI-2005-2, Woods Hole Oceanographic Institution, Woods Hole, 2005); https://darchive.mblwhoilibrary.org/bitstream/1912/55/1/WHOI-2005-02.pdf (Last viewed October, 2008).
[10]J. C. Swallow, "A neutral-buoyancy float for measuring deep currents," Deep-Sea Res. **3**, 93–104 (1955).
[11]D. Roemmich and W. B. Owens, "The Argo Project: Global ocean observations for understanding and prediction of climate variability," Oceanogr. **13**, 45–50 (2000).
[12]O. Klatt, O. Boebel, and E. Fahrbach, "A profiling float's sense of ice," J. Atmos. Ocean. Technol. **24**, 1301–1308 (2007).
[13]S. Riser, "Low-frequency acoustics (RAFOS) and profiling floats," ONR Progress Report No. N00014-03-1-0615, Office of Naval Research, Arlington, 2006); http://www.onr.navy.mil/sci_tech/32/reports/docs/06/poriser4.pdf (Last viewed October, 2008).
[14]Developing arctic modeling and observing capabilities for long-term environmental studies (DAMOCLES) (2005–2009) is an integrated project funded by the EU Sixth framework program.
[15]S. E. Dosso, M. R. Fallat, B. J. Sotirin, and J. L. Newton, "Array element localization for horizontal arrays via Occam's inversion," J. Acoust. Soc. Am. **104**, 846–859 (1998).
[16]S. E. Dosso and M. Riedel, "Array element localization for towed marine seismic arrays," J. Acoust. Soc. Am. **110**, 955–966 (2001).
[17]S. E. Dosso, N. E. B. Collison, G. J. Heard, and R. I. Verrall, "Experimental validation of regularized array element localization," J. Acoust. Soc. Am. **115**, 2129–2137 (2004).
[18]S. E. Dosso and B. J. Sotirin, "Optimal array element localization," J. Acoust. Soc. Am. **106**, 3445–3459 (1999).
[19]S. E. Dosso and N. E. B. Collison, "Acoustic tracking of a freely drifting sonobuoy field," J. Acoust. Soc. Am. **111**, 2166–2177 (2002).
[20]E. K. Skarsoulis, "Multi-section matched-peak tomographic inversion with a moving source," J. Acoust. Soc. Am. **110**, 786–797 (2001).
[21]B. Cornuelle, "Simulation of acoustic tomography array performance with untracked or drifting sources and receivers," J. Geophys. Res. **90**, 9079–9088 (1985).
[22]F. Gaillard, "Ocean acoustic tomography with moving sources or receivers," J. Geophys. Res. **90**, 11891–11898 (1985).
[23]T. F. Duda, R. A. Pawlowicz, J. F. Lynch, and B. D. Cornuelle, "Simulated tomographic reconstruction of ocean features using drifting acoustic receivers and a navigated source," J. Acoust. Soc. Am. **98**, 2270–2279 (1995).
[24]W. Munk, P. Worcester, and C. Wunsch, *Ocean Acoustic Tomography* (Cambridge University Press, New York, 1995).
[25]T. F. Duda, B. M. Howe, and J. H. Miller, "Acoustics in global process ocean observatories," Sea Technol. **48**, 35–38 (2007).
[26]J. F. Lynch, R. C. Spindel, C. S. Chiu, J. H. Miller, and T. G. Birdsall, "Results from the 1984 marginal ice zone experiment preliminary tomography transmissions: Implications for marginal ice zone, Arctic, and surface wave tomography," J. Geophys. Res. **92**, 6869–6885 (1987).
[27]T. F. Duda, "Analysis of finite-duration wide-band frequency sweep signals for ocean tomography," IEEE J. Ocean. Eng. **18**, 87–94 (1993).
[28]A. Tarantola, *Inverse Problem Theory* (Elsevier, Amsterdam, 1987).
[29]W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, *Markov Chain Monte Carlo in Practice* (Chapman and Hall, New York, 1998).

Measurements of three-dimensional propagation in a continental shelf environment

Kevin D. Heaney[a] and James J. Murray
Ocean Acoustical Services and Instrumentation Systems, Inc., 11006 Clara Barton Drive, Fairfax Station, Virginia 22039

(Received 20 October 2008; revised 12 December 2008; accepted 15 December 2008)

Although a significant amount of theoretical and numerical modeling effort has been put into the study of three-dimensional (3D) acoustic propagation on a coastal wedge, including the development of the ASA 3D benchmark problem set, there have been few observations of the predicted 3D propagation effects. Significant horizontal multipath arrivals were observed in a pair of acoustic transmission tests on the continental shelf off the east coast of Florida in September 2007 and February 2008. For many transmissions, arrivals were received coming from nearly the global positioning system (GPS) bearing of the ship, as well as up to 30 deg inshore of the true bearing. The inshore path was up to 25 dB stronger than the direct path in some cases. The experimental waveforms transmitted included continuous-wave transmissions ranging in frequency from 24 to 415 Hz as well as wideband linear frequency modulation pulses (20–420 Hz). Horizontal multipath arrivals were observed for source ranges from 10 to 80 km, source depths of 20 and 100 m, and along several different bearings (inshore and along the 250 m isobath). It is a conclusion of this paper that the bearing bias and multiple horizontal arrivals are the result of 3D propagation due to the local shoaling bathymetry. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3075558]

PACS number(s): 43.30.Xm, 43.30.Bp, 43.30.Cq [AIT] Pages: 1394–1402

I. INTRODUCTION

The phenomenon of three-dimensional (3D) acoustic propagation in a continental shelf environment is well understood both theoretically and from a numerical modeling perspective. The propagation feature of interest in the wedge environment is the horizontal refraction of an acoustic ray (or equivalent normal mode) due to repeated small angle changes from specular reflections off the sloping seafloor. Phenomena predicted include the horizontal refraction of sound traveling at oblique angles relative to the shoreline, an acoustic shadow zone with a leading edge caustic, intermode interference and focusing, as well as the selective cutoff of acoustic normal-modes with water depth. The impenetrable wedge was addressed via acoustic normal-modes by Pierce[1] and Bradley[2] and then extended to the penetrable wedge problem by Buckingham.[3] Harrison[4] looked at the problem of basin scale 3D propagation and horizontal shadow zones[5] via ray invariants and extended this to a wave field solution.[6] In 1974 Weinburg and Burridge[7] presented the definitive paper on horizontal ray theory using vertical normal modes. This paper presented the idea of using vertical normal mode phase speeds as the background sound speed field in which to trace horizontal acoustic rays. Numerically this was extended to vertical modes and the horizontal parabolic equation (PE) by Collins,[8] applied to the basin scale 3D propagation problem[9] and adapted to include coupled modes by Abawi and Kuperman.[10] Recent development of the fully 3D PE solution has confirmed our previous physical intuition.[11] In 1995 the Acoustical Society of America generated the "ASA wedge benchmark" test cases that were solved using multiple methods, including theoretical approaches,[12] rays, normal modes,[13] and a set of 3D parabolic equations.[14] Predictions were made that 3D effects would affect matched field processing[15] and shallow water propagation.[16]

Quantitative measurements of 3D propagation, however, are rare in literature. Recently Frank *et al.*[17] described 3D propagation measurements from the scattering of sound by non-linear internal waves. Heaney and Baggeroer[18] observed 3D propagation due to sloping bathymetry off the coast of Kauai in 2004. In this paper, we present results from a set of experiments where 3D propagation effects were not only observed, but dominated acoustic propagation. Measurements in the summer of 2007 taken on the continental shelf 20 km off the east coast of Florida indicated a bearing error (signal arrival direction vs source true bearing) of up to 30 deg for a 206 Hz source at broadside at a range of 40 km. For this and many subsequent runs, observations of a bearing bias as well as multiple horizontal arrivals were taken. In addition to presentation of experimental results, we postulate here that the propagation is indeed 3D, consistent with our understanding of the wedge problem. A subsequent paper on the comparison of these data with theoretical and numerical modeling is under preparation.

This paper is organized as follows. In Sec. II a description of the calibration operations test (CALOPS) experiments is presented. The narrowband results are presented in Sec. III. In Sec. IV observations of 3D propagation with broadband signals are presented. Section V is the summary and conclusion.

[a]Electronic mail: heaney@oasislex.com

 0001-4966/2009/125(3)/1394/9/$25.00

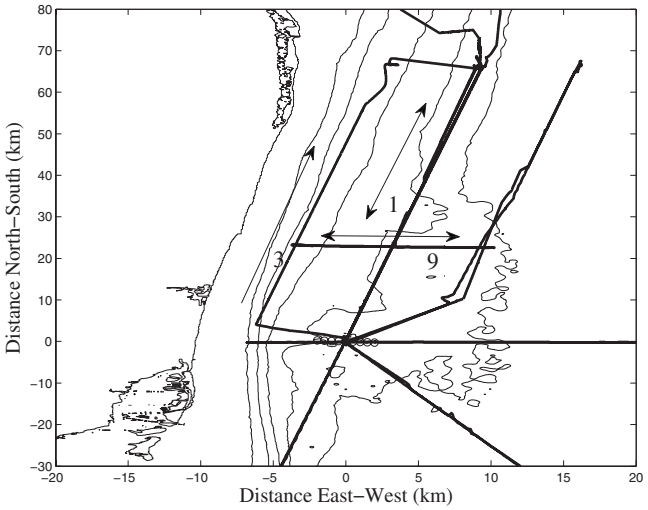


FIG. 1. Bathymetry contours and GPS ship tracks for Summer CALOPS. The run numbers analyzed in this paper are shown as is the nominal site and orientation of the array (o). The depth contours are 0, 100, 150, 200, 250, and 300 m.

II. ACOUSTIC OBSERVATORY ARRAY AND CALOPS EXPERIMENTS

In the summer of 2007 a horizontal line array, cabled to a shore recording facility, was deployed on the bottom by the Office of Naval Research 20 km off the east coast of Florida. The array concept was developed in the 1990s over the course of deliberations by the Acoustic Observatory Working Group, which followed up on the idea of a permanent acoustic observatory installation first put forward by Munk and Wunsch.[19] The array contained 120 elements, with $\frac{1}{2}$ wavelength spacing at 450 Hz (1.75 m spacing). The data were lowpass-filtered, digitized (at 1 kHz), and cabled to shore for real time processing and storage. The array was oriented downslope, providing a broadside look along the 250 m isobath. This resulted in an orientation of 8 deg relative to true north. The slope of the seafloor from 100 to 250 m is approximately 1 deg. As is evident in the bathymetry contour map shown in Fig. 1, there is a relatively flat shelf at 250 m near the array site. Two separate weeklong tests were conducted to calibrate the acoustic propagation environment—in the late summer (September 7–15) of 2007 and in the winter (February 19–25) of 2008. The tests will be referred to as CALOPS-S and CALOPS-W for summer and winter.

A. Ship tracks and signals

Bathymetry measurements, which are crucial to the understanding of propagation in this environment, were compiled by Bill Baxley (South Florida Test Facility) and include a National Oceanic and Atmospheric Administration multibeam data set. The bathymetry of the region is shown in Fig. 1, with global positioning system (GPS) ship track positions from the CALOPS-S overlaid. The CALOPS tests were designed to calibrate the acoustic propagation environment. Primary emphasis was given to transmission loss (TL) and impulse response as a function of range, frequency, bearing (and therefore bathymetry), and source depth. Characterization of the propagation environment involves various measures of coherence as well as the development of a suitable geo-acoustic parametrization. With this in mind, a series of transects, generally along isobaths, was constructed to the north, south, in-shore, and offshore of the array. In addition to these source tows, a pair of stationary transmissions was performed at ranges of 10 and 80 km in the summer and 10, 20, and 40 km in the winter. The winter towed source geometries were designed to repeat the summer tracks.

Due to the strong northward currents of the region (associated with the Gulf stream), all northbound runs were conducted with a source depth of 100 m. Almost all southbound runs had a source depth of 20 m to approximate the sound produced by an interfering ship. For all of the source tows conducted in CALOPS-S, a 2-min repeating sequence was transmitted. This included a 60-s cw comb with triplets centered at 24, 52.5, 106, 206, and 415 Hz. The source power at each peak was estimated, using a calibrated monitor hydrophone, to be 170, 171, 171, 171.6, and 171 dB/μPa at 1 m, respectively. A lower signal (10 dB down) was transmitted at ± 8 Hz from each center tone. A 30-s multi-band set of five linear frequency modulations (LFMs) was then transmitted. The signal duration was 6 s and the swept bands were 20–50, 50–100, 120–180, 200–300, and 320–420 Hz. The final signal (following an 8-s silence) was a wideband up-down LFM with a 7-s upsweep from 20 to 420 Hz, followed by 8 s of silence and then a 7-s downsweep from 420 to 20 Hz. For the CALOPS-W test this 2-min sequence was used for all southbound transits (with the source at 20 m depth) and the cw tone set was transmitted continuously for the northbound events. Two stationary transmissions were conducted at ranges of 10 and 80 km during CALOPS-S. During the stationary events, a 3-h sequence of signals was used. The cw comb was transmitted continuously for the first hour. A set of three maximal length sequences (m-sequence) was transmitted for the second hour. The wideband up/down LFM was transmitted every 30 s for the final hour.

B. Environmental measurements

The sound speed field was measured extensively during both tests with both expendable bathythermographs and conductivity-temperature-depth (CTD) observations. In this region the water column is strongly downward refracting due to surface heating and the advection of Antarctic intermediate water below the Gulf stream. The average sound speed field for each experiment taken at the array site is shown in Fig. 2. The primary difference between summer and winter conditions is the significantly faster surface water in summer due to surface heating. Below 100 m the sound speed fields are comparable. The presence of a very strong thermocline in both seasons means that this is effectively a two-layer ocean or a bi-linear profile. From the array site toward shore, it was observed that the sound speed field could be represented by truncating the 250 m sound speed profile (SSP) at the local water depth. In deeper water the clear thermocline disappeared and the SSP approximated a linear downward refracting profile.

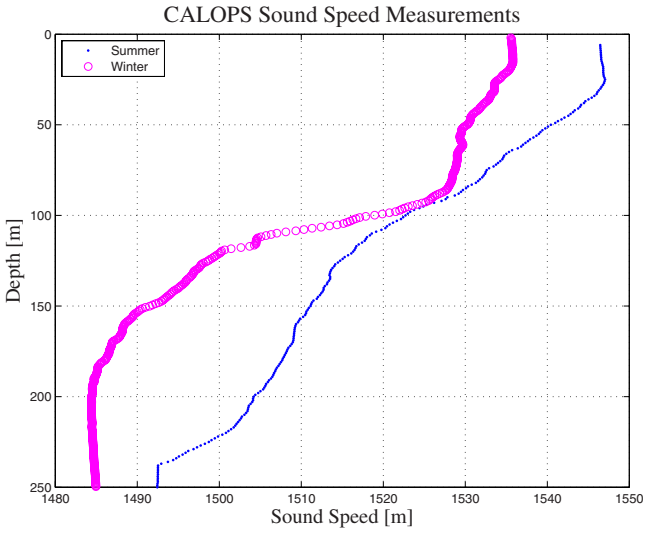


FIG. 2. (Color online) Summer and winter CALOPS sound speed measurements (from CTD casts) taken near the array location.

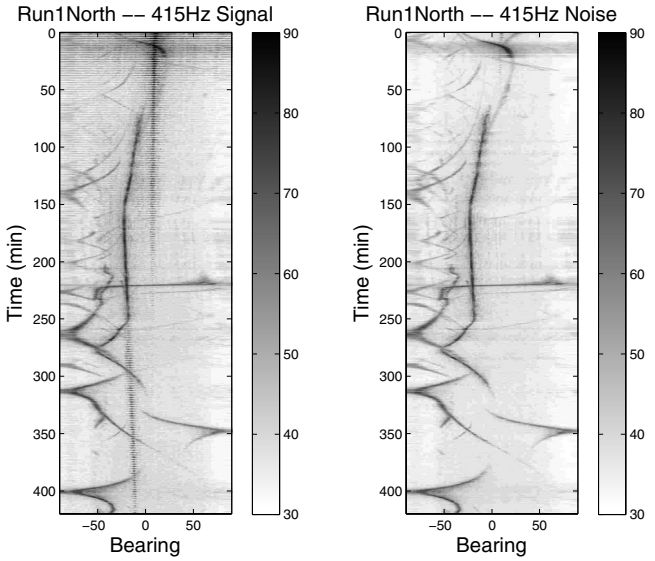


FIG. 3. BTR for CALOPS-S run 1N: (left) signal bin (415 Hz) and (right) noise bin (410 Hz). The source signal is visible at 8 deg for times 0–220 min and then appears significantly inshore at 250 min.

III. NARROWBAND EXPERIMENTAL RESULTS

In this section we present the beamforming results obtained by processing the narrowband signals. Standard techniques for developing narrowband bearing-time-record (BTR) analysis of the data were performed. The first step in data processing is a fast Fourier transform (FFT) of each hydrophone time series to transform to the frequency domain. An 8192 FFT using a Kaiser–Bessel function window (β=2.5) was applied with a 50% overlap, producing a complex spectrum output every 4.096 s. The peak bin in the signal band was determined and considered the signal bin. In order to develop an estimate of the non-signal energy (noise), an average of four edge bins (±5 Hz from the signal) was recorded as well. Conventional beamforming was performed on the complex spectra data vector using a Hanning shaded (H) plane wave replica given by

$$w = \exp(i\vec{k}\cdot\vec{r})/\sqrt{N},$$

$$B(\theta,\omega) = w'(\theta,\omega)H'xx'Hw(\theta,\omega),$$

$$(Hw)'d = 1 \qquad (1)$$

where N is the number of elements, k is the horizontal wavenumber (ω/c) in the look direction θ, r is the array element location, and B is the output beam power for a data snapshot x. The beamformer is normalized to provide a unit beam power for a unit signal vector ($d'd$=1). Beamforming therefore conserves the signal energy and reduces the noise level (DI=10 log N for N elements in uncorrelated noise). This normalization is required to compute accurate TL levels. Near field effects (beam smearing) are expected to occur but are not considered significant in the identification of horizontal multi-path for source ranges beyond 10 km. A range-focused replica is a necessity if the exact value of the beam response is required.

A note on terminology is required. For all of the acoustic runs below we will observe two distinct arrivals, with different propagation paths and arrival angles at the receiver. In general we will refer to two arrivals, one coming near the GPS bearing of the ship and the other from an angle inshore of the ship bearing. The nomenclature for these two paths is not standard and can be confusing. We will follow an analogy of simple ray propagation near a source where there is a direct path and a surface reflected (or if in deep water a refracted) path. The arrival coming in near the source ship bearing we will label the "direct path." This path does contain multipath propagation in the vertical. Other options for labeling this arrival are "in-plane path," "2D path," and "GPS ship-bearing path," though we prefer direct path. The path identified as having propagated up the slope and back to the receiver will be labeled the "inshore path." This path has traveled a greater distance than the direct path and has refracted due to multiple bottom interactions to arrive at the receiver later and from an inshore angle. Other options for this path label are "3D path" or "horizontally refracted path."

A. CALOPS-S run 1 north

For CALOPS-S run 1N, the source was deployed at 100 m and towed from the array site north along the 250 m isobath. This is along a bearing of 8 deg. The narrowband BTR for the signal (415 Hz corrected for Doppler) and the noise bin (~410 Hz) is presented in Fig. 3. The 415 Hz results are presented because this frequency has the highest signal-to-noise ratio (SNR) due to the fall-off of surface shipping noise levels with frequency. Plotting the signal bin and the noise bin is necessary to permit identification (through visual analysis) of energy received from the source. It is expected that the shipping noise not associated with the source transmissions will be the same at 410 and 415 Hz. The x-axis in this plot and all subsequent BTRs is degrees true (positive clockwise relative to north), as opposed to bearing relative to the array. The y-axis is time during the run. The run was conducted at a constant speed of 3 km to a range of 80 km. The 2-min on-off cw signal assists in the identification of source signal energy. The source transmission is clearly vis-

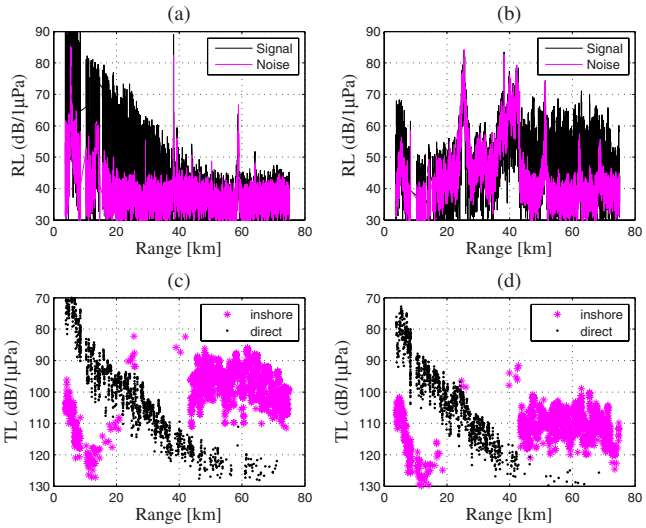


FIG. 4. (Color online) RL and TL for CALOPS-S run 1N: (a) 415 Hz RL for direct path bearing—signal (dark) and noise (light), (b) 206Hz RL for inshore bearing—signal (dark) and noise (light), (c) 206 Hz TL for direct (dark dots) and inshore (light stars), and (d) 415 Hz TL for direct (dark dots) and inshore path (light stars).

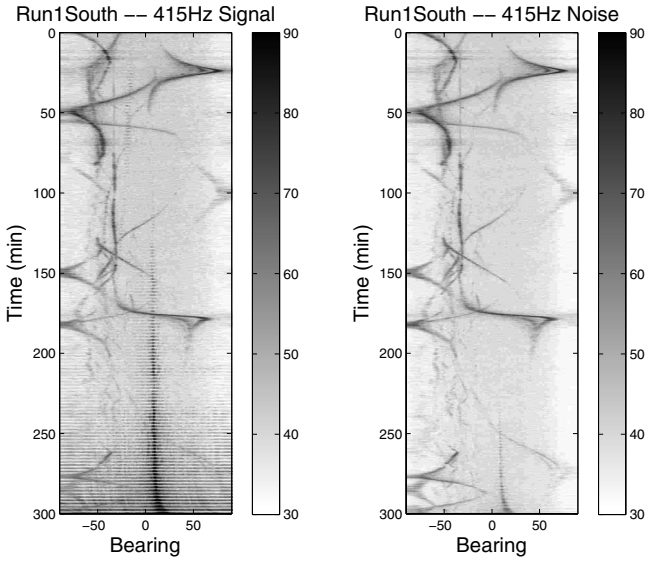


FIG. 5. BTRs for signal/noise bins for CALOPS-S run 1S at 415 Hz. The source moves toward the receiver at 8 deg T. The refracted arrival is visible, coming in from −15 deg after 250 min.

ible at the beginning of the run at the GPS bearing of 8 deg, as expected. The signal level drops off with range, disappearing at about 220 min into the run, which equates to a range of 30 km. At about 250 min, a signal appears at −30 deg that is not in the noise record. The signal is strong enough to permit classification as a transmitted signal. The 2-min sequence is visible out to 400 min, or 70 km. The bearing of this arrival changes slowly over the last 200 min from −25 to −5 deg. Initial interpretation of these results focused on the "jump" from the GPS bearing up to a range of 30 km to the "false bearing" beyond a range of 40 km. These angles are always inshore of the ship, as is expected and predicted by theory and modeling. For this particular run, the inshore arrival is blocked by an interfering ship at precisely the same bearing and time for much of the run.

The received level (RL) and TL of each of these arrivals are now examined. The narrowband RL and TL values for CALOPS-S run 1N are shown in Fig. 4. For this and subsequent level plots, the beam output time series for the signal and the noise level is plotted. The beamformer is normalized so that when the source signal is dominant, the beam output level is equal to the omni level. This has been confirmed but is not shown. The SNR is the difference between the two lines and is shown to be larger than 30 dB for ranges around 20 km. At 38 and 55 km, an interfering ship runs over the array and dominates this frequency band. Beyond 40 km, the SNR is negligible and the identification of the signal is problematic. The in-shore arrival level is generated by hand selecting the bearing and plotting the signal and noise levels for this bearing as a function of time/range. In Fig. 4(b) the signal (black) and noise (gray) are plotted as a function of range for this inshore path. At 25 and 40 km, noise from an interfering ship again masks the signal levels. Beyond 45 km, the RL of the inshore path is on the order of 60 dB and is nearly constant with range, except for a fall-off at 75 km. The 2-min transmission cycle (with the cw on for 1 min and then off for a minute) is clearly visible in this time series, though it cannot be seen in this plot due to scale of the *x*-axis.

To estimate the TL, a two-pass peak picker was used to automatically estimate arrival levels. The first pass selected arrivals with a SNR greater than 10 dB. The second pass required that the neighboring peaks were within 5 dB. This second pass is required to filter out time periods where the weaker cw signals transmitted during the LFM sequences come through. For low source ranges, even these weak signals (10 dB down from the strong cw comb) have SNR greater than 10 dB. The TLs resulting from the application of the two-pass filter, using the source levels listed in Sec. II, are presented in Figs. 4(c) and 4(d) for 206 and 415 Hz, respectively.

This analysis clearly demonstrates the arrival of inshore energy. At ranges from 40 to 70 km, the 206 Hz TL for this horizontal multipath is nearly constant with range (on the order of 88–95 dB), compared with the direct path TL of 125 dB. 3D arrivals are also observed for closer ranges (10–20 km), but are significantly lower in level than the direct path. It is unfortunate that an interfering ship passed over the bearing of the inshore path while the source ship was 40 km away because this masks the transition zone. For other runs this was not the case. To summarize, during CALOPS-S run 1 North, multiple horizontal arrivals were observed, with a bearing separation at the receiver of up to 30 deg. The inshore path dominated the direct path levels at ranges beyond 30 km.

B. CALOPS-S run 1 south

For the CALOPS-S run 1S, the source was towed at a depth of 20 m along the 250 m isobath toward the array from a distance of 60 km to the north. (Although the run began at 80 km, the shore recording facility was not fully operational until the source ship was 60 km north.) The signal and noise BTR results are plotted in Fig. 5. For this run the signal is

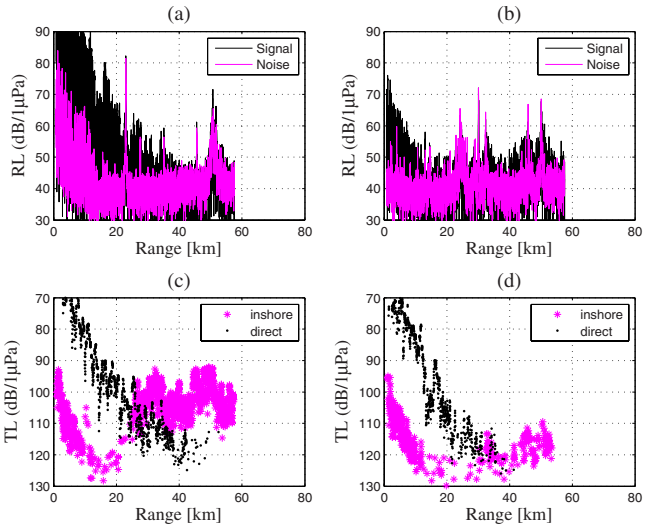


FIG. 6. (Color online) RL and TL for CALOPS-S run 1S: (a) 415 Hz RL for direct path bearing—signal (dark) and noise (light), (b) 206 Hz RL for inshore bearing—signal (dark dots) and noise (light stars), (c) 206 Hz TL for direct path (dark dots) and inshore path (light stars), and (d) 415 Hz TL for GPS (dark dots) and 3D bearing (light stars).

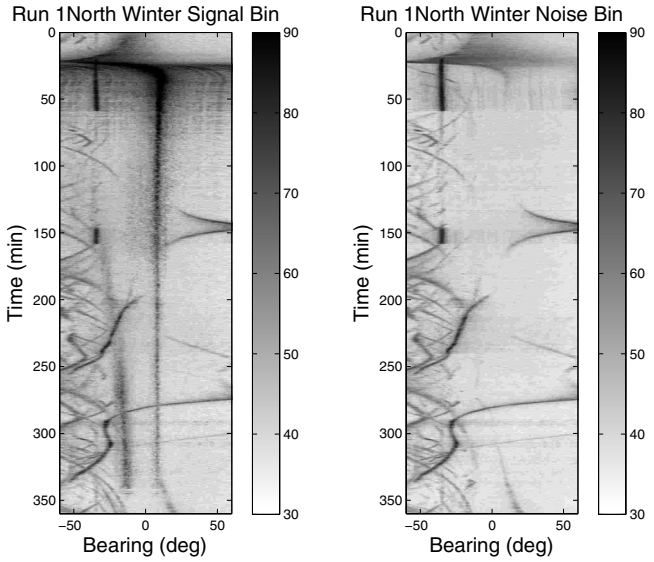


FIG. 7. BTRs for 415 Hz cw transmission during CALOPS-W run 1N—signal (left) and noise (right). The source was on continuously. The source ship towed away from the receiver along the 250 m isobath (bearing 8 deg T). The inshore signal is visible from ranges of 25–60 km.

observable along the true GPS path from the beginning of the run and becomes dominant as the source ship approaches the array. The inshore path is also visible from the beginning of the run (−5 deg) up to about 100 min (30 km), where it drops out. The angle of the inshore path moves from −5 to −30 deg as the source ship moves from 60 to 30 km. In the noise bin BTR on the right side of Fig. 5 the ship is visible as it passes over the array.

The RL and TL results for this run are shown in Fig. 6. In Fig. 6(a), the RLs at 206 Hz, the beam-noise levels along the GPS bearing indicate high SNR for ranges out to 40 km, with interfering ships passing over at 21 and 50 km. The inshore tracked bearing is weaker, with an interfering ship passing over at a source range of 25 km. The TL plots in panels (c) and (d) confirm that there is a transition zone at 30 km where both GPS and 3D signals have similar TL. For the inshore path, the 206 Hz TL is nearly uniform with range (beyond 20 km) at 95–100 dB, with the signal as much as 20 dB stronger than the direct path. The 415 Hz TL plot indicates much weaker signal beyond 20 km, as compared with 206 Hz. These results confirm not only the presence but also the dominance of 3D propagation. It is also apparent that the shallow source, which only excites higher angle propagating energy, has a weaker refracted path than the 100 m source depth.

C. CALOPS-W run 1 north

The source TL for the winter northbound runs was a continuous narrowband comb of the same frequencies as the summer test described above. The continuous narrowband comb was chosen to permit accurate TL measurements without the 1-min LFM signal dropouts. For CALOPS-W run 1N, the source was suspended at a depth of 100 m and towed from the receiver north to a range of 60 km. The results from this event are clearer than for CALOPS-S run 1N for two reasons. Winter propagation conditions are better due to a less strongly downward refracting profile (see Fig. 2). There are also no nearby dominant interfering ships when the source ship is at a range of 20–50 km from the array. From Fig. 7, we see that the 3D horizontal multipath arrivals have significant energy from ranges of 25 km out to the longest range of 60 km. We also see that the refracted path arrives at the array sweeping from bearings of approximately −40 deg (t=160 m corresponding to 30 km) to −10 deg (t=340 m corresponding to r=60 km). This reduction in the bearing difference between the true and the refracted (reflected) path is consistent with expectations as the distance between the source and receiver increases. Consider the thought experiment with a candle and a mirror, easily solved by the method of images. As the candle moves out in range, the angle between the true source and the image source reduces.

D. CALOPS-S run 3 north

For CALOPS-S run 3N, the source was towed at a depth of 100 m inshore along the 120 m isobath. The BTR for run 3N is shown in Fig. 8. For this run, the inshore path appears strongly at a time of 50 min, corresponding to a range of 10 km. The angular difference between the inshore path and the direct path is much smaller than for the run 1 geometry. The double arrival (direct and inshore path) is clear out to a time of 250 min, after which a single spread arrival is observed. Propagation conditions are significantly better in this run than in either run 1N or tun 1S, to the point where the source transmission is evident in the noise bin BTR, apparently a glint from each time the LFM passes through the frequency bins of interest. Two mechanisms proposed to explain these improved propagation conditions (reduced TL) are downslope focusing (in a downward refracted environment which deepens, energy is trapped near the bottom) and the change in acoustic interaction angles with the bottom due to the warmer water at the seafloor in shallow water.

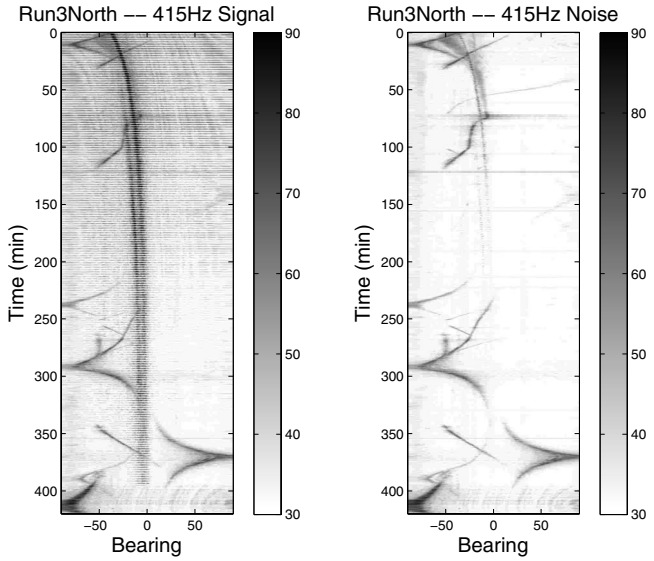


FIG. 8. BTRs for 415 Hz cw transmission during CALOPS-S run 3N—signal (left) and noise (right). The source was on continuously. The source ship towed away from the receiver along the 120 m isobath (bearing 8 deg T). The 3D signal is visible beyond a range of 10 km, or 50 min.

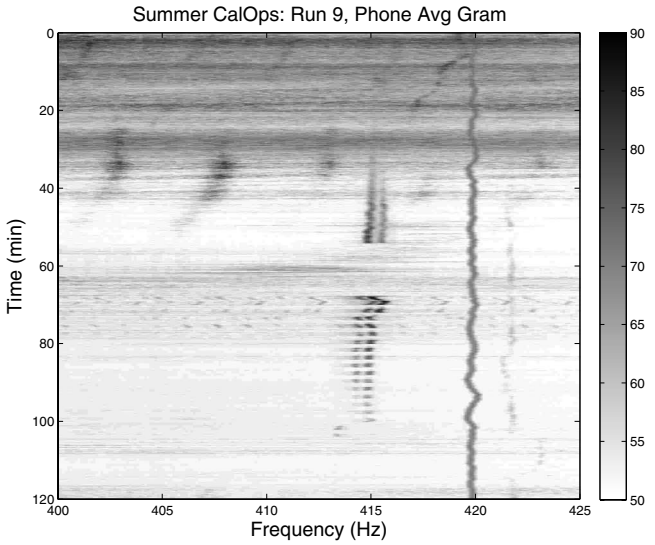


FIG. 10. Phone averaged spectrogram for CALOPS-S run 9. Note the split signals at 415 Hz. A cw tone at 415 Hz was transmitted continuously for the westbound leg (30–50 min). The 2-min sequence (1 min of cw) was transmitted for the eastbound leg (65–100 m).

The RL and TL levels for run 3N are shown in Fig. 9. Note how much higher the RLs are for this run, compared with runs 1N and 1S, when the source was towed along the deeper (250 m) isobath. The SNR for nearly the entire run is greater than 20 dB, particularly for the inshore arrival. The inshore path starts to appear at 10 km, is equivalent to the direct path levels at 25 km, and becomes much stronger out to 80 km. At some ranges, the inshore path is as much as 20 dB stronger.

E. CALOPS-S run 9

For CALOPS-S run 9, the source was towed up-shelf from deep to shallow water (9W) and then back (9E) at a range of approximately 30 km from the source. A cw narrowband comb was transmitted during the westbound leg (toward shallow water) and the standard 2-min sequence of pulses was transmitted for the subsequent eastbound leg. For this geometry, the range-rate of the source is nearly zero. With near-zero Doppler, the 415 Hz transmitted signal is expected to be in the 415 Hz bin. The search for the signal and noise bins produced an interesting result. The single element spectrogram near the 415 Hz cw signal is shown in Fig. 10. The source signal is clear at 415 Hz for both the continuous transmission (9W) and the 2-min sequence (9E). What is surprising is the presence of a Doppler-shifted signal in addition to the expected zero-Doppler signal. For the 9W leg, the additional signal is Doppler-shifted up on the order of 1 Hz, while for the 9E leg, the additional signal is shifted down a similar amount. This Doppler shift is consistent with the geometry being a refracted 3D propagation path. For this path, there is a positive range-rate for 9W and a negative range-rate for 9E.

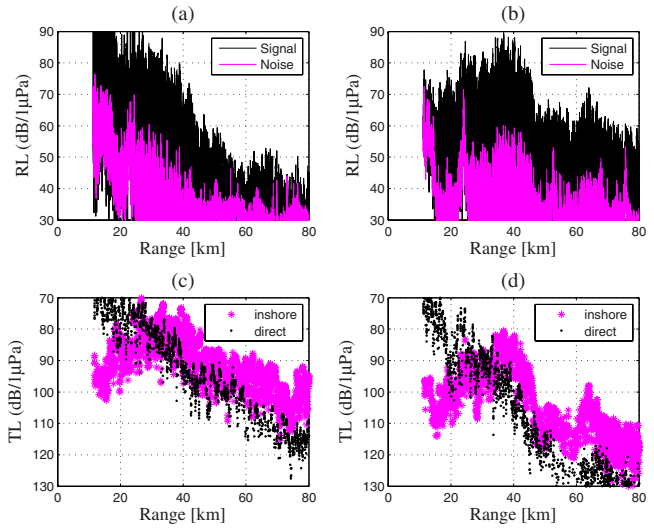


FIG. 9. (Color online) RL and TL for CALOPS-S run 3N: (a) 415 Hz RL for direct path bearing—signal (dark) and noise (light), (b) 206 Hz RL for inshore bearing—signal (dark) and noise (light), (c) 206 Hz TL for direct (dark dots) and inshore path (light stars), and (d) 415 Hz TL for direct (dark dots) and inshore path (light stars).

The BTRs for the center bin and the up-shifted Doppler bin (noise) during run 9W are shown in Fig. 11. Over this timeframe, the source ship traveled toward shore, from a bearing of 40 deg to due north of the array. The right panel of Fig. 11 is BTR in the signal bin. There is no clear double arrival, but there is energy spread over many angles. This out-of-plane scattering is significant to the problem of quiet target detection in shallow water because it means that energy from a nearby (and/or very loud) interfering ship is spread over many angles and cannot be canceled by standard adaptive beamforming algorithms. We refer to this energy as out-of-plane 3D scattering. It is evident that there would be very strong bearing-dependent reverberation for active antisubmarine warfare in this wedge environment. The left panel of Fig. 11 shows the BTR of the Doppler-shifted signal bin (noise bin). This energy arriving at −25 deg is identifiable with the source in that it ceases 50 min into the run, coincident with the termination of source transmissions. The in-

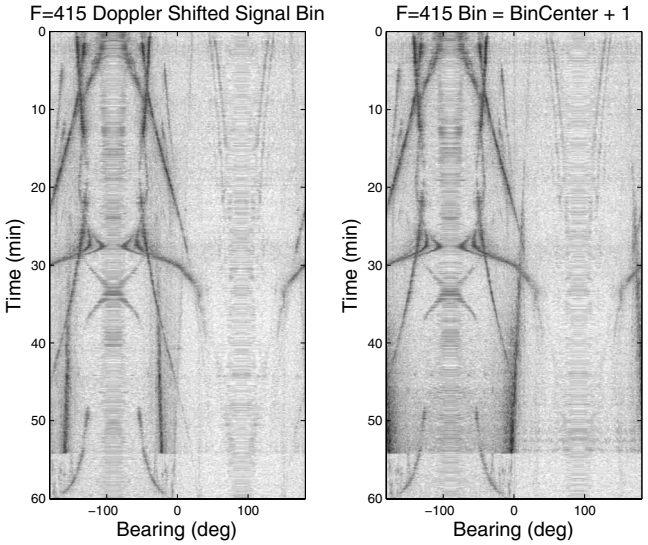


FIG. 11. BTR for CALOPS-S run 9W. (Left) BTR of positive Doppler shifted energy. There is source energy seen coming from inshore due to 3D propagation. (Right) BTR for the zero-Doppler source showing the move from 40 to 0 deg. Scattering to multiple angles near the source (reverberation) is visible.

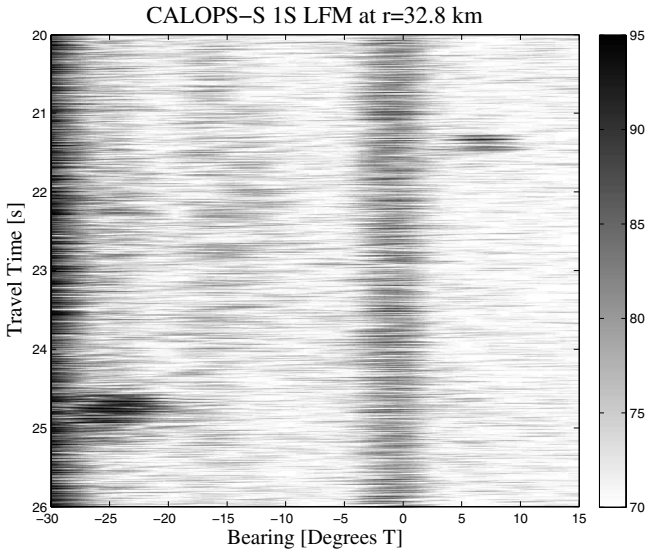


FIG. 12. Beam matched filter results for CALOPS-S run 1S for LFM signals transmitted when the source was at 32.8 km north of the array along the 250 m isobath.

shore 3D path appears just after 20 min and by 40 min is cleaner than and as strong as the direct propagation path. The positive Doppler (despite a near-zero range rate in this geometry) is an indication that the sound traveled inshore, refracting from the sloping bathymetry, then propagated away from shore to arrive at the array at an angle of −25 deg. For this horizontal path, the source appears to be moving toward the ship, giving it a positive Doppler.

IV. BROADBAND EXPERIMENTAL RESULTS

A. CALOPS-S run 1 LFM results

The reception of broadband transmissions on a horizontal line array permits the estimation of travel time of each horizontal arrival. To estimate the arrival time of energy from a particular direction, matched filtering of a transmitted LFM signal is applied to the beam time-series, which is formed by conducting a Fourier transform then narrowband beamforming and finally an inverse Fourier transform (IFT). The hydrophone complex spectrum was calculated by performing an unwindowed 1-s FFT with 50% overlap. Beamforming was applied to each frequency between 20 and 450 Hz according to Eq. (1). The IFT was then performed yielding a time series for each arrival angle. Each beam time series was then matched filtered using the 30-s upsweep LFM signal replica. In order to compute absolute travel time, a monitor hydrophone was used to record the signal transmitted into the water. These data were recorded in stereo with the IRIG-B channel GPS time recorded on the other channel. Matched-filter processing of the monitor hydrophone permitted the determination of the exact time the source LFM transmission began. The data at the recording facility are also time-synced to GPS.

The beam time series for CALOPS-S run 1S, when the source ship was at a range of 32.8 km from the receiver, is presented in Fig. 12. The multi-path arrival for the direct path arrives at an angle of 7 deg at a time of 21.2 s. The inshore 3D arrival is clearly visible at −24 deg arriving at 24.8 s. The energy at 0 deg is the common-mode electronic noise. The energy at 30 deg is an interfering ship, as can be seen in the BTR for CALOPS-S run 1S in Fig. 5.

To compare the arrival time and relative intensities of the direct path and the inshore path, line-cuts at the signal maximum bearings (−24 and 7 deg) are taken. Note that the arrival angle of the direct path is not exactly equal to the GPS ship bearing of 8 deg T. Time cuts along these two arrival bearings are shown in Fig. 13. The inshore path produces an arrival approximately 3.2 s later than the direct path. The peak level is 7 dB higher, and it is clear that there is significantly more energy in the later arriving inshore path. This is consistent with the 25 dB higher received narrowband level for 206 Hz for run 1S.

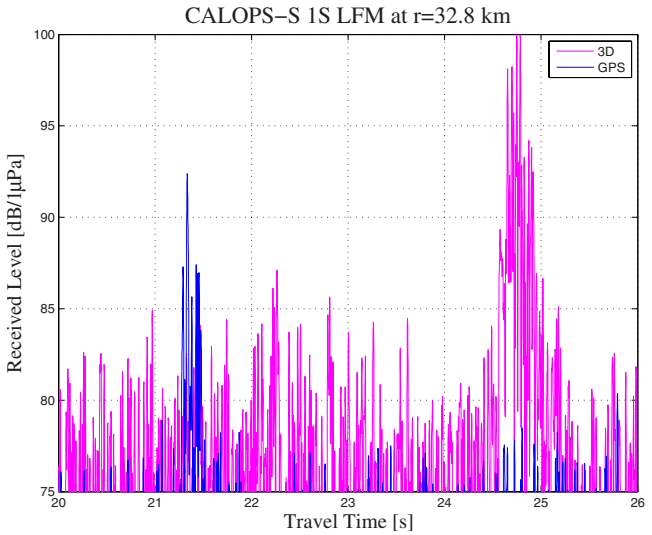


FIG. 13. (Color online) Horizontal multipath arrivals (−24 and 7 deg T) for a 20–420 Hz LFM transmitted from a source 40 km along a bearing of 8 deg T. The inshore path arrives 3.2 s later than the direct path bearing energy and is up to 10 dB stronger (in peak level, there are more peaks).

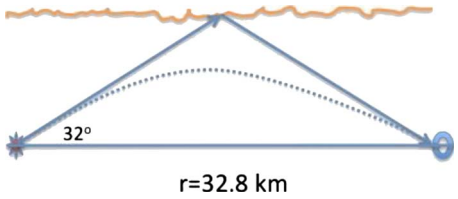


FIG. 14. (Color online) Geometry for two hypothesized propagation paths to explain the observed inshore arrivals. The solid line is a reflection from a bathymetric feature such as a coral wall. The curved path is the refracted path due to many bottom interactions.

B. Refraction or reflection?

One hypothesis for the 3D inshore path results is the specular reflection of acoustic energy from an inshore bathymetric feature, such as a coral wall. The standard 3D interpretation is for continuous refraction of the energy as a result of repeated specular interactions with the sloping seafloor. We have the measurements needed to differentiate between these two effects. In particular, we have the source-receiver range, the arrival angle of both the direct path and the inshore path, and the travel-time difference between them. The geometry of the two postulated paths is shown in Fig. 14.

The distance from the source/receiver to the scattering wall (x) is the range ($R=32.8$ km) divided by 2 tan θ, where θ is the arrival angle at the array (32 deg). This leads to a total path length of 38.67 km, which is 5.88 km longer than the observed source-receiver separation. Using the measured group velocity of the direct path (1520 m/s), this would lead to a predicted travel time of 3.86 s, whereas the observed travel-time difference was 3.2 s. This gives a strong indication that the path-length separation must be less than the straight-line path from the source to the scattering wall to the receiver, as is the case for the continuously refracted path. The group velocity assumption that the inshore path has a group velocity (vg) of 1520 m/s is certainly incorrect, but the required 20% reduction in the group velocity is larger than is expected in shallow water, particularly since the net group velocity is the reciprocal of the average group slowness ($1/vg$).

V. SUMMARY AND CONCLUSION

A pair of nearly identical acoustic propagation experiments were conducted on the continental shelf in the summer of 2007 and winter of 2008. A set of acoustic transmissions, including narrowband cw tones and broadband LFM, was transmitted from a towed source and recorded on a bottom-mounted horizontal line array. This permitted the spatial and temporal resolution of the received acoustic signals. The presence of multiple horizontal arrivals, sometime separated by up to 30 deg in azimuthal arrival angle, was observed. For summer and winter conditions, multiple source depths, and multiple source bearings, observations of multiple horizontal arrivals were recorded. For many of the receptions beyond 20 km, the inshore path was substantially stronger than the direct path along the measured bearing from the receiver to the ship. A rather odd feature of the measurements is the presence of a uniform RL vs range for paths from 40 to 80 km. This apparent lack of loss is inconsistent with the rapid falloff of received energy in the first 40 km and is indicative of some kind of horizontal focusing.

Most source tows were either toward the receiver or away, producing nearly similar range-rates for the direct path and the inshore path. This permitted the observation of the ship GPS path and the inshore 3D path from BTR plots processed at the signal frequency bin. It also means that both acoustic paths coherently interacted with each other, presumably producing significant effects on measures of spatial and temporal coherence. For one run (CALOPS-S run 9E-W), however, the source was towed cross-slope with very little range closure. For this reception the direct path has near-zero Doppler. It was observed that the inshore path has an up-Doppler shift when the source ship moves toward shore and a down-Doppler shift when the source ship moves toward deep water. This is explained by the geometry of refraction in a wedge environment. This novel result means that multiple frequency lines can be observed in a 3D environment when only a single line is emitted. Broadband processing of a LFM, transmitted when the source was at a range of 32 km, indicated the clear separation of the two paths, consistent with horizontal refraction due to the sloping bathymetry.

The observations of 3D propagation presented in this paper should not be surprising since they are consistent with theoretical and numerical modeling results dating back to the late 1960s. What is surprising is the strength of the inshore energy. For most environments, propagation in shallower water involves more frequent bottom interaction and therefore the shallow water refracted paths are expected to be weaker than the along-isobath path. This is clearly not the case in this environment. These results have important implications relevant to the accuracy of bearing-based localization such as target motion analysis in the continental shelf environment. They also imply that interfering ships may be expected to impact multiple bearings. Data from these tests should provide a useful experimental database to validate 3D propagation models in the future.

ACKNOWLEDGMENT

This work was funded by NAVSEA IWS-5 and the Office of Naval Research. The authors would like to acknowledge Theo Kooij, Judith Bishop, Jeff Vuono, Phil Dinalpho, Bill Baxley, Harry Cox, Rich Gramman, Kent Scarborough, and the captain and crew of the R/V Walton Smith and R/V Seward Johnson. The bathymetry plotting code was developed by Richard Campbell at Portland State University.

[1]A. D. Pierce, "Extension of the method of normal modes to sound propagation in an almost stratified medium," J. Acoust. Soc. Am. **37**, 19–27 (1965).
[2]D. L. Bradley, PhD thesis, The Catholic University of America, 1970.
[3]M. J. Buckingham, "Theory of three-dimensional acoustic propagation in a wedgelike ocean with a penetrable bottom," J. Acoust. Soc. Am. **82**, 198–210 (1987).
[4]C. H. Harrison, "Three-dimensional ray paths in basins, troughs, and near seamounts by use of rays," J. Acoust. Soc. Am. **62**, 1382–1388 (1977).
[5]C. H. Harrison, "Acoustic shadow zones in the horizontal plane," J. Acoust. Soc. Am. **65**, 56–61 (1979).
[6]C. H. Harrison, "Wave solutions in three-dimensional ocean environments," J. Acoust. Soc. Am. **93**, 1826–1840 (1993).

[7]H. Weinberg and R. Burridge, "Horizontal ray theory for ocean acoustics," J. Acoust. Soc. Am. **55**, 63–79 (1974).
[8]M. D. Collins, "The adiabatic mode parabolic equation," J. Acoust. Soc. Am. **94**, 2269–2278 (1993).
[9]M. D. Collins, B. E. McDonald, K. D. Heaney, and W. A. Kuperman, "Three-dimensional effects in global acoustics," J. Acoust. Soc. Am. **97**, 1567–1575 (1995).
[10]A. T. Abawi and W. A. Kuperman, "The coupled mode parabolic equation," J. Acoust. Soc. Am. **102**, 233–238 (1997).
[11]F. Sturm and J. A. Fawcett, "On the use of higher-order azimuthal schemes in 3-D PE modeling," J. Acoust. Soc. Am. **113**, 3134–3145 (2003).
[12]G. B. Deane and M. J. Buckingham, "Two approximate solutions for the 3-D field in the ASA benchmark wedge," J. Acoust. Soc. Am. **109**, 2332 (2001).
[13]C. T. Tindle and Z. Y. Zhang, "An adiabatic normal mode solution for the benchmark wedge," J. Acoust. Soc. Am. **101**, 606–609 (1997).
[14]F. Sturm, "Numerical study of broadband sound pulse propagation in three-dimensional oceanic waveguides," J. Acoust. Soc. Am. **117**, 1058–1079 (2005).
[15]P. A. Baxley and H. Bucker, "The effect of three-dimensional bathymetry on array processor performance," J. Acoust. Soc. Am. **100**, 2852–2853 (1996).
[16]K. B. Smith, C. W. Miller, A. F. D'Agostino, B. Sperry, J. H. Miller, and G. R. Potty, "Three-dimensional propagation effects near the mid-Atlantic Bight shelf break (L)," J. Acoust. Soc. Am. **112**, 373–377 (2002).
[17]S. D. Frank, M. Badiey, J. F. Lynch, and W. L. Siegmann, "Experimental evidence of three-dimensional acoustic propagation caused by nonlinear internal waves," J. Acoust. Soc. Am. **118**, 723–734 (2005).
[18]K. D. Heaney and A. B. Baggeroer, presented at the OCEANS'06 Asia Pacific, Singapore, 2006.
[19]W. Munk and C. Wunsch, "Ocean acoustic tomography: A scheme for large scale monitoring," Deep-Sea Res., Part A **26**, 123–161 (1979).

Array element localization using ship noise

Michael G. Morley,[a] Stan E. Dosso, and N. Ross Chapman
School of Earth and Ocean Sciences, University of Victoria, Victoria, British Columbia V8W 3P6, Canada

(Received 11 August 2008; revised 5 January 2009; accepted 8 January 2009)

This paper describes a method of estimating hydrophone positions in a receiver array using the noise from a passing ship. Relative arrival times of the ship-noise signal between pairs of hydrophones are obtained from several time windows of data (corresponding to different ship locations) by cross-correlating the band-pass filtered time series. The relative arrival times are used as data in an array element localization inversion to estimate both the hydrophone and ship locations based on iterated linearization of the acoustic ray equations. The inversion applies the method of regularization to include prior information such as approximate location estimates and uncertainties for the source and receivers and the expectation that the array shape and/or source tracks are smooth functions of position. Linearized and nonlinear (Monte Carlo) estimates of the position errors are in good agreement and indicate a high degree of confidence in the receiver positions (relative uncertainties of approximately 0.2 m in the horizontal and 0.05–0.1 m in the vertical). The ability to improve upon the initial source position estimates depends on the geometry of the problem, as investigated with simulations. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3076201]

PACS number(s): 43.30.Xm, 43.60.Pt [AIT] Pages: 1403–1409

I. INTRODUCTION

Ocean acoustic array processing methods, such as beamforming[1] and matched-field processing and inversion,[2,3] require accurate knowledge of the locations of individual elements in a receiver array. However, sufficiently accurate receiver locations are generally not known after array deployment at sea, and array element localization (AEL) surveys are often required.[4] AEL is usually based on inverting acoustic arrival-time measurements from a series of controlled sources at (approximately) known positions to the receivers to be localized.[4–14] However, the uncertainties in source positions can be an important source of error in AEL.[5,7] A regularized ray-based AEL inversion, which treats both source and receiver locations as unknown and determines the simplest (smoothest) array shape subject to fitting the acoustic data, has been developed and applied to a variety of source/receiver geometries, such as bottom-moored horizontal arrays[7,14] and vertical arrays,[8,14] towed arrays,[9,10] two-dimensional arrays,[13] and freely-drifting sonobuoy fields.[11,12]

AEL is usually carried out using impulsive sources, which produce identifiable acoustic arrivals from which arrival times can be measured. The continuous nature of broadband noise sources such as a ship makes it difficult to directly observe coherent arrivals across an array. However, relative arrival times can be extracted from recordings of a noise source by cross-correlating the time series at spatially separated hydrophone pairs. This paper shows that estimating relative arrival times from ship noise at several positions can provide a sufficient data set for high-precision AEL.

Cross-correlation techniques have been used in ocean acoustic applications such as ship tracking,[15] marine mammal tracking,[16] and geoacoustic inversion.[17] More recently, cross-correlation of ocean ambient noise has been used to estimate Green's functions between bottom-moored hydrophones for positioning and time synchronization.[18] Array-shape estimation has also been performed by a matched-field approach using the noise from ships of opportunity.[19] However, it does not appear that cross-correlation of ship noise has been applied previously to estimate individual receiver and source positions in an AEL survey. An interesting aspect of using a continuously-moving source such as a ship (as opposed to a series of impulsive sources at a grid of locations) is that the smoothness regularization can be applied to the ship tracks (as well as array shape) to minimize unrealistic zig-zags or jumps between adjacent source positions, which could be introduced by inaccurate prior information and noisy acoustic data.

Motivation for the present work comes from an AEL survey of a bottom-moored vertical line array (VLA) in which controlled-source recordings intended for AEL proved unusable (described in Sec. II), and the use of relative arrival times extracted from broadband ship noise was the only recourse to localize the array. However, there are a number of potential advantages of ship-noise AEL, including reductions in cost, effort, and environmental impact of not using impulsive sources; efficient coverage of wide areas (particularly if the nominal array position is poorly known); and the ability to carry out AEL covertly, using either a known ship or a ship of opportunity.

II. EXPERIMENT AND DATA

In October 2003, a scientific cruise to the Mississippi Canyon, lease block 798 (MC798) in the northern Gulf of Mexico, was undertaken to collect acoustic data for seabed gas-hydrate research. A bottom-moored VLA consisting of 16 hydrophones spaced at approximately 12.5-m intervals was deployed from the *R/V Pelican* and allowed to free-fall

[a]Present address: MGM Geosciences, 420-900 Tolmie Avenue, Victoria, British Columbia V8X 3W6, Canada.

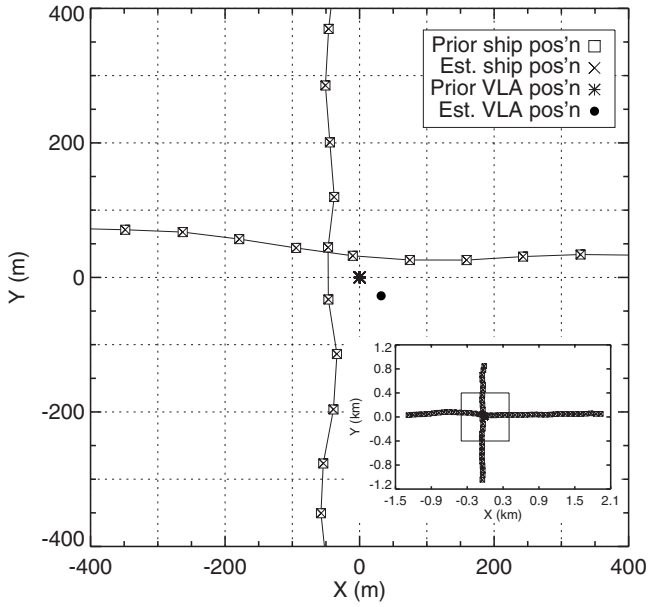


FIG. 1. Plan view showing initial (prior) and estimated (est.) ship and VLA locations relative to the deployment position. Inset shows ship positions along both track lines of *R/V Pelican*.

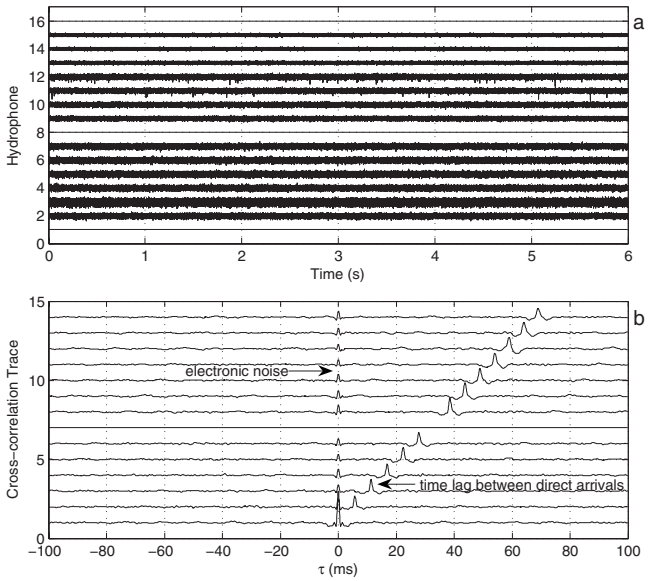


FIG. 3. (a) Raw ship-noise time series (hydrophones 1, 8, and 16 zeroed due to poor data). (b) Results of cross-correlation processing with hydrophone 2 (hydrophone numbers increment from the bottom of the array).

to the seafloor in ~800 m of water. Concrete weights anchored the array to the seafloor, and its shape was maintained by means of glass floats at the top of the array. 6-s samples of acoustic data were recorded every 18 s at a 10-kHz sampling rate as the ship transited along two radial track lines centered near the nominal array position (Fig. 1). Data were recorded first while towing an impulsive source (water gun) and then with no sound source except the survey ship. The water-column sound speed profile was measured prior to running the track lines and is shown in Fig. 2. Due to badly-clipped direct arrivals and large correlated timing errors, the water-gun data were unusable for AEL inversion; therefore, a method was devised to use the ship-noise recordings for this purpose.

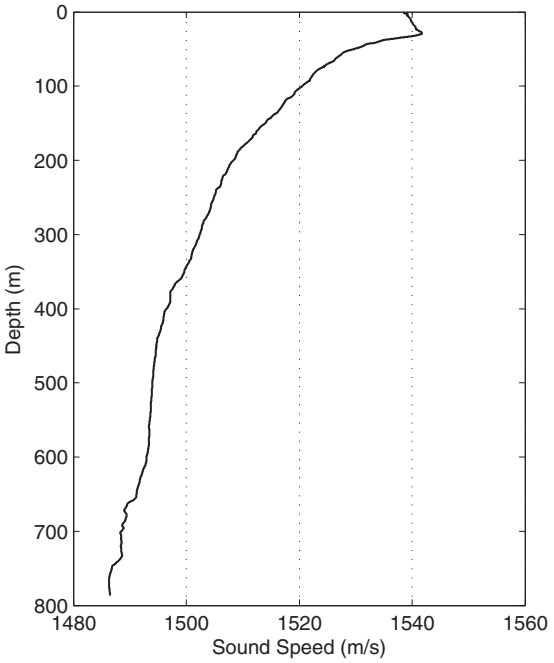


FIG. 2. Water-column sound speed profile measured near VLA position.

An example of the acoustic time series recorded at the VLA as the ship transited along line 1 (north-south track line) is shown in Fig. 3(a) (data for hydrophones 1, 8, and 16 were of poor quality due to seawater leakage and are not considered here). Direct observation of coherent arrivals on the noise time series is difficult or impossible. However, the time differences of energy arriving along the array can be extracted from the cross-correlation of signals recorded at pairs of receivers. The cross-correlation between signals $s_i(t)$ and $s_j(t)$ at hydrophones i and j over T samples is

$$R_{i,j}(\tau) = \sum_{t=0}^{T-1-\tau} s_i(t)s_j(t+\tau), \tag{1}$$

where τ is the lag time (in samples). The lag at the cross-correlation maximum corresponds to the best estimate of the time delay of coherent acoustic energy between the receivers,

$$\Delta t_{i,j} = \arg_{\max}[R_{i,j}(\tau)]. \tag{2}$$

Figure 3(b) shows cross-correlation functions for hydrophone pairs from Fig. 3(a). Trace 1 represents the autocorrelation of the data received at hydrophone 2 (bottom of the array), and traces 2–14 are the cross-correlations between hydrophone 2 and the other hydrophones on the array. The cross-correlation functions are normalized such that the autocorrelation at zero lag is unity. The maximum correlation occurs when the time delay is equal to the difference between arrival times of the ship noise at the two hydrophones. This time delay increases with separation between the hydrophones in Fig. 3(b); the cross-correlation peaks at $\tau=0$ are an artifact due to electrical noise common to all the receivers.

The presence of signals in the data that are not associated with the survey ship can degrade the correlation processing. Prior to computing cross-correlations, the data were

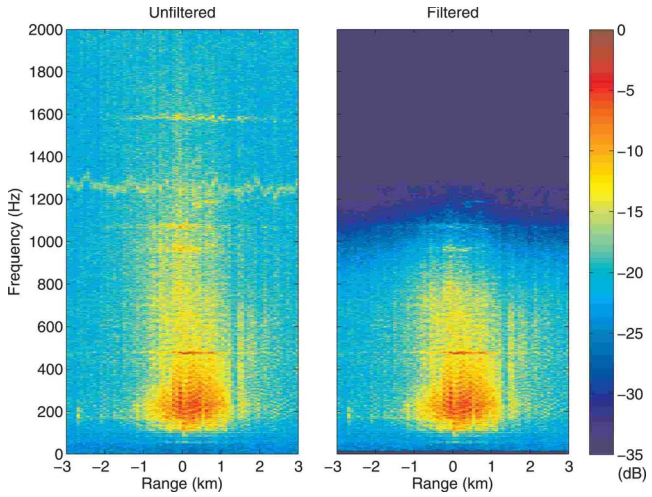


FIG. 4. (Color online) Spectrograms of unfiltered (left) and filtered (right) ship noise for line 2. Spectra are normalized to maxima at 0 dB.

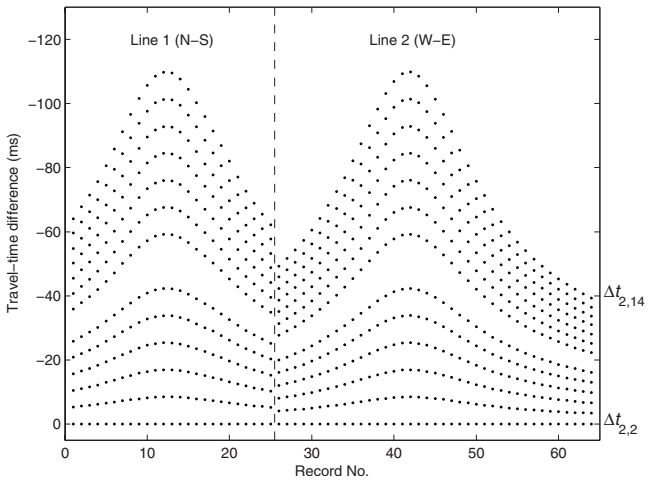


FIG. 6. Relative travel-time picks from cross-correlation for all hydrophones and both ship lines.

band-pass filtered to exclude signals outside the band containing the majority of the ship-noise power. Normalized spectrograms for the acoustic data recorded at hydrophone 2 along line 2 are shown in Fig. 4. In the raw-data spectrogram (left panel), a tonal band near 1250 Hz is not associated with the ship since the signal strength does not diminish with increasing range from the nominal array position. After some experimentation, a filter with a pass band of 20–600 Hz and a 3-dB/octave roll-off was found to yield a good balance between peak resolution and noise cancellation in the cross-correlation functions. The spectrogram of the filtered data is also shown in Fig. 4 (right panel).

A raster image showing the cross-correlation values for hydrophones 2 and 6 for all recordings along line 1 is shown in Fig. 5. The time delay of the direct arrival is readily identified as a dark band in the image that moves out hyperbolically with range from the nominal array position. Faint hyperbolic bands observed at larger time delays in the image are likely associated with water-column multiples of the ship noise; however, the order of these multiples is not obvious. If

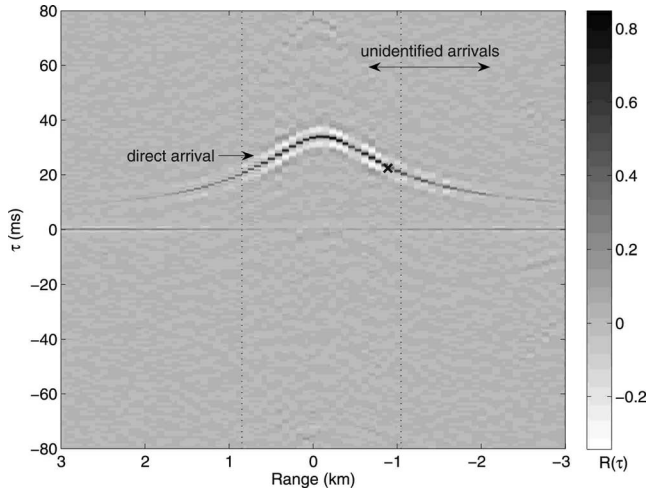


FIG. 5. Cross-correlation values for hydrophones 2 and 6 and all ship-noise recordings of line 1. Dotted lines indicate the data used in the AEL inversion. The **x** indicates the data sample in Fig. 3.

the paths of these arrivals were identified, they could potentially be included in the inversion to provide additional information in determining the sensor depths. However, the multi-path arrivals would add little information on the horizontal positions of the sensors since they lie in the same vertical plane as the direct arrival (considered in previous AEL work).[8,14] In this case the arrivals were too weak to pick reliably and are not considered.

Using hydrophone 2 as the reference, cross-correlation functions were computed with the usable hydrophones of the array for the two orthogonal ship tracks shown in Fig. 1. Time delays corresponding to the cross-correlation maxima were picked for 13 receivers and 64 source positions over the two lines (25 from line 1, 39 from line 2) for a total of 832 data, as shown in Fig. 6 (since the reference hydrophone is at the bottom of the array, the travel-time differences are negative). Based on examining the cross-correlation plots at high-magnification, the uncertainty in picking the correlation peaks was estimated to be 0.2 ms. This uncertainty did not appear to vary across the array.

Nominal (prior) estimates of the location of the array elements in x (east-west) and y (north-south) were obtained from the global positioning system (GPS) position of the ship when the array was deployed; the (vertical) z coordinate estimates were based on the chart water depth at this location and the nominal hydrophone heights above the array anchor. The uncertainties of the hydrophone-position estimates were conservatively taken to be 1000 m in x and y and 100 m in z (these large values allow the data information to dominate the solution for receiver localization). The position of the ship at the start of each 6-s recording was extracted from the navigation data and corresponds to the location of the GPS antenna near the fore-aft center of the vessel. Traveling at about 5 m/s, the ship moved ~30 m over a 6-s recording and ~90 m between records. The ship position at the center point of each recording was determined by interpolating the GPS positions, assuming that the ship traveled in a straight line at a constant speed. Based on this procedure and the GPS accuracy, the uncertainties in the x and y coordinates of

the prior source positions were estimated to be 15 m. The z coordinates for all source positions were estimated to be 3 m with an uncertainty of 3 m based on the draft of the survey vessel.

III. INVERSION METHOD

Regularized AEL inversion is briefly summarized in this section; more complete descriptions are found elsewhere.[7,13,14] The AEL inverse problem considered here involves estimating three-dimensional position variables (x,y,z) for N_r VLA receivers and N_s source (ship) locations via linearized inversion of the acoustic ray equations,

$$r = \int_{z_s}^{z_r} \frac{pc(z)dz}{[1-p^2c^2(z)]^{1/2}}, \tag{3}$$

$$t = t_0 + \int_{z_s}^{z_r} \frac{dz}{c(z)[1-p^2c^2(z)]^{1/2}}, \tag{4}$$

where r represents the source-receiver range, t is the arrival time, t_0 is the source transmission instant, and $c(z)$ is the measured sound speed profile. In Eqs. (3) and (4), $p=\cos\theta(z)/c(z)$ [where $\theta(z)$ is the grazing angle] is the ray invariant, which defines the take-off angle at the source. The eigenray for a direct path from source to receiver is found by determining the p value that gives the correct range according to Eq. (3); this can be determined efficiently using Newton's method.[7,14]

The measured arrival times $\mathbf{t}$ can be written in vector form as

$$\mathbf{t} = \mathbf{t}(\mathbf{m}) + \mathbf{n}, \tag{5}$$

where the model $\mathbf{m}$ represents the unknown parameters $(x,y,z$ for sources and receivers and t_0 values) and $\mathbf{n}$ represents data errors (noise), with the assumption that the error n_i on t_i is due to a Gaussian-distributed random process of standard deviation σ_i. The AEL inverse problem of estimating the model $\mathbf{m}$ that fits the data $\mathbf{t}$ is functionally nonlinear. A linearized estimate can be obtained by expanding $\mathbf{t}(\mathbf{m})=\mathbf{t}(\mathbf{m}_0+\delta\mathbf{m})$ in a Taylor series to first order about a starting model $\mathbf{m}_0$. Rearranging terms, the expansion can be written as

$$\mathbf{Jm} = \mathbf{t} - \mathbf{t}(\mathbf{m}_0) + \mathbf{Jm}_0 \equiv \mathbf{d}, \tag{6}$$

where $\mathbf{J}$ is the Jacobian matrix of partial derivatives $J_{ij}=\partial t_i(\mathbf{m}_0)/\partial m_j$ (analytic ray derivatives are given in Ref. 7), and $\mathbf{d}$ consists of known quantities and may be considered modified data. Equation (6) represents a linear inverse problem, which can be solved for $\mathbf{m}$ (discussed below). Since nonlinear terms are neglected, the inversion must be repeated iteratively until convergence.

Including both source and receiver positions as unknowns leads to an ill-conditioned inverse problem, which can be stabilized by including prior information via the method of regularization. This involves minimizing an objective function ϕ that includes the data misfit and regularizing terms:

$$\phi = |\mathbf{G}(\mathbf{Jm}-\mathbf{d})|^2 + \mu_1|\mathbf{H}_1(\mathbf{m}-\hat{\mathbf{m}})|^2 + \mu_2|\mathbf{H}_2\mathbf{m}|^2. \tag{7}$$

In Eq. (7), the first term on the right represents the data misfit for the linearized problem with $\mathbf{G}=\mathrm{diag}[1/\sigma_i]$, and the last two terms represent regularizations; μ_1 and μ_2 are trade-off parameters that control the relative importance of the regularizations in the minimization. The first regularization term in Eq. (7) is used to apply a prior estimate $\hat{\mathbf{m}}$ for the source and receiver locations based on knowledge of the deployment procedure; $\mathbf{H}_1=\mathrm{diag}[1/\xi_j]$ is a regularization matrix that weights prior estimate $\hat{m}_j$ according to its estimated standard deviation ξ_j, assuming Gaussian-distributed prior uncertainties. The second regularization term can be used to apply the prior expectation that the array shape and/or source tracks are smooth functions of position, where $\mathbf{H}_2$ is a discrete approximation to the second derivative operator with respect to position along the array/track.[14] The term $|\mathbf{H}_2\mathbf{m}|^2$ then provides a measure of the total curvature or roughness, and the regularization ensures that the solution has the simplest array shape and/or source track that is consistent with the data and prior parameter estimates.[7]

The regularized solution is obtained by setting $\partial\phi/\partial\mathbf{m}=0$, leading to

$$\mathbf{m} = \hat{\mathbf{m}} + [\mathbf{J}^T\mathbf{G}^T\mathbf{GJ} + \mu_1\mathbf{H}_1^T\mathbf{H}_1 + \mu_2\mathbf{H}_2^T\mathbf{H}_2]^{-1}[\mathbf{J}^T\mathbf{G}^T\mathbf{Gd} - \mathbf{J}\hat{\mathbf{m}}]. \tag{8}$$

From a starting model $\mathbf{m}_0$ (usually coinciding with prior $\hat{\mathbf{m}}$), Eq. (8) is applied iteratively until convergence is achieved. Convergence is based on two criteria: fitting the data to a statistically appropriate level such that the nonlinear data misfit $\chi^2=|\mathbf{G}(\mathbf{d}-\mathbf{d}(\mathbf{m}))|^2$ achieves its expected value of $\langle\chi^2\rangle=N$ for N data and achieving a stable solution in which the change in model parameters between successive iterations is small. Fitting the data to $\chi^2>N$ under-fits the data (i.e., under-constrains the solution); fitting the data to $\chi^2<N$ over-fits the data (i.e., fits noise on the data) at the expense of fitting the prior information.[7,8] Assigning values to the trade-off parameters, μ_1 and μ_2, to achieve a balance between the data misfit and the prior information requires a two-dimensional search, which can be carried out efficiently as described in Refs. 7 and 14.

An important component of any inverse problem involves estimating the uncertainty of the solution. For linearized problems with Gaussian-distributed errors and priors, the posterior model covariance matrix can be approximated,

$$\mathbf{C_m} = [\mathbf{J}^T\mathbf{G}^T\mathbf{GJ} + \mathbf{H}_1^T\mathbf{H}_1]^{-1}, \tag{9}$$

where $\mathbf{J}$ is evaluated at the final model, with the ith diagonal element of $\mathbf{C_m}$ representing the variance of the ith recovered parameter. The validity of this approach depends on the degree of nonlinearity but has generally been found to be a good approximation for AEL inversion.[14] Fully nonlinear uncertainty estimates can be computed using a Monte Carlo approach at higher computational cost. In the Monte Carlo approach, the source and receiver positions determined via inversion of the measured data are assumed to define the true positions for a synthetic inverse problem, and acoustic arrival-time data are computed. A series of independent inversions is then carried out, each with different random er-

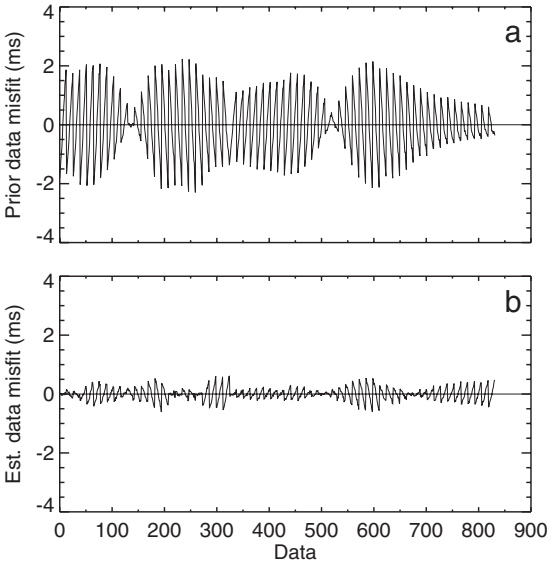


FIG. 7. Data residuals for (a) prior hydrophone positions and (b) recovered hydrophone positions.

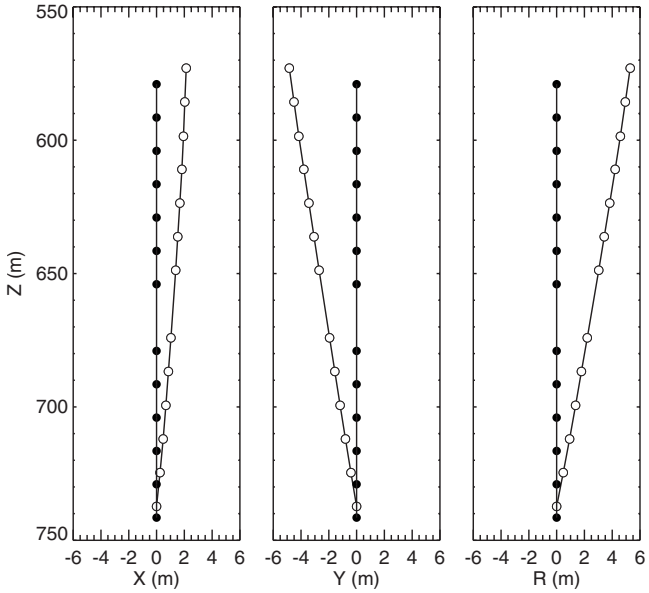


FIG. 8. Recovered array shape in x, y, and R indicated by open circles; filled circles indicate prior estimates. Receiver locations are relative to hydrophone 2 shifted to $x=y=0$.

rors applied to the computed data and to the prior position estimates (errors are drawn from Gaussian distributions with standard deviations equivalent to the corresponding uncertainties of the data and priors). Standard deviations about the true sensor positions can then be computed from the ensemble of inversion results. An advantage of the Monte Carlo approach is that it can be used to estimate both absolute localization errors (i.e., relative to the geographic coordinate system) and relative localization errors (in array-based coordinates), while linearization provides only absolute uncertainties. Relative uncertainties are more relevant in many applications because position errors common to all receivers are equivalent to a simple rigid-body translation and/or rotation of the array, while relative position errors introduce inter-receiver timing or phase errors which degrade array processing. Relative uncertainties are computed by correcting each realization of the Monte Carlo process for the optimal translation and rotation estimate[13,14] and can be expressed either referenced to the array centroid or as receiver-to-receiver uncertainties.

IV. SHIP-NOISE AEL RESULTS

The regularized inversion procedure described in Sec. III was applied to the data and prior estimates described in Sec. II to estimate locations for each source and receiver (the inverse problem involved 832 data and 295 unknown parameters). The algorithm converged in seven iterations, achieving a data misfit of $\chi^2=831$ (the expected value is 832), which is more than 23 times smaller than the misfit computed for the starting model of $\chi^2=19{,}400$. Figure 7 shows data residuals (difference between measured and predicted data) for the prior and recovered models. The residuals are reduced from a maximum of 2.2 ms for the prior model to less than 0.6 ms for the recovered model. The small prior residuals near data points 140 and 525 correspond to the ship passing nearly over the array on lines 1 and 2, respectively. At these points the ship is near endfire to the VLA, and the relative travel-times are sensitive only to the vertical hydrophone separation (i.e., modest errors in the x-y hydrophone positions do not affect the data significantly).

The x and y locations of the VLA and several nearby sources as estimated by the inversion are shown in Fig. 1, along with the corresponding prior estimates (the position of the VLA is represented by hydrophone 2). The VLA location is approximately 33 m east and 29 m south of the measured deployment position. Figure 8 shows the VLA shape in x, y, and $R=[x^2+y^2]^{1/2}$ relative to the position of hydrophone 2 shifted to $x=y=0$. The results indicate that the array is approximately 4 m shallower in depth than the prior estimate. Hydrophone 2 is nominally 31.5 m above the array mooring; therefore, using the estimated z value of 737.3 m, the depth of the water at the array location is estimated to be 768.8 m. The water depth determined from bathymetric charts for the array position from the inversion is 769.0 m, indicating excellent agreement. Figure 8 shows that the array (hydrophones 2–14) is tilted ~5.5 m in a south-southeasterly direction. This agrees with the average direction of the current (~0.1 m/s to the southeast) over the course of the experiment, measured by an upward-looking acoustic Doppler current profiler mounted near the base of the VLA. The array shape is smooth and physically reasonable; a slight curvature is apparent, which is consistent with the expected hydrodynamic behavior of a moored VLA in the presence of a current.

The inter-element spacing from the AEL solution varied from 12.4 to 12.9 m with a mean of 12.6 m, compared to the nominal spacing of 12.5 m. The total distance between hydrophone elements 2 and 14 from the AEL solution is 1.9 m longer than the nominal length. To check if the regularization stretched the array to minimize curvature, the inversion was re-run without applying the smoothing regularization to the array; however, there was no difference in the recovered array length. The VLA manufacturer verified that the combi-

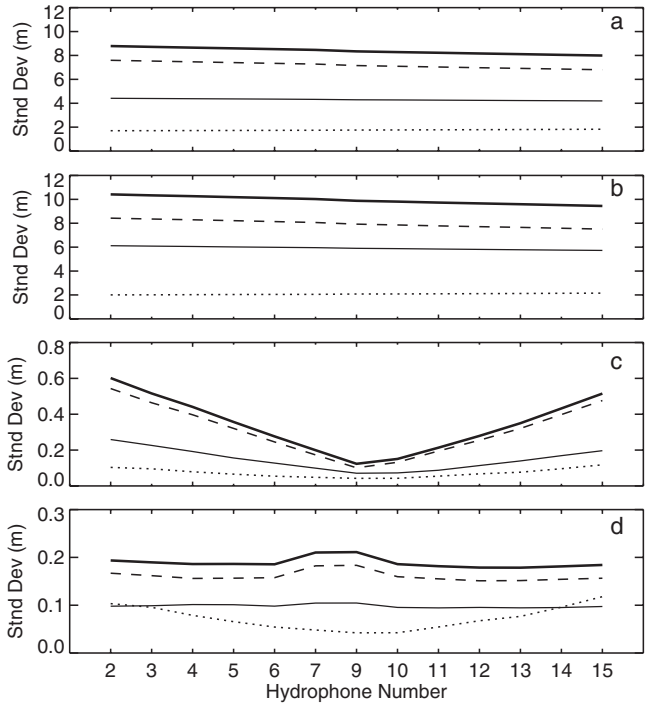


FIG. 9. AEL standard deviation estimates in x, y, z, and R indicated by solid, dashed, dotted, and heavy solid lines, respectively. Linearized (absolute) estimates are shown in (a). Nonlinear estimates are shown in (b) as absolute uncertainties, in (c) as relative uncertainties referenced to the array centroid, and in (d) as receiver-to-receiver relative uncertainties. Note that hydrophones 1, 8, and 16 are not included due to poor data.

nation of cable stretch under load and the precision to which the cable was manufactured and hydrophones positioned along the cable could easily result in an actual inter-element spacing of 12.6 m or more.

The estimated hydrophone localization uncertainties are given in Fig. 9. Figures 9(a) and 9(b) show that similar absolute uncertainties are estimated via linearization and the nonlinear Monte Carlo approach, respectively, with horizontal (R) uncertainties of approximately 10 m and vertical (z) uncertainties of about 2 m. Relative uncertainties are shown in Fig. 9(c) referenced to the array centroid. Finally, Fig. 9(d) shows the relative uncertainties between consecutive hydrophones, with R uncertainties of approximately 0.2 m and z uncertainties of about 0.05–0.1 m. The slightly higher standard deviations for hydrophones 7 and 9 result from the longer inter-receiver spacing caused by omitting hydrophone 8 from the analysis.

The inversion results for the source locations did not differ greatly from their prior estimates, likely because the data do not contain sufficient information to reposition both sources and receivers for this geometry (the role of source-receiver geometry in refining source locations is considered in Sec. V). The mean (absolute) standard deviations for the source positions estimated by the Monte Carlo method are 11.3 m in x, 11.9 m in y, and 2.5 m in z (linearized estimates are similar). These values are slightly smaller than the prior uncertainties (15 m in x and y, and 3 m in z).

V. SOURCE REPOSITIONING: SYNTHETIC STUDY

An interesting aspect of the AEL inversion developed in this paper is the application of the smoothing regularization

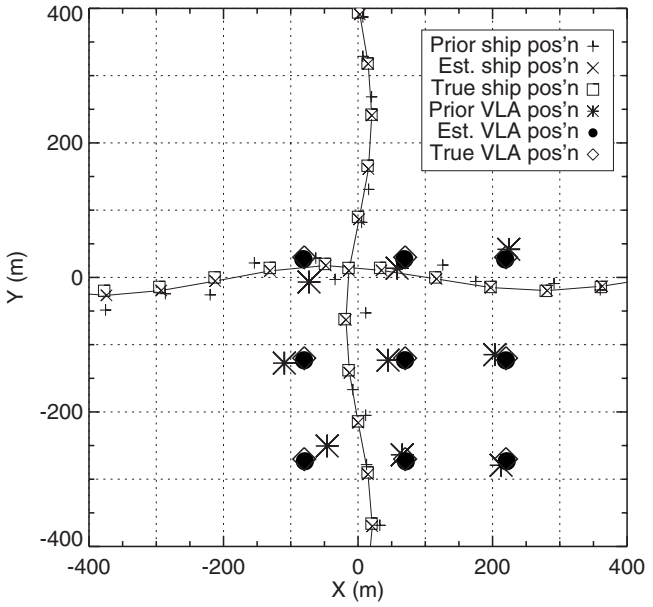


FIG. 10. Plan view showing true, prior, and estimated (est.) ship and hydrophone (h/ph) locations for the synthetic example. Only portions of the ship tracks are shown.

to the unknown source positions along the tracks, as well as to the array shape. In many AEL surveys, the source positions are considered independent of each other; however, here positions are extracted from continuous ship tracks, and it is desirable for the solution to minimize unrealistic track roughness. In the previous section, the measured data did not contain sufficient information to adjust the source positions significantly due to the limited spatial aperture of the array. In this section, a synthetic case is considered to illustrate how source position estimates can be improved in cases involving an array with sufficient horizontal aperture.

The geometry used for the synthetic example consists of nine equi-spaced receivers arranged in a grid pattern on the seafloor, as indicated by the diamond symbols in Fig. 10 (configurations like this are commonly used for ocean bottom seismometers in geophysical experiments). The array elements are 150 m apart at 800 m depth, and the center of the pattern is offset from the intersection of the source lines by 70 m in x and −120 m in y. A total of 64 source positions along two orthogonal lines are used (simulating the ship tracks of the MC798 experiment). The north-south source line is 2750 m long with 25 source positions, and the east-west line is 4480 m long with 39 positions. A sine function with 20-m amplitude modulates the source tracks to simulate a slowly meandering ship (Fig. 10). Synthetic (relative) arrival-time data were calculated for this source-receiver geometry, and Gaussian-distributed random errors of standard deviation σ=0.2 ms were added. Prior estimates of the source positions were generated by adding random Gaussian errors with standard deviations of 15 m in x and y and 3 m in z to the true positions. The prior estimates of the receiver positions included errors with standard deviations of 25 m in x and y and 5 m in z.

The inversion was carried out with the data and prior source position uncertainties outlined above. Smoothing regularization was applied to the source tracks but not to the

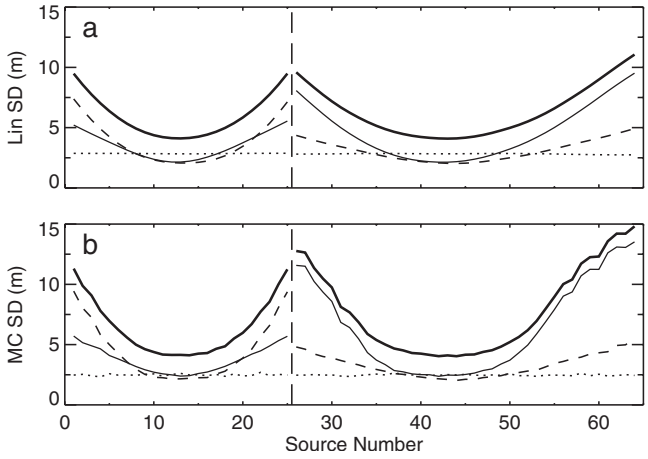


FIG. 11. Source position standard-deviation estimates in x, y, z and R indicated by solid, dashed, dotted, and heavy solid lines, respectively, for the synthetic example. Linearized estimates are shown in (a), and Monte Carlo absolute estimates in (b). Sources 1–25 correspond to the east-west line, and 26–64 to the north-south line.

receiver positions (which do not form a linear array). Figure 10 shows that both the source and receiver positions estimated via inversion are significantly closer to the true positions than the prior estimates. Figure 11 shows the linearized and nonlinear estimates of source position uncertainties. The vertical (z) localization errors are approximately 2.5 m for all source positions (compared to 3-m prior uncertainty). However, the horizontal uncertainties for all source positions are substantially smaller than the prior uncertainties of 15 m in x and y (or $\sim$21 m in R). The horizontal errors are smallest ($<$5 m in R) when the source is directly over the array due to a larger apparent array aperture; errors are largest ($\sim$10 m in R) at track end points where the apparent array aperture is smallest.

VI. SUMMARY

This paper demonstrated high-precision AEL using relative arrival times extracted from broadband ship noise, rather than from a standard AEL survey based on impulsive sources at distinct locations. Arrival-time delays between hydrophone pairs were extracted from ship noise using cross-correlation of the time-series recordings. Linearized inversion of the acoustic ray equations was applied to determine both source and receiver locations that fit the acoustic data and prior position estimates within uncertainties. Additional regularizations for the smoothest array shape and for the smoothest source tracks (relevant for a continuously-moving source such as a ship) were also applied.

Ship-noise AEL inversion results for a $\sim$160-m VLA in 800 m of water provided an array location and shape that is consistent with available information. Nonlinear (Monte Carlo) error analysis indicated absolute positioning uncertainties of approximately 10 m in the horizontal and 2 m in the vertical. Relative receiver-to-receiver uncertainties were 0.05–0.1 m in the vertical and about 0.2 m in the horizontal, which is much smaller than the receiver displacements along the VLA resolved by the AEL inversion.

The VLA geometry provided limited ability to resolve the source positions from their prior estimates; however, a synthetic study showed that array geometries with a greater horizontal aperture can provide significantly improved locations for sources as well as receivers.

[1]W. S. Burdic, *Underwater Acoustic System Analysis* (Prentice-Hall, Englewood Cliffs, NJ, 1994).
[2]A. Tolstoy, *Matched Field Processing for Underwater Sound* (World Scientific, Singapore, 1993).
[3]A. B. Baggeroer, W. A. Kuperman, and P. N. Mickhalevsky, "An overview of matched field methods in ocean acoustics," IEEE J. Ocean. Eng. **18**, 401–424 (1993).
[4]B. D. Steinberg, *Principles of Aperture and Array Systems Design* (Wiley, New York, 1976).
[5]K. C. Creager and L. M. Dorman, "Location of instruments on the seafloor by joint adjustment of instrument and ship positions," J. Geophys. Res. **87**, 8379–8388 (1982).
[6]B. J. Sotirin and J. A. Hildebrand, "Acoustic navigation of a large aperture array," J. Acoust. Soc. Am. **87**, 154–167 (1990).
[7]S. E. Dosso, M. R. Fallat, B. J. Sotirin, and J. L. Newton, "Array element localization for horizontal arrays via Occam's inversion," J. Acoust. Soc. Am. **104**, 846–859 (1998).
[8]S. E. Dosso, G. H. Brooke, B. J. Kilistoff, B. J. Sotirin, V. K. McDonald, M. R. Fallat, and N. E. Collison, "High-precision array element localization of vertical line arrays in the Arctic Ocean," IEEE J. Ocean. Eng. **23**, 365–379 (1998).
[9]S. E. Dosso and M. Riedel, "Array element location for towed marine seismic arrays," J. Acoust. Soc. Am. **110**, 955–966 (2001).
[10]S. E. Dosso and N. E. Collison, "Regularized inversion for towed-array shape estimation," *Inverse Problems in Ocean Acoustics*, edited by M. I. Taroudakis and G. N. Makrakis (Springer-Verlag, New York, 2001), pp. 77–103.
[11]S. E. Dosso and N. E. Collison, "Acoustic tracking of a freely drifting sonobuoy field," J. Acoust. Soc. Am. **111**, 2166–2177 (2002).
[12]N. E. Collison and S. E. Dosso, "Acoustic tracking of a freely-drifting sonobuoy field: Experimental results," IEEE J. Ocean. Eng. **28**, 554–561 (2003).
[13]S. E. Dosso, N. E. B. Collison, G. J. Heard, and R. I. Verrall, "Experimental validation of regularized array element localization," J. Acoust. Soc. Am. **115**, 2129–2137 (2004).
[14]S. E. Dosso and G. R. Ebbeson, "Array element localization accuracy and survey design," Can. Acoust. **34**, 3–13 (2006).
[15]E. K. Westwood and D. P. Knobles, "Source line localization via multipath correlation matching," J. Acoust. Soc. Am. **102**, 2645–2654 (1997).
[16]L. E. Freitag and P. L. Tyack, "Passive acoustic localization of the Atlantic bottlenose dolphin using whistles and echolocation clicks," J. Acoust. Soc. Am. **93**, 2197–2205 (1993).
[17]B. F. Kuryanov and A. I. Vedenev, "Investigation of the bottom sediment profiling method in the deep ocean by means of ship noise analysis," Oceanology (Engl. Transl.) **34**, 563–570 (1995).
[18]K. G. Sabra, P. Roux, A. M. Thode, G. L. D'Spain, W. S. Hodgkiss, and W. A. Kuperman, "Using ocean ambient noise for array self-localization and self-synchronization," IEEE J. Ocean. Eng. **30**, 338–347 (2005).
[19]W. S. Hodgkiss, P. Gerstoft, and J. J. Murray, "Array shape estimation from sources of opportunity," in *Oceans '99 MTS/IEEE Conference Proceedings*, Vol. **1**, pp. 582–585 (1999).

Chirp excitation technique to enhance microbubble displacement induced by ultrasound radiation force

Yi Hu and Dong Zhang[a)]
Key Laboratory of Modern Acoustics, Ministry of Education, Institute of Acoustics, Nanjing University, Nanjing 210093, China

Hairong Zheng
Institute of Biomedical and Health Engineering, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518067, China

Xiufen Gong
Key Laboratory of Modern Acoustics, Ministry of Education, Institute of Acoustics, Nanjing University, Nanjing 210093, China

(Received 1 April 2008; revised 31 October 2008; accepted 8 December 2008)

Ultrasound radiation force has been proposed to increase the targeting efficiency in ultrasonic molecular imaging and drug delivery. A chirp excitation technique is proposed to increase the radiation force induced microbubble displacement and might potentially be used for enhancing the targeting efficiency of microbubble clouds. In this study, a modified Rayleigh–Plesset equation is used to estimate the radius-time behavior of insonified microbubbles, and the translation of insonified microbubbles is calculated by using the particle trajectory equation. Simulations demonstrate that the chirp excitation is superior to the sinusoidal one in displacing microbubbles with a wide-size distribution, and that the performance is dependent on the parameters of the chirp signal such as the center frequency and frequency range. For Gaussian size distributed microbubble clouds with mean diameter of 3.5 μm and variance of 1, a 2.25 MHz chirp with frequency range of 1.5 MHz induces about 59.59% more microbubbles over a distance of 10 μm during 200 μs insonification, compared to a 2.25 MHz sinusoidal excitation with equal acoustic pressure.
 [DOI: 10.1121/1.3075548]

PACS number(s): 43.35.Mr, 43.35.Wa [DLM] Pages: 1410–1415

I. INTRODUCTION

Ultrasound contrast microbubbles, which are highly compressive and thus highly echogenic due to their construction, are usually injected into the blood stream in order to provide increased contrast between blood-filled organ and surrounding tissues.[1–3] Recent studies show that microbubbles could be utilized for ultrasonic molecular imaging as well as gene and drug delivery.[4–8] In the ultrasonic molecular imaging, targeted agents selectively attach molecular markers expressed on diseased endothelium and increase contrast in the area such as thrombus and inflammation.[9–18] For this purpose, it is desired to localize targeted microbubbles near the binding site, typically the vessel wall, for bond formation. However, microscopic observations show that microbubbles tend to distribute toward the axis of the vessel while circulating, similar to that of erythrocytes.[19–22] Previous efforts have shown that ultrasound radiation force can manipulate encapsulated microbubbles and displace them off the vessel axis in blood stream toward the vessel wall.[23] In addition, displacement to the side of the vessel could decrease the velocity of the microbubbles, allowing more time for bond formation. In their study, the radiation force induced by the sinusoidal excitation is significantly affected by the variation of resting radius, and the maximum displacement occurs to bubbles at their resonance frequency. Considering that contrast bubbles used usually have a wide-size distribution, the efficiency of targeting bubbles to the vessel wall with radiation force might be limited. The idea of using chirps to drive microbubbles has been investigated by several researchers with respect to contrast fundamental and harmonic imaging modalities.[24–26] Compared with the traditional pulse excitation, the chirp excitation offers higher Signal-to-Noise Ratio (SNR) and leads to better image quality and a significant increase in penetration depth in tissue.[26,27]

In this report, the effectiveness of a chirp excitation method is studied mathematically in producing acoustic radiation force and the displacing of microbubble clouds.

II. THEORETICAL MODEL

Microbubble dynamics in an incompressible liquid has been studied widely using various types of modified Rayleigh–Plesset (RP) equations.[3,28–32] Based on the modified RP equation considering shell properties and acoustic radiation damping, Morgan *et al.*[33] proposed a model to describe radius-time oscillations of insonified microbubbles.

[a)]Author to whom correspondence should be addressed. Electronic mail: dzhang@nju.edu.cn

 0001-4966/2009/125(3)/1410/6/$25.00

$$\rho_l\left(R\ddot{R}+\frac{3}{2}\dot{R}^2\right)=\left(P_0+\frac{2\sigma}{R_0}+\frac{2\chi}{R_0}\right)\times\left(\frac{R_0^3-R_0^3\left(\frac{b}{V_m}\right)}{R^3(t)-R_0^3\left(\frac{b}{V_m}\right)}\right)^{\gamma}\left(1-\frac{3\gamma}{c}\dot{R}\right)-4\frac{\mu\dot{R}}{R}-\frac{2\sigma}{R}\left(1-\frac{1}{c}\dot{R}\right)-\frac{2\chi}{R}\left(\frac{R_0}{R}\right)^2\times\left(1-\frac{3}{c}\dot{R}\right)-12\mu_{\mathrm{sh}}\varepsilon\frac{\dot{R}}{R(R-\varepsilon)}-(P_0+P_{\mathrm{driv}}(t)),\quad(1)$$

where R is the instantaneous radius of microbubble; $\dot{R}$ and $\ddot{R}$ are the first and second time derivatives of the instantaneous radius, respectively; R_0 is the resting radius; P_0 is the hydrostatic pressure; b is the van der Waals constant; V_m is the universal molar volume; ρ_l, σ, χ, γ, c, μ, μ_{sh}, and ε are the density, the surface tension of the liquid surrounding the bubble, elastic modulus of lipid shell, the polytropic gas exponent, the speed of sound in liquid, medium viscosity, viscosity of lipid shell, and thickness of lipid shell, respectively; and P_{driv} is the time-varying acoustic pressure driving the bubble.

As indicated by Marmottant *et al.*,[34] there is a problem in the model of Morgan *et al.*,[33] where the analysis of surface tension of microbubble fails to describe a coated microbubble. The quantity $\sigma(R)$ is an effective surface tension, which should be expressed as $\sigma(R)\simeq\sigma(R_0)+2\chi(R/R_0-1)$ at elastic state, rather than simply a constant. Using this expression of surface tension, we can get a modified equation as

$$\rho R\ddot{R}+\frac{3}{2}\rho\dot{R}^2=\left(P_0+\frac{2\sigma_0}{R_0}\right)\left(\frac{R_0}{R}\right)^{3\gamma}\left(1-\frac{3\gamma}{c}\dot{R}\right)-\frac{4\mu\dot{R}}{R}-\frac{2\sigma_0}{R}\left(1-\frac{1}{c}\dot{R}\right)-\frac{4\chi}{R_0}+\frac{4\chi}{R}\left(1-\frac{\dot{R}}{c}\right)-12\mu_{\mathrm{sh}}\varepsilon\frac{\dot{R}}{R(R-\varepsilon)}-(P_0+P_{\mathrm{driv}}(t)).\quad(2)$$

Equation (2) is equivalent to Eq. (3) in the work of Marmottant *et al.*[34] The translation motion of a microbubble in a fluid during insonification is calculated by solving a particle trajectory equation[23,31]

$$\rho_{b|}V_b\frac{du_b}{dt}=-V_b\frac{dP_l}{dx}+\frac{1}{2}\rho_l|u_r|u_rA\frac{24}{\frac{2R|u_l-u_b|}{\upsilon}}\times\left(1+0.197\left(\frac{2R|u_l-u_b|}{\upsilon}\right)^{0.63}+2.6\times10^{-4}\left(\frac{2R|u_l-u_b|}{\upsilon}\right)^{1.38}\right)+\left[\frac{1}{2}\rho_lV_b\frac{d}{dt}(u_l-u_b)\right]+\left[\frac{3}{2}\frac{\rho_lV_b(u_l-u_b)}{R}\frac{dR}{dt}\right]+(V_b(\rho_l-\rho_b)g-\mu R^2(u_l-u_b)^2\rho_l)\mu_F,\quad(3)$$

where ρ_b is the bubble density, υ is kinematic viscosity, V_b is the bubble volume, μ_F is the coefficient of friction, and P_l is the pressure in liquid. The value of u_r is defined by the difference between the bubble velocity and the liquid velocity, $u_r=u_b-u_l$. The term on the left-hand is the overall force on the microbubble, which is equal to the product of the microbubble mass and its acceleration. The first term on the right-hand is the radiation force on a highly compressible microbubble. The second term on the right-hand is the quasistatic drag force introduced by Johnson and Hsieh,[35] and further modified by Meyer *et al.*[36] The translation equation requires a Reynolds number of less than 100 for a satisfactory simulation. In this study, since the Reynolds number $\mathrm{Re}=U_0L/\upsilon=2R|u_l-u_b|/\upsilon$ has a small value of U_0 (about 10 mm/s) and a small value of L (several microns), the Reynolds number is less than 80, which is suitable for this simulation. The third term on the right is the added mass due to microbubble translation, and the fourth term is added mass due to microbubble oscillation. The fifth term is incorporated to consider the frictional force prominent in this experimental system.

The velocity of liquid is calculated from the acoustic pressure field as

$$\frac{\partial u_l}{\partial t}=-\frac{1}{\rho_l}\frac{\partial P_l}{\partial x}.\quad(4)$$

Pressure in liquid is calculated by[32]

$$P_l=P_0+P_a\cos\left(\frac{\omega}{c}x\right)\sin(\omega t),\quad(5)$$

where ω is the angular frequency of the acoustic signal.

A chirp is a special type of coded excitation signal with a long frequency modulated burst. A quadratic chirp signal with a sinusoidal envelope is usually described as

$$p(t)=A(t)\times\cos(2\pi f_0t+\phi(t)),\quad(6)$$

where $A(t)$ describes the envelope. $\phi(t)=\pi\beta t^2$ is a quadratic phase modulation function at the center frequency f_0. Accordingly, a linearly increasing instantaneous frequency $f_i(t)=f_0+(1/2\pi)(d\phi(t)/dt)=f_0+\beta t$ is obtained. Assume that the frequency range is from $f_0-\Delta f$ to $f_0+\Delta f$, the duration of the chirp signal is from $-T/2$ to $T/2$, then $\beta=2\Delta f/T$.

III. RESULTS AND DISCUSSIONS

In this study, the microbubbles are assumed to be Gaussian distributed, with a mean radius of 3.5 μm and a variance of 1, as shown in Fig. 1. The microbubble clouds have a concentration of 1×10^6 microbubbles/ml, and 1 ml solution is taken for numerical calculation. In order to compare with previous study,[23] we use the same shell parameters of the agent MP1950. The values of microbubble parameters in the numerical calculation are given in Table I.

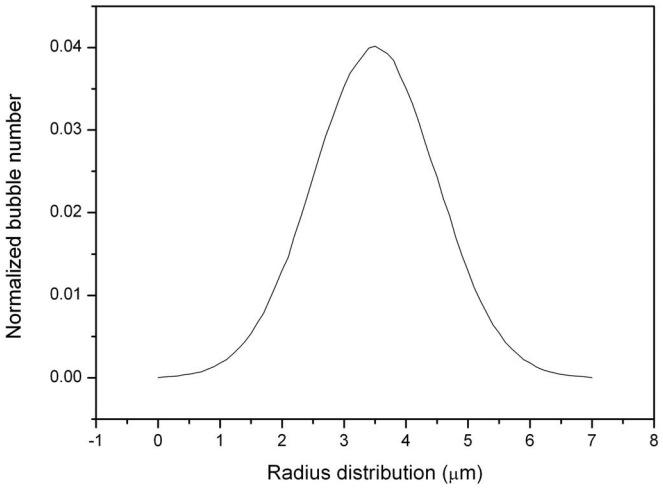


FIG. 1. Normalized size distribution of microbubble clouds.

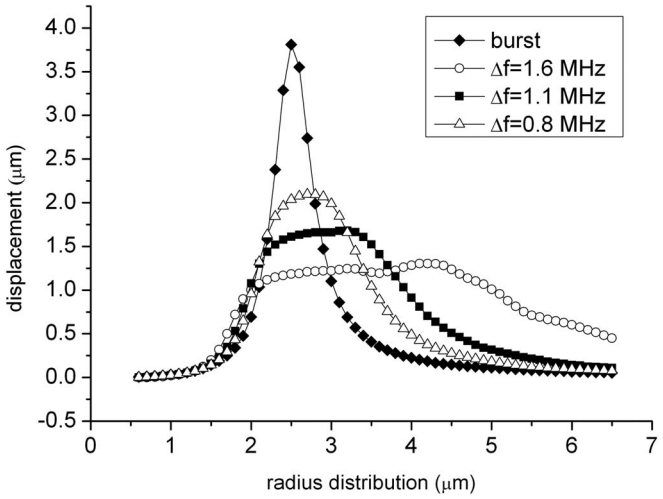


FIG. 2. Microbubble displacements insonified with 10 μs sinusoidal excitation or linear chirp excitation.

A. Microbubble displacement induced by chirp excitation

First, the displacements of the Gaussian distributed microbubbles excited by chirp signals are investigated numerically. Equations (2) and (3) are numerically solved by the fourth-order Runge–Kutta method with the initial values $R=R_0$, $R'=0$, $u_b=0$, and $du_b/dt=0$ at $t=0$. Forced oscillation and displacement of individual microbubbles are examined with the linear chirp excitation $(A(t)=1)$.

Figure 2 compares the displacements of microbubbles driven by a sinusoidal signal and linear chirp signals, in which the sinusoidal signal is at 2.25 MHz with acoustic pressure amplitude of 100 kPa; the linear chirp signals with the equal acoustic pressure amplitudes have the center frequency of 2.25 MHz and frequency ranges of 0.8, 1.1, and 1.5 MHz, respectively. For the case of sinusoidal excitation, there is a sharp peak of displacement for microbubble with the radius of 2.5 μm that is close to the resonance frequency of 2.25 MHz. It suggests that the sinusoidal excitation is efficient in displacing microbubbles at resonance. For the case of chirp excitation, the shapes of displacement versus radius curves are relatively flat without sharp peaks. The maximum displacement is obviously decreased compared to that by sinusoidal excitation. For example, the displacement induced by the chirp with the frequency shift of 0.8 MHz is about 59.46% of that by the sinusoidal excitation. However, the chirp excitation has the capability of displacing microbubbles with a distribution. Larger frequency shift performs better in inducing the movements of microbubbles with wider size distribution.

TABLE I. Values of microbubble parameters in the numerical calculation.

Symbol	Description	Value
b	van der Waals constant	0.1727
ε	Thickness of lipid shell (m)	2×10^{-9}
c	Acoustic velocity in liquid (m/s)	1540
P_0	Hydrostatic pressure (Pa)	1.013×10^5
ρ_b	Bubble density (kg/m^3)	11.2
σ_0	Surface tension (N/m)	0.051
μ_{sh}	Viscosity of lipid shell (Pa s)	1.27
μ_F	Coefficient of friction	0.5
μ	Medium viscosity (Pa s)	0.001
υ	Kinematic viscosity (Pa s)	1×10^{-6}
V_m	Universal molar volume (l)	22.4
χ	Elastic modulus of lipid shell	1

B. Optimization of chirp excitations

To obtain optimized parameters of the chirp excitation, the effects of the center frequency and frequency range on the displacement response from microbubbles excited by the chirp signal are investigated numerically. Figure 3 plots the bubble displacements for chirp excitations about 10 μs at center frequencies of 1, 2.25, and 5 MHz, respectively. The radii corresponding to linear resonance frequencies of 5, 2.25, and 1 MHz of microbubbles are approximately 0.8, 1.6, and 3.25 μm, respectively.[34] It is observed that increasing of the center frequency leads to the decrease in the magnitude of maximum translation. For example, the maximum translations of microbubble are 6, 3.3, and 1.75 μm for chirps at center frequencies of 1, 2.25, and 5 MHz, respectively. On the other hand, as frequency shift increases, the maximum translation of microbubbles decreases, while there is an increase in radius distribution of bubbles that travel away from the transducer for a certain distance. As shown in Fig. 3(c), the translational displacement distribution induced by chirp with frequency shift of 0.3 MHz is much sharper than that by chirp with frequency shift of 4.1 MHz.

C. Comparison of sinusoidal and chirp excitations

Figure 4 compares the efficiency of displacement of microbubble clouds over a fixed distance induced by sinusoidal and chirp excitations, where the percentage number of microbubbles whose displacements over 30, 40, and 50 μm during 200 μs are calculated by dividing by the overall number of the initial microbubbles. The center frequency and acoustic pressure amplitude for both sinusoidal signal and chirp signal are 1 MHz and 100 kPa. Note that the sinusoidal signal can be considered as a chirp signal with no frequency

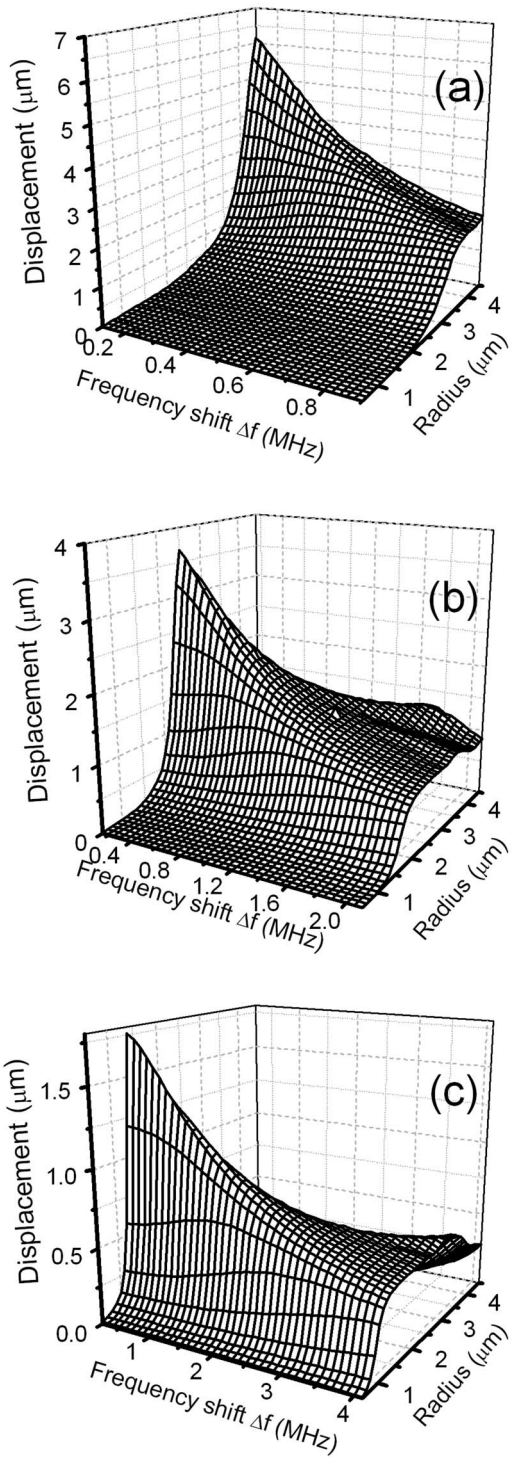


FIG. 3. Effect of center frequency and frequency shift range on microbubble displacements induced by linear chirp excitation with acoustic pressure amplitude of 100 kPa: (a) center frequency of 1 MHz and frequency shift range of 0.14–0.9 MHz, (b) center frequency of 2.25 MHz and frequency shift range of 0.2–2.1 MHz, and (c) center frequency of 5 MHz and frequency shift range of 0.3–4.1 MHz.

shift. It is observed that the chirp signal is superior to sinusoidal signal in displacing bubbles over the same distance, and the larger frequency range displaces more bubbles. For example, the chirp with frequency shift of 0.75 MHz induces about 56%, 52%, and 0% microbubbles moving over 30, 40, and 50 μm; the chirp with frequency shift of 0.25 MHz

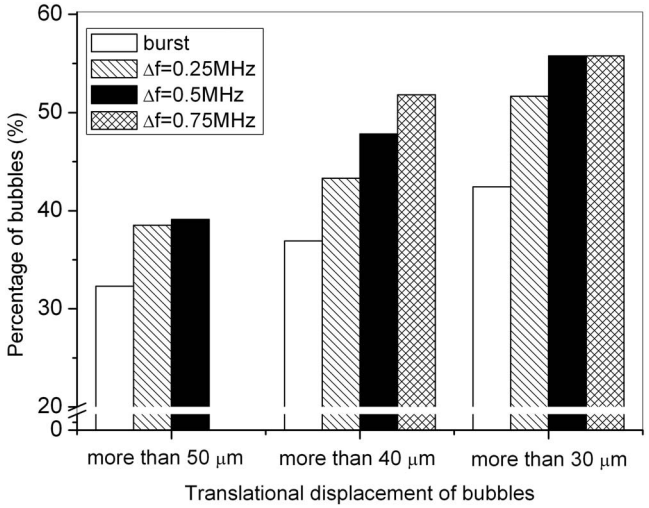


FIG. 4. Comparison of efficiency of displacement of microbubble clouds over a fixed distance between 1 MHz sinusoidal and chirp excitations with acoustic pressure is 100 kPa for 200 μs insonification.

induces about 50%, 43%, and 38% microbubbles moving over 30, 40, and 50 μm; while the sinusoidal signal induces about 46%, 36%, and 32% microbubbles over 30, 40, and 50 μm. Although the maximum displacement of resonant bubble decreases, more off-resonant bubbles are displaced. It demonstrates that when the exciting frequency is close to the resonant frequency of the mean radius of bubble clusters, they show better performance for both sinusoidal and chirp signals, and chirp excitation displaces more bubbles than sinusoidal excitation.

Figure 5 shows the calculated percentage of microbubbles translating over a fixed distance during 200 μs insonified by 2.25 MHz sinusoidal and chirp signals with equal acoustic pressure of 100 kPa. Similar to Fig. 4, chirp excitation exhibits better capability in displacing microbubble clouds than sinusoidal excitation, for example, chirp signal with 1.5 MHz frequency shift displaces 94.77% over 10 μm, much more than 35.18% for sinusoidal excitation. Percentage numbers of microbubbles translating more

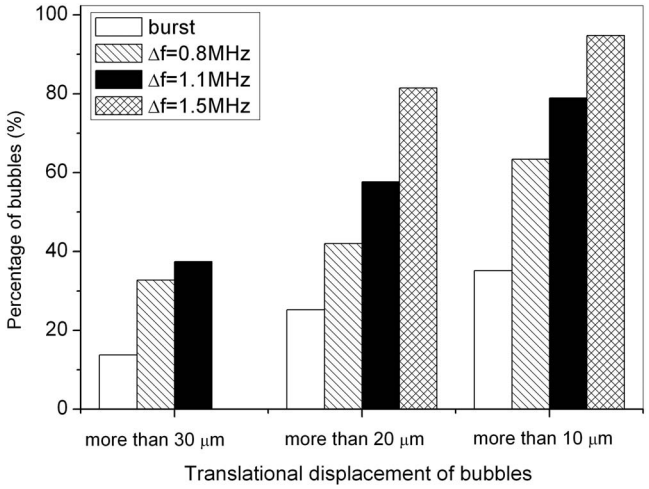


FIG. 5. Comparison of efficiency of displacement of microbubble clouds over a fixed distance between 2.25 MHz sinusoidal and chirp excitations with acoustic pressure is 100 kPa for 200 μs insonification.

than 30 μm insonified by sinusoidal and chirps are, respectively, 13.8%, 32.8%, and 37.46%, which are much smaller than those in Fig. 4. It demonstrates that when the resonant size is far from the size distribution center of the bubble cluster, both sinusoidal and chirp excitations generally decrease in performance, but chirps with larger frequency shift show more capability in displacing more bubbles over considerable distances.

In the previous study, the equation of Morgan *et al.*[33] was utilized to estimate the radiation force under sinusoidal excitation. Our study modified the surface tension expression according to the discussion of Marmottant *et al.*[34] Although both the model of Morgan *et al.*[33] and the modified model demonstrate that chirp excitation gets a shape of relatively flat without sharp peak for the displacement versus radius curve as shown in Fig. 2, the modified model generates slightly larger displacement, for example, about 10% when frequency shift is 0.8 MHz. As frequency shift increases, displacement gradually becomes close to that obtained from the model of Morgan *et al.*[33] Similarly, both models indicate that the chirp signal is much more capable in displacing microbubbles with wide-size distribution over sinusoidal excitation, but the results differ in quantity although the general trends are similar. Additionally, it should be noted that the radial and translational equations are coupled together. Depending on the size and shell property of the microbubble, the pressure amplitude plays an important factor for the coupled relationship. According to the experimental study within the work by Zheng *et al.*,[37] the pressure amplitude used in this study is still in normal range and is not expected to change the coupled bubble dynamics. Detailed descriptions on the effect of translation on radial oscillation can be found in the works by Doinikov[38] and Kuznetsova *et al.*,[39] where an additional term provides feedback between translation and radial oscillations.

IV. CONCLUSIONS

Acoustic radiation force has shown promise in facilitating targeted microbubble adhesion. In this paper, we compared the displacements of microbubbles excited by sinusoidal and chirp excitations, and explored the possibility of maximizing the adhesion efficiency of a cluster with different sized microbubbles insonified by linear chirp signal. Based on the modified RP equation and particle trajectory equation, we provided an evaluation of microbubble displacement induced by chirp excitation and investigation of optimization of chirp excitation respected to center frequency and frequency range. Results indicated that translational displacement of individual bubbles insonified with chirp signals is dependent on the center frequency, frequency shift, and acoustic pressure. Also, both sinusoidal and chirp excitations with center frequency close to the mean radius of the bubble cluster show better performance. In addition, chirp signal is much more capable in displacing microbubbles with wide-size distribution over sinusoidal excitation. Further experimental study is ongoing to examine the feasibility of this technique.

ACKNOWLEDGMENTS

This work is supported by Program for New Century Excellent Talents in University (Grant No. 06-0450), National Natural Science Foundation of China (Grant No. 10774071), National Natural Science Foundation of Jiangsu Province (Grant No. BK2007518), and State Key Laboratory of Acoustics.

[1] C. X. Deng and F. L. Lizzi, "A review of physical phenomena associated with ultrasonic contrast agents and illustrative clinical applications," Ultrasound Med. Biol. **28**, 277–286 (2002).
[2] A. L. Klibanov, "Ultrasound contrast agents: Development of the field and current status," Top. Curr. Chem. **222**, 73–106 (2002).
[3] N. De Jong, R. Cornet, and C. T. Lancee, "Higher harmonics of vibrating gas-filled microspheres. I. Simulations," Ultrasonics **32**, 447–453 (1994).
[4] P. A. Dayton and K. W. Ferrara, "Targeted imaging using ultrasound," J. Magn. Reson Imaging **16**, 362–377 (2002).
[5] A. L. Klibanov, "Microbubble contrast agents-Targeted ultrasound imaging and ultrasound-assisted drug-delivery applications," Invest. Radiol. **41**, 354–362 (2006).
[6] G. M. Lanza and S. A. Wickline, "Targeted ultrasonic contrast agents for molecular imaging and therapy," Curr. Probl Cardiol. **28**, 625–653 (2003).
[7] J. R. Lindner and S. Kaul, "Delivery of drugs with ultrasound," Echocardiogr. **18**, 329–337 (2001).
[8] E. C. Unger, T. O. Matsunaga, T. McCreery, P. Schumann, R. Sweitzer, and R. Quigley, "Therapeutic applications of microbubbles," Eur. J. Radiol. **42**, 160–168 (2002).
[9] G. E. Weller, E. M. Tom, M. M. Felix, W. R. Wagner, and F. S. Villanueva, "Multi-targeting enhances adhesion of ultrasound contrast microbubbles to inflammatory endothelium," Circulation **108**, 516–517 (2003).
[10] S. Zhao, M. Borden, S. H. Bloch, D. Kruse, K. W. Ferrara, and P. A. Dayton, "Radiation-force assisted targeting facilitates ultrasonic molecular imaging," Mol. Imaging **3**, 135–148 (2004).
[11] M. J. Shortencarier, P. A. Dayton, S. H. Bloch, P. A. Schumann, T. O. Matsunaga, and K. W. Ferrara, "A method for radiation-force localized drug delivery using gas-filled lipospheres," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **51**, 822–831 (2004).
[12] M. A. Borden, M. R. Sarantos, S. M. Stieger, S. I. Simon, K. W. Ferrara, and P. A. Dayton, "Ultrasound radiation force modulates ligand availability on targeted contrast agents," Mol. Imaging **5**, 139–147 (2006).
[13] J. J. Rychak, A. L. Klibanov, and J. A. Hossack, "Acoustic radiation force enhances targeted delivery of ultrasound contrast microbubbles: In vitro verification," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **52**, 421–433 (2005).
[14] A. F. Lum, M. A. Borden, P. A. Dayton, D. E. Kruse, S. I. Simon, and K. W. Ferrara, "Ultrasound radiation force enables targeted deposition of model drug carriers loaded on microbubbles," J. Controlled Release **111**, 128–134 (2006).
[15] H. Leong-Poi, J. Christiansen, A. L. Klibanov, S. Kaul, and J. R. Lindner, "Noninvasive assessment of angiogenesis by ultrasound and microbubbles targeted α_v–integrins," Circulation **107**(3), 455–460 (2003).
[16] G. E. Weller, E. Lu, M. M. Csikari, A. L. Klibanov, D. Fischer, W. R. Wagner, and F. S. Villanueva, "Ultrasound imaging of acute cardiac transplant rejection with microbubbles targeted to intercellular adhesion molecule-1," Circulation **108**, 218–224 (2003).
[17] G. E. Weller, M. K. K. Wong, R. A. Modzelewski, E. X. Lu, A. L. Klibanov, W. R. Wagner, and F. S. Villanueva, "Ultrasonic imaging of tumor angiogenesis using contrast microbubbles targeted via the tumor-binding peptide arginine-arginine-leucine," Cancer Res. **65**, 533–539 (2005).
[18] P. M. Winter, S. D. Caruthers, X. Yu, S. K. Song, J. J. Chen, B. Miller, J. W. M. Bulte, J. D. Robertson, P. J. Gaffney, S. A. Wickline, and G. M. Lanza, "Improved molecular imaging contrast agent for detection of human thrombus," Magn. Reson. Med. **50**, 411–416 (2003).
[19] J. R. Lindner, A. R. Jayaweera, J. Sklenar, and S. Kaul, "Microvascular rheology of Definity microbubbles after intra-arterial and intravenous administration," J. Am. Soc. Echocardiogr **15**, 396–403 (2002).
[20] S. Ismail, A. R. Jayaweera, G. Camarano, L. W. Gimple, E. R. Powers, and S. Kaul, "Relation between air-filled albumin microbubble and red blood cell rheology in the human myocardium: Influence of echocardiographic systems and chest wall attenuation," Circulation **94**, 445–451 (1996).

[21]A. R. Jayaweera, N. Edwards, W. P. Glasheen, F. S. Villanueva, R. D. Abbot, and S. Kaul, "In vivo myocardial kinetics of air filled albumin microbubbles during myocardial contrast echocardiography: Comparison with radiolabeled red blood cells," Circ. Res. **74**, 1157–1165 (1994).
[22]M. W. Keller, S. S. Segal, S. Kaul, and B. R. Duling, "The behavior of sonicated albumin microbubbles within the microcirculation: A basis for their use during myocardial contrast echocardiography," Circ. Res. **65**, 458–467 (1989).
[23]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).
[24]Y. Sun, S. K. Zhao, P. A. Dayton, and K. W. Ferrara, "Observation of contrast agent response to chirp insonation with a simultaneous optical-acoustical system," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **53**, 1130–1137 (2006).
[25]T. X. Misaridis, K. Gammelmark, C. H. Jorgensen, N. Lindberg, A. H. Thomsen, M. H. Pedersen, and J. A. Jensen, "Potential of coded excitation in medical ultrasound imaging," Ultrasonics **38**, 183–189 (2000).
[26]T. Misaridis and J. A. Jensen, "Use of modulated excitation signals in medical ultrasound. Part I: Basic concepts and expected benefits," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **52**, 177–191 (2005).
[27]J. M. Borsboom, C. T. Chin, and N. de Jong, "Nonlinear coded excitation method for ultrasound contrast imaging," Ultrasound Med. Biol. **29**, 277–284 (2003).
[28]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).
[29]X. Yang and C. C. Church, "A model for the dynamics of gas bubbles in soft tissue," J. Acoust. Soc. Am. **118**, 3595–3606 (2005).
[30]D. B. Khismatullin and A. Nadim, "Radial oscillations of encapsulated microbubbles in viscoelastic liquids," Phys. Fluids **14**, 3534–3557 (2002).
[31]T. Watanabe and Y. Kukita, "Translational and radial motions of a bubble in an acoustic standing-wave field," Phys. Fluids A **5**, 2682–2688 (1993).
[32]H. J. Rath, "The translational motion of a nonlinear oscillating gas bubble in a compressible liquid in the presence of an inhomogeneous sound field," Acustica **44**, 148–155 (1980).
[33]K. E. Morgan, J. S. Allen, P. A. Dayton, J. E. Chomas, K. 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).
[34]P. Marmottant, S. van der Meer, M. Emmer, and M. Versluis, "A model for large amplitude oscillations of coated bubbles accounting for buckling and rupture," J. Acoust. Soc. Am. **118**, 3499–3505 (2005).
[35]V. E. Johnson and T. Hsieh, "The influence of the trajectories of gas nuclei on cavitation inception," Sixth Naval Hydrodynamics Symposium (1966), pp. 163–182.
[36]R. S. Meyer, M. L. Billet, and J. W. Holl, "Freestream nuclei and traveling bubble cavitation," ASME Trans. J. Fluids Eng. **114**, 672–679 (1992).
[37]H. R. Zheng, P. A. Dayton, K. Caskey, S. K. Zhao, S. P. Qin, and K. W. Ferrara, "Ultrasound driven microbubble oscillation and translation within microvessels," Ultrasound Med. Biol. **33**, 1978–1987 (2007).
[38]A. A. Doinikov, "Translational motion of a spherical bubble in an acoustic standing wave of high intensity," Phys. Fluids **14**, 1420–1425 (2002).
[39]L. A. Kuznetsova, S. Khanna, N. N. Amso, W. T. Coakley, and A. A. Doinikov, "Cavitation bubble-driven cell and particle behavior in an ultrasound standing wave," J. Acoust. Soc. Am. **117**, 104–112 (2005).

Acoustic field in a quasi-spherical resonator: Unified perturbation model

Cécile Guianvarc'h[a] and Laurent Pitre
Institut National de Métrologie, 61 Rue du Landy, 93210 La Plaine Saint Denis, France

Michel Bruneau and Anne-Marie Bruneau
Laboratoire d'Acoustique de l'Université du Maine, UMR CNRS 6613, Avenue Olivier Messiaen, 72085 Le Mans Cedex 9, France

(Received 23 July 2008; revised 31 December 2008; accepted 31 December 2008)

Gas-filled quasi-spherical resonators are excellent tools for the measurement of thermophysical properties of gas and have also been retained for the determination of the Boltzmann constant with a low uncertainty, which can be derived from measurements of both the speed of sound in a noble gas and the volume of the resonator. To achieve this, a detailed modeling of the acoustic field in quasi-spherical resonators is of importance. Several phenomena and perturbations must be taken into account, including, among inertia and compressibility, heat conduction, viscosity, the shape of the resonator, small irregularities on the wall, and so on. The aim of this paper is to provide improvements to the current models of the acoustic field in such resonator. Namely, the model given here takes into account all the different perturbing elements together in a unique formalism, including the coupling between the different perturbing elements and the resulting modal coupling in a consistent manner. The first results obtained from this analytical model on a simple configuration show that the effect of modal coupling is small but should not be neglected regarding the accuracy required here, even if several improvements could still be provided to this new unified model. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3075764]

PACS number(s): 43.35.Ud, 43.20.Ks [RR] Pages: 1416–1425

I. INTRODUCTION

The International Committee for Weights and Measures (CIPM) recently recommended the redefinition of the International System Units (SI), where the kelvin would be linked to an exactly defined value of the Boltzmann constant.[1] The advantage would be that the definition of the kelvin would not depend on any temperature nor on any method for its measurement. The value to be chosen for the Boltzmann constant must be known with a relative uncertainty of 10^{-6}. Gas-filled quasi-spherical resonators are excellent tools for the determination of the Boltzmann constant, which can be derived from measurements of both the speed of sound in a noble gas and the volume of the resonator. The choice of a quasi-spherical shape for the shell allows us to have a high quality factor for the acoustic resonances and non-degenerate electromagnetic modes.

Actually, an accurate design of the shell and a detailed modeling of the acoustic field in spherical or quasi-spherical resonators are of importance to measure thermophysical properties of gas and the Boltzmann constant. Several phenomena must be taken into account, including inertia, compressibility, heat conduction and shear viscosity (in the boundary layers and in the bulk of the fluid), bulk viscosity, the real shape of the resonator (in fact, quasi-spherical shape), the acoustic input impedance of small acoustic elements flush-mounted on the wall that are necessary for the measurements (tubes and transducers), geometrical irregularities, roughness of the wall, and shell motion due to external vibration sources and to the coupling between the fluid and the shell. All these phenomena, which act as perturbations to an idealized acoustic fluid filling a perfect rigid spherical shell, need to be taken into account to determine with a low uncertainty their influence on the resonance properties of the resonator and then on the value obtained for the Boltzmann constant.

Groundbreaking experiments and a number of significant theoretical studies have already been made by Moldover *et al.*,[2] Mehl and co-workers,[3–6] and Trusler.[7] In particular, they suggested means to take into account several perturbations in the cavity. It is worth noting that these perturbations are taken into account separately, then neglecting the coupling between them. These perturbations are the thermal and viscous effects (in the boundary layers and the bulk of the gas), the deformation of the cavity, the influence of small elements located on the wall of the resonator (tubes, slits, and transducers), and the vibroacoustic coupling between the fluid and the shell.

The aim of this paper is to suggest unified modeling able to gather all the different types of perturbing factors in a unique formalism, including the coupling between the different perturbation factors through the modal coupling) neglected until now, which would be of interest to have a better interpretation of the measurement results in the quasi-spherical resonator. Here, the acoustic field is expressed by the coupling between Neumann modes of an ideal, unperturbed resonator that bounds outwardly the perturbated surface of the resonator, the coupling being due to energy trans-

[a]Electronic mail: cecile.guianvarch@cnam.fr

 0001-4966/2009/125(3)/1416/10/$25.00

fer between modes induced by every perturbation). This model is then used on the most simple experimental conditions: a perfectly spherical cavity, filled with a dissipative gas (argon), with only a receiver and a transmitter microphone flush-mounted on the wall. Assuming perfectly spherical cavity in the application considered in the present paper enables us to evaluate the effect of modal coupling due to the viscous and thermal effects in the bulk and in the boundary layers (modeled at the lowest order) and to the microphones.

The fundamental equations of acoustic propagation in dissipative fluid are given in Sec. II. In Sec. III, the solution of the fundamental problem (propagation equation associated to the boundary conditions) for the acoustic pressure in the quasi-spherical resonator is expressed as an expansion on the spherical eigenfunctions, making use of the integral formulation. Finally in Sec. IV, results are given and discussed. These results are obtained with this model applied on the most simple experimental configuration which has been carried out in 1988 at the NIST.[2]

II. FUNDAMENTAL EQUATIONS OF ACOUSTICS IN DISSIPATIVE FLUID

The formalism used in the following lies on the works of Kirchhoff, gathered later by Rayleigh,[8] Morse and Ingard,[9] Pierce,[10] and Bruneau and Scelo.[11] The presentation of the fundamental equations of acoustics chosen here refers to those of Bruneau and Scelo.

The system considered is a quasi-spherical cavity because its shape retains certain advantages of spherical acoustic resonators (high quality factor) while simplifying the measurement of the resonator volume using microwave resonances (electromagnetic modes are not degenerate for the shape chosen here). The variables describing the dynamic and thermodynamic states of the fluid are the pressure variation p, the particle velocity $\boldsymbol{v}$, the density variation ρ', the entropy variation σ, and the temperature variation τ. The parameters which specify the properties and the nature of the fluid are the values of the density ρ_0, the static pressure P_0, the shear viscosity coefficient μ, the bulk viscosity coefficient η, the coefficient of thermal conductivity λ, the specific heat coefficient at constant pressure and constant volume per unit of mass C_P and C_V, respectively, the specific heat ratio γ, and the increase in pressure per unit increase in temperature at constant density $\hat{\beta}=(\partial P/\partial T)_V$ $(\hat{\beta}\gamma=\alpha\rho_0 c_0^2 = 1/V(\partial V/\partial T)_P$, α being the volume thermal expansivity and c_0 the adiabatic speed of sound). A complete set of linearized homogeneous equations governing small amplitude disturbances of the fluid includes the following:

- the Navier–Stokes equation,

$$\frac{1}{c_0}\frac{\partial \boldsymbol{v}}{\partial t} + \frac{1}{\rho_0 c_0}\,\mathbf{grad}\, p = \ell_v\,\mathbf{grad}\,\mathrm{div}\,\boldsymbol{v} - \ell_v'\,\mathbf{rot}\,\mathbf{rot}\,\boldsymbol{v}, \quad (1)$$

where the characteristic lengths ℓ_v and ℓ_v' are defined as follows,

$$\ell_v = \frac{1}{\rho_0 c_0}\left(\frac{4}{3}\mu + \eta\right) \quad \text{and} \quad \ell_v' = \frac{\mu}{\rho_0 c_0},$$

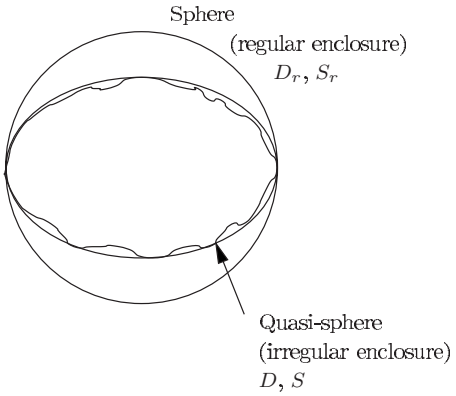


FIG. 1. Sphere and quasi-sphere with irregular boundary conditions.

- the conservation of mass equation, taking into account the thermodynamic law expressing the density variation as function of the independent variables p and τ,

$$\rho_0 c_0\,\mathrm{div}\,\boldsymbol{v} + \frac{\gamma}{c_0}\frac{\partial}{\partial t}(p - \hat{\beta}\tau) = 0, \quad (2)$$

- the Fourier equation for heat conduction, taking into account the thermodynamic law expressing the entropy variation as function of the independent variables p and τ,

$$\left(\frac{1}{c_0}\frac{\partial}{\partial t} - \ell_h\Delta\right)\tau = \frac{\gamma-1}{\hat{\beta}\gamma}\frac{1}{c_0}\frac{\partial p}{\partial t}, \quad (3)$$

the operator Δ being the Laplacian, where the characteristic length ℓ_h is defined as

$$\ell_h = \frac{\lambda}{\rho_0 c_0 C_P}.$$

Equations (1)–(3), associated to the boundary conditions, describe the acoustic field in the quasi-spherical cavity. The purpose of the following section is to find the propagation equations for the acoustic pressure, the temperature variation, and the particle velocity.

III. THE HARMONIC FIELD IN A QUASI-SPHERICAL CAVITY

In this section a standard analytic procedure is developed whereby the acoustic field inside the cavity bounded by perturbed surface (irregular enclosure, domain D, surface S) is expressed as a sum over the eigenmodes (Neumann) of a cavity having separable geometry (regular enclosure, domain D_r, surface S_r) that bounds outwardly the perturbed enclosure considered (Fig. 1).

The resulting acoustic modeling presented in Sec. III B thus includes in a coherent manner all types of perturbation in the resonator (quasi-spherical shape, dissipation in the bulk of the fluid, and acoustic elements on the wall) and the resulting modal coupling.

A. The boundary problem with sources

The acoustic pressure in the perturbed enclosure (domain D bounded by surfaces S) is governed by the set of

equations, including the propagation equation with harmonic source term $f(\boldsymbol{r})$ on the right hand side and the boundary conditions with harmonic boundary source term $\sigma(\boldsymbol{r})$ on the right hand side, which takes the following form:

$$(\Delta + k_a^2)p(\boldsymbol{r}) = -f(\boldsymbol{r}) \quad \text{in } D, \tag{4a}$$

$$\left[\frac{\partial}{\partial n} + jk_0\beta(\boldsymbol{r})\right]p(\boldsymbol{r}) = \sigma(\boldsymbol{r}) \quad \text{on } S, \tag{4b}$$

where the symbol $\partial/\partial n$ implies a normal derivative outwardly directed, $k_0=\omega/c_0$ is the adiabatic wavenumber, k_a is the acoustic wavenumber including the dissipation in the bulk of the fluid (Appendix A), and the parameter $\beta(\boldsymbol{r})$ is the specific admittance of the wall. The specific admittance $\beta(\boldsymbol{r})$ can include the viscous and thermal boundary layer effects and any kind of discontinuities and perturbation on the wall of the resonator (tubes, slits, transducers, roughness, etc.). For the problem considered, there is no source in the domain D, then the term $f(\boldsymbol{r})$ is zero.

B. The associated eigenvalue problem

The modal wave functions of an ideal spherical cavity (domain D_r bounded by a surface S_r), which encloses the perturbed enclosure, are solutions of the homogeneous Helmholtz equation subject to the Neumann boundary condition at the wall, namely,

$$(\Delta + k_m^2)\psi_m^t(\boldsymbol{r}) = 0 \quad \text{in } D_r, \tag{5a}$$

$$\frac{\partial}{\partial n}\psi_m^t(\boldsymbol{r}) = 0 \quad \text{on } S_r, \tag{5b}$$

where the subscript m stands for three quantum numbers (ν,μ,η) and the superscript t (denoted by either c or s) represents the two degenerate eigenfunctions which depend on the azimuthal angle φ. The eigenfunctions ψ_m^t expressed in the spherical coordinates (r,θ,φ) are

$$\psi_m^c(r,\theta,\varphi) = \frac{1}{a_m} j_\nu(k_{\nu\eta}r)P_{\mu\nu}(\cos\theta)\cos\mu\varphi, \tag{6a}$$

$$\psi_m^s(r,\theta,\varphi) = \frac{1}{a_m} j_\nu(k_{\nu\eta}r)P_{\mu\nu}(\cos\theta)\sin\mu\varphi, \tag{6b}$$

where the functions $P_{\mu\nu}$ are Legendre functions, and the coefficients a_m are normalization coefficients (see Appendix B).

Expanding the pressure field $p(\boldsymbol{r})$ on the eigenfunctions $\psi_p^{t'}(\boldsymbol{r})$, leads to

$$p(\boldsymbol{r}) = \sum_{p,t'} c_p^{t'}\psi_p^{t'}(\boldsymbol{r}). \tag{7}$$

Then, multiplying Eq. (4a), where $f=0$, by the eigenfunction $\psi_m^t(\boldsymbol{r})$ and integrating all over the domain D, i.e.,

$$\int\int\int_D \psi_m^t(\boldsymbol{r})(\Delta + k_a^2)p(\boldsymbol{r})dD = 0, \tag{8}$$

solution of the posed problem for the acoustic pressure field is subsequently achieved with the help of Green's integral theorem which states

$$\int\int\int_D p(\boldsymbol{r})\Delta\psi_m^t(\boldsymbol{r})dD - \int\int\int_D \psi_m^t(\boldsymbol{r})\Delta p(\boldsymbol{r})dD = \int\int_S p(\boldsymbol{r})\frac{\partial}{\partial n}\psi_m^t(\boldsymbol{r})dS - \int\int_S \psi_m^t(\boldsymbol{r})\frac{\partial}{\partial n}p(\boldsymbol{r})dS. \tag{9}$$

The expressions of $\Delta\psi_m^t$ and $(\partial/\partial n)\psi_m^t$ being required to satisfy Eqs. (5a) and (5b), respectively, and the solutions p being expressed by Eq. (7), Eq. (9) gives straightforwardly the following equation:

$$\sum_{p,t'} c_p^{t'}[(k_m^2 - k_a^2)N_{mp}^{tt'} + A_{mp}^{tt'} + E_{mp}^{tt'}] = S_m^t, \tag{10}$$

where

$$N_{mp}^{tt'} = \int\int\int_D \psi_p^{t'}(\boldsymbol{r})\psi_m^t(\boldsymbol{r})dD, \tag{11a}$$

$$A_{mp}^{tt'} = \int\int_S \psi_p^{t'}(\boldsymbol{r})\frac{\partial}{\partial t}\psi_m^t(\boldsymbol{r})dS, \tag{11b}$$

$$E_{mp}^{tt'} = \int\int_S jk_0\beta(\boldsymbol{r})\psi_p^{t'}(\boldsymbol{r})\psi_m^t(\boldsymbol{r})dS, \tag{11c}$$

$$S_m^t = \int\int_S \sigma(\boldsymbol{r})\psi_m^t(\boldsymbol{r})dS, \tag{11d}$$

which is equivalent to the matrix equation, using the notation $[k^2N]$ for the matrix whose elements are $k_m^2N_{mp}^{tt'}$,

$$([k^2N] - k_a^2[N] + [A] + [E])[C] = [S]. \tag{12}$$

The matrix $[N]$ expresses the influence on the acoustic field of the depth of the deformation: the radial coordinate of the cavity depends on the angles θ and φ in spherical coordinates (the effect expressed by the matrix $[N]$ can be called "bulk" or "global"[12] modal coupling, corresponding to energy transfer between the mode labeled m and the mode labeled p). The matrix $[A]$ expresses the influence on the acoustic field of the slope of the deformation: the normal to the wall of the resonator depends on θ and φ, too. The matrix $[E]$ expresses the effects on the acoustic field of the perturbations on the wall, taking also into account the shape of the resonator (the effects expressed by the matrices $[A]$ and $[E]$ can be called "boundary" or "local"[12] modal coupling, also corresponding to energy transfer between the mode labeled m and the mode labeled p). The vector $[S]$ expresses the effect of the acoustic source set on the wall of the resonator. Finally, the matrix $[k^2N]-k_a^2[N]+[A]+[E]$ in Eq. (12) involves coupling terms between the different modes (in the cross products $\psi_p^{t'}(\boldsymbol{r})\psi_m^t(\boldsymbol{r})$, weighed or not by the admittance β, and $\psi_p^{t'}(\boldsymbol{r})(\partial/\partial n)\psi_m^t(\boldsymbol{r})$), whereas the right hand side involves the modes created by the source.

The eigenvalues of the matrix in the left hand side of Eq. (12) include then together all the perturbations we would take into account in the calculation of the resonance characteristics (resonance frequencies and half-widths) of the cavity: the effect of the deformation is included in the matrices

$[N]$ and $[A]$, and all the perturbing factors on the wall of the cavity are included in the specific admittance β of the matrix $[E]$ for the passive acoustic effects and elements (boundary layer effects, tubes, transducers, irregularities, roughness, vibroacoustic coupling between the fluid and the shell, etc.) and in the matrix $[S]$ for the active acoustic effects and elements (acoustic source and shell motion due to external vibration sources).

Remark. In previous works by Mehl,[3,4] using the general formalism of Morse and Feshbach[13] and focusing only on the effects of the small deformations of the cavity (quasi-spherical shape) from the regular shape (perfect spherical shape), all the other perturbations on the wall of the resonator being taken into account by other means, the terms E_{mp} and S_m in Eq. (10) are then zero, which leads to, neglecting terms $N_{mp}^{tt'}$ for $m \neq p$ and $t \neq t'$ as made by Mehl[3] [Eq. (8)],

$$[(k_a^2 - k_m^2)N_{mm}^{tt} - A_{mm}^{tt}]c_m^t = \sum_{p\neq m, t\neq t'} c_p^{t'} A_{mp}^{tt'}.$$

The coefficients c_m^t are then given by

$$c_m^t = \frac{\Sigma_{p\neq m, t\neq t'} A_{mp}^{tt'} c_p^{t'}}{(k_a^2 - k_m^2)N_{mm}^{tt} - A_{mm}^{tt}}, \tag{13}$$

and the difference $(k_a^2 - k_m^2)$ by

$$k_a^2 - k_m^2 = \frac{1}{N_{mm}^{tt} c_m^t} \sum_{p\neq m, t\neq t'} A_{mp}^{tt'} c_p^{t'} + \frac{A_{mm}^{tt}}{N_{mm}^{tt}},$$

that is

$$k_a^2 - k_m^2 = \frac{A_{mm}^{tt}}{N_{mm}^{tt}} + \frac{1}{N_{mm}^{tt} c_m^t} \sum_{p\neq m, t\neq t'} A_{mp}^{tt'} \frac{\Sigma_{q\neq p, t''\neq t'} A_{pq}^{t't''} c_q^{t''}}{(k_a^2 - k_p^2)N_{pp}^{t't'} - A_{pp}^{t't'}}, \tag{14}$$

then, neglecting the non-diagonal terms in the sum over q and assuming that $(k_a^2 - k_p^2)N_{pp}^{t't'} \gg A_{pp}^{t't'}$ and $A_{mp}^{tt'} = A_{pm}^{t't}$,

$$k_a^2 - k_m^2 \approx \frac{A_{mm}^{tt}}{N_{mm}^{tt}} + \sum_{p\neq m, t\neq t'} \frac{A_{mp}^{tt'\,2}}{(k_a^2 - k_p^2)N_{pp}^{t't'} N_{mm}^{tt}}. \tag{15}$$

In expression (15) of the frequency shifts due to the deformation of the cavity, derived from Morse and Feshbach formalism,[13] the coupling terms $N_{mp}^{tt'}$ $(m \neq p$ and $t \neq t')$ are neglected (the bulk modal coupling is neglected, assuming quasi-orthogonality), whereas the terms $A_{mp}^{tt'}$ are kept; i.e., the boundary coupling is taken into account (due to the boundary slope). Furthermore, in these previous works, the symmetry of the matrix $[A]$ is assumed but not demonstrated in the general case considered (every shape of the cavity). It is the aim of the remainder of this paper to keep all the coupling terms that was not taken into account in previous works in order to know more accurately the influence of each one.

IV. APPLICATION: ACOUSTIC FIELD IN A SPHERICAL RESONATOR

The case studied here is the spherical cavity used at the NIST to measure the gas constant.[2] This spherical cavity (mean radius $R = 8.890\,143 \times 10^{-2}$ m) is filled with a dissipative gas (argon) with a receiving microphone (specific admittance β_r and surface ΔS_r) and a transmitting microphone (sound source, vibration velocity V_σ, and surface ΔS_σ). The dissipation in the bulk of the fluid is taken into account in the complex wavenumber k_a. The boundary layer effects are expressed in terms of specific admittance β_{vh} at the lowest order (order 1/2 of the characteristic lengths ℓ_h and ℓ'_v). This specific admittance is uniform on the total area of the wall S.

The shell motion can be taken into account in the matrices E and S in an appropriate manner for the determination of the Boltzmann constant only if the spatial repartition of this shell motion is known or modeled with great accuracy. Previous works[14] allow us to determine the influence of the shell motion on the acoustic resonance frequencies of a spherical cavity with a simple shape. However, regarding the complexity of the real shape of the resonator's shell, it does not seem possible, at short- or mid-term, to carry out a vibroacoustical model with the accuracy required for this application. Then, currently, the acoustical measurements are limited to the frequency ranges for which the effect of the shell motion is not significant in the experimental results (this corresponds to frequencies far from the structural resonance frequencies of the shell). The shell motion is thus neglected here in the global modeling of the acoustic field.

For the perfect sphere considered below, the domain D and the surface S are the same as the regular domain D_r and surface S_r, respectively. There is no modal coupling due to the deformation of the domain and the enclosure. As a consequence, the matrix $[N]$ is the identity and $[A]$ is zero. Modal coupling in such case is then due to dissipation and perturbations on the wall of the cavity only (matrix $[E]$). The integral equation (10) leads then to

$$\sum_{p,t'} c_p^{t'} [(k_m^2 - k_a^2)\delta_{mp}\delta_{tt'} + E_{mp}^{tt'}] = S_m^t, \tag{16}$$

$\gamma_{\nu,\eta}$ being the nth root of the first derivative of the spherical Bessel function j_ν and where [Eqs. (11c) and (11d)]

$$S_m^t = \int\int_S \sigma(\boldsymbol{r}) \psi_m^t(\boldsymbol{r}) dS, \tag{17a}$$

$$E_{mp}^{tt'} = \int\int_S jk_0 \beta(\boldsymbol{r}) \psi_p^{t'}(\boldsymbol{r}) \psi_m^t(\boldsymbol{r}) dS. \tag{17b}$$

The small surfaces ΔS_r and ΔS_σ are much smaller than the total area S and sufficiently small to assume that the associated specific admittances, velocity, and eigenfunctions are uniform on them, giving for Eqs. (17a) and (17b)

$$S_m^t = \sigma \psi_m^t(R, \theta_\sigma, \varphi_\sigma) \Delta S_\sigma, \tag{18a}$$

and

$$E_{mp}^{tt'} = {}^{vh}E_{mp}^{tt'} + {}^{r}E_{mp}^{tt'} = jk_0[\beta_{vh} \langle \psi_p^{t'} | \psi_m^t \rangle_S + \beta_r \Delta S_r \langle \psi_p^{t'} | \psi_m^t \rangle_c], \tag{18b}$$

where the notation $\langle \psi_p^{t'} | \psi_m^t \rangle_i$ stands for the integral of the product of the eigenfunctions $\psi_p^{t'}$ and ψ_m^t over the surface either S or ΔS_i.

TABLE I. θ and φ coordinates of the small elements flush-mounted on the wall of the resonator.

	θ_i	φ_i
Transmitting microphone	π	0
Receiving microphone	$\pi/2$	$3\pi/4$

The contributions in the modal coupling of the boundary layer effects (acting on the whole surface S of the wall) and of each small element flush-mounted on the wall are calculated in Secs. IV A and IV B, respectively, when the resonator is excited by a local sound source, whose influence (terms S_m) is calculated in Sec. IV C. The coordinates θ and φ of the small elements flush-mounted on the wall are given in Table I, corresponding to the design of the cavity used at INM.

A. Viscous and thermal boundary layer effects

The specific admittance on the surface S due to the viscous and the thermal boundary layer effects has the form [11] (at the order $1/2$ of the characteristic lengths)

$$\beta_{vh} \approx \frac{1+j}{\sqrt{2}}\sqrt{k_0}\left[\left(1-\frac{k_r^2}{k_a^2}\right)\sqrt{\ell_v'}+(\gamma-1)\sqrt{\ell_h}\right], \tag{19}$$

where the wavenumber k_a is related to the radial and azimuthal wavenumbers k_r and k_w by $k_a^2=k_r^2+k_w^2$. In the expression of this admittance, which acts as a small correction on the acoustic field, the complex wavenumber k_a, including the dissipation in the bulk of the fluid (Sec. III A), is replaced by the adiabatic wavenumber k_0, the dissipation in the bulk being much smaller than in the boundary layers.

The term $(1-k_r^2/k_a^2)$ is related to the incidence of the acoustic wave on the wall, which depends on the acoustic modes that occur in the cavity. Then, because of the modal coupling in the cavity, while the acoustic source generates an acoustic field at the angular frequency ω, different acoustic modes are excited at this angular frequency in the cavity. The specific admittance associated to each acoustic mode in Eq. (19) is then a function of the wavenumber $k_0=\omega/c_0$ and of the factor related to this mode, namely,

$$\beta_{vh}=\beta_{\nu\eta}\approx\frac{1+i}{\sqrt{2}}\sqrt{k_0}\left[\frac{\nu(\nu+1)}{\gamma_{\nu\eta}^2}\sqrt{\ell_v'}+(\gamma-1)\sqrt{\ell_h}\right]. \tag{20}$$

On the other hand, the integral $\langle\psi_p^{t'}|\psi_m^t\rangle_S$ on the whole surface S is given by

$$\langle\psi_p^{t'}|\psi_m^t\rangle_S=\int_0^{2\pi}\int_0^{\pi}\psi_p^{t'}(R,\theta,\varphi)\psi_m^t(R,\theta,\varphi)R^2\sin\theta d\theta d\varphi$$

that gives straightforwardly (because of the orthogonality of the Legendre's functions and of functions sin and cos), for the influence of the viscous and thermal boundary layers in the resonator,

$${}^{vh}E_{mp}^{tt'}=jk_0\beta_{\nu\eta}R^2\frac{j_\nu(\gamma_{\nu\eta})j_\nu(\gamma_{\nu\eta'})}{\mathcal{J}_{\nu\eta}\mathcal{J}_{\nu\eta'}}\delta_{\nu\nu'}\delta_{\mu\mu'}\delta_{tt'}, \tag{21}$$

$\mathcal{J}_{\nu\eta}$ being the norm of the function $j_\nu(k_{\nu\eta}r)$ (see Appendix B).

B. Receiving microphone flush-mounted on the wall

On the small surface ΔS_r, the integral $\langle\psi_p^{t'}|\psi_m^t\rangle_r$ reduces to

$$\langle\psi_p^{t'}|\psi_m^t\rangle_r=\psi_m^t(R,\theta_r,\varphi_r)\psi_p^{t'}(R,\theta_r,\varphi_m ic)\Delta S_r \tag{22}$$

that gives, for the coordinates $\theta_r=\pi/2$ and $\varphi_r=3\pi/4$ (Table I), making use of the relations $k_{\nu\eta}R=\gamma_{\nu\eta}$,

$${}^{r}E_{mp}^{tt'}=jk_0\frac{1}{a_m a_p}\Delta S_r\beta_r j_\nu(\gamma_{\nu\eta})j_{\nu'}(\gamma_{\nu'\eta'})P_{\nu\mu}(\cos\theta_r)$$
$$\times P_{\nu'\mu'}(\cos\theta_r)\phi^t(3\pi\mu/4)\phi^{t'}(3\pi\mu'/4), \tag{23}$$

the function ϕ^t standing for either cos or sin.

The expression chosen here for the specific admittance β_r of the receiving microphone is the expression used in the study done at the NIST,[2]

$$\beta_r=j\omega\rho_0c_0\chi, \tag{24}$$

χ being the compliance per unit area of the membrane of the microphone (nominal value of 1.5×10^{-10}).

C. Sound source

The influence of the source on the acoustic field is expressed by [Eq. (18a)]

$$S_m^t=\sigma\psi_m^t(R,\theta_\sigma,\varphi_\sigma)\Delta S_\sigma, \tag{25}$$

where $\sigma=j\omega\rho_0V_\sigma$, V_σ being the vibration velocity of the source that gives

$$S_m^t=jk_0\rho_0c_0V_\sigma\Delta S_\sigma\psi_m^t(R,\theta_\sigma,\varphi_\sigma). \tag{26}$$

The sound source being an electrostatic microphone (used as a transmitter), the volume velocity $V_\sigma\Delta S_\sigma$ of the sound source is then derived from the following transduction equation:

$$M_{p\sigma}i=Y_\sigma p(R,\theta_\sigma,\varphi_\sigma)+\Delta S_\sigma V_\sigma, \tag{27}$$

where $M_{p\sigma}$ and Y_σ are, respectively, the pressure sensitivity and the acoustic admittance of the microphone, and i is the electric current through the microphone. Making use of Eqs. (27) and (7) in Eq. (26) gives

$$S_m^t=jk_0\rho_0c_0M_{p\sigma}i\psi_m^t(R,\theta_\sigma,\varphi_\sigma)-\sum_{p,t'}c_p^{t'\sigma}E_{mp}^{tt'}, \tag{28a}$$

with

$${}^{\sigma}E_{mp}^{tt'}=jk_0\rho_0c_0Y_\sigma\langle\psi_p^{t'}|\psi_m^t\rangle_\sigma \tag{28b}$$

and

$$\langle\psi_p^{t'}|\psi_m^t\rangle_\sigma=\psi_m^r(R,\theta_\sigma,\varphi_\sigma)\psi_p^{t'}(R,\theta_\sigma,\varphi_\sigma)\Delta S_\sigma. \tag{28c}$$

The influence of the source in the acoustic field is then composed of an active part linked to the electrical excitation applied on the microphone and a reactive part due to the acoustic input admittance of the microphone flush-mounted on the wall.

Expressions (23) and (28a)–(28c) of the effects of small elements flush-mounted on the wall of the cavity show that

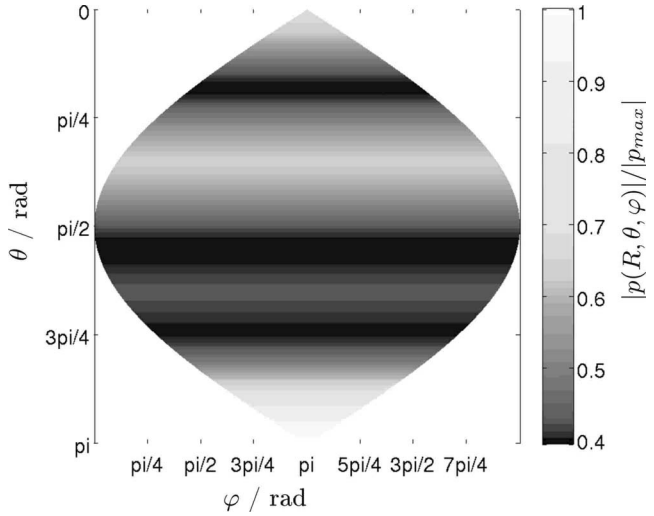


FIG. 2. Acoustic pressure on the wall of a spherical cavity at the resonance frequency of the (0,2) radial mode at the static pressure of 25 kPa.

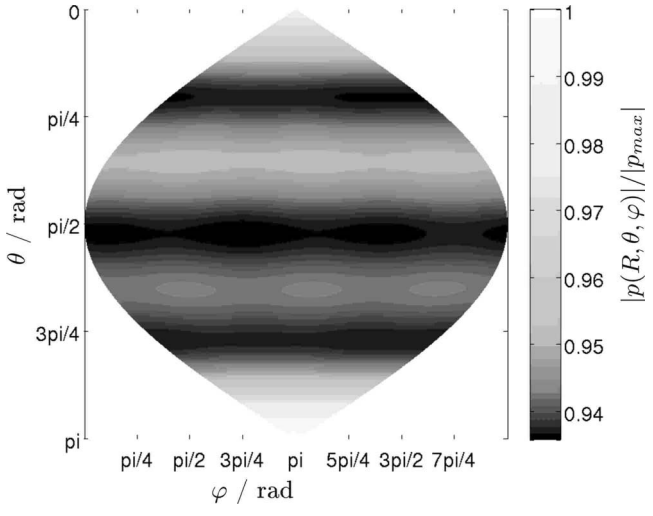


FIG. 4. Acoustic pressure on the wall of a spherical cavity at the resonance frequency of the (0,2) radial mode at the static pressure of 500 kPa.

these effects are strongly dependent on the relative position of these elements on the wall (neglected in the previous works).

D. Results

1. Effect of modal coupling on the acoustic pressure

The theoretical effect of modal coupling on the bulk behavior of the acoustic pressure in a spherical cavity filled with argon gas, at the triple point of water for different static pressures, is shown in Figs. 2–5, representing the cartographic projection of the amplitude of the acoustic pressure on the wall of the cavity, with respect to the spherical coordinates φ and θ, when the transmitting microphone generates a harmonic signal at the resonance frequency of a radial mode. The cavity has only a transmitting and a receiving microphone on its wall.

These almost qualitative representations in the lower static pressure range, in Figs. 2 and 3, show that the acoustic field is quasi-axisymmetric around the axis of the transmitting microphone because the influence of the receiving microphone is minimum. The maxima at $\theta=\pi$ and $\theta=0$ are due to the transmitting microphone. The other deviations from a uniform acoustic pressure that should be observed at the resonance frequency of a radial mode are due to the influence of the viscous and thermal effects on non-radial modes, which are immediately close to the radial mode expected. Then, the (0,2) mode is mostly perturbed by the (3,1) mode, and the (0,5) mode by the (12,1) mode, all of them without azimuthal components because of the axisymmetry of the system for such low static pressures.

In the higher static pressure range, the viscous and thermal effects are less important; then the deviations from a uniform acoustic pressure on the wall of the cavity due to modal coupling are smaller, which can be seen in Figs. 4 and 5. The impedance of the receiving microphone decreases, and then its influence on the acoustic pressure in the cavity increases (but remains small), generating also azimuthal modes by modal coupling.

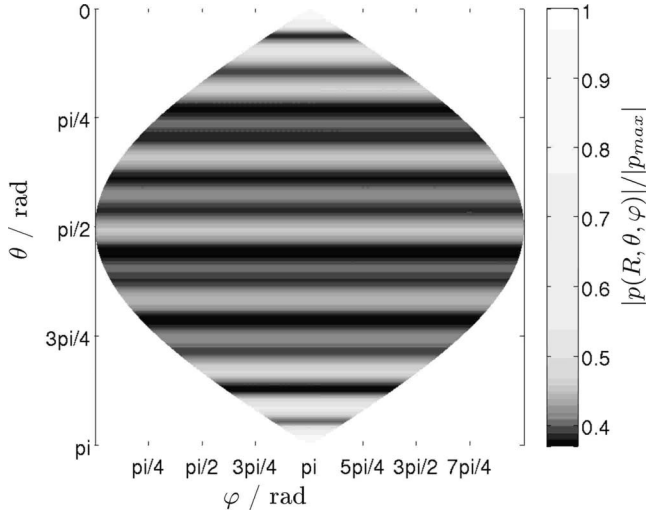


FIG. 3. Acoustic pressure on the wall of a spherical cavity at the resonance frequency of the (0,5) radial mode at the static pressure of 25 kPa.

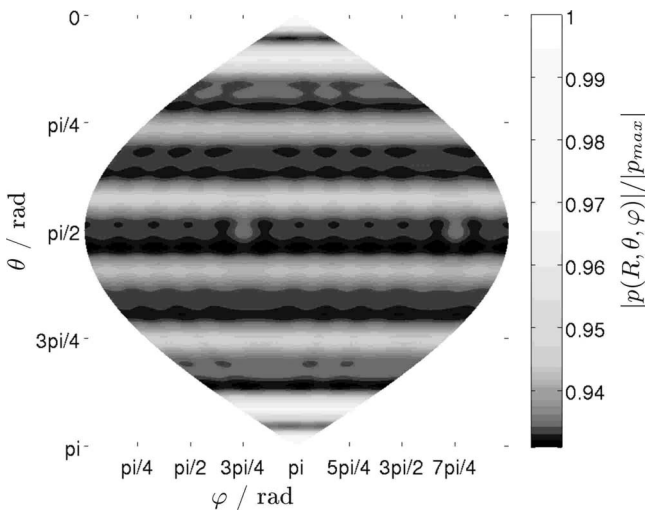


FIG. 5. Acoustic pressure on the wall of a spherical cavity at the resonance frequency of the (0,5) radial mode at the static pressure of 500 kPa.

Such maps of the acoustic field on the wall of the cavity could be used to support the design of an acoustic resonator for the determination of the Boltzmann constant or for acoustic thermometry. Indeed, they show us the optimal places on the wall for the measurement devices to maximize the signal to noise ratio and reduce the perturbations effects. For example, here, the best place for the receiving microphone would be in front of the transmitting microphone (maximum of acoustic pressure for all the acoustic modes and static pressures of interest, axisymmetric acoustic system).

Moreover, the particular results presented in this section show that to elucidate the individual physical phenomena in the formalism, it is necessary to consider separately the physical phenomena involved and to do simplifying assumptions.

2. Influence of modal coupling on the acoustic resonance properties of the cavity

For the determination of the Boltzmann constant, the value of k is derived from the extrapolation at zero pressure of the speed of sound determined from measurements of the acoustic resonance frequencies of radial modes in the cavity, on which are applied correction terms calculated from the model of the acoustic field in the cavity. The comparison between the experimental and theoretical half-widths allows us to check the validity of the model used to calculate the corrections to be applied on the resonance frequencies.

The global model presented here allows us to calculate the acoustic frequency response of the resonator. Fitting a resonant function (Lorenzian function) with this calculated transfer function around the resonance frequency of an acoustic mode gives us the resonance frequency and the half-width of this mode according to the global analytical model.

Since the global formalism presented here includes modal coupling (non-diagonal terms in the matrices), a direct analytic expression of the perturbation effects on the resonance properties of the cavity cannot be derived from this global modeling without doing several simplifying assumptions. Then, only a global and numerical comparison between the current and previous theoretical results is presented here. However, it is worthy to notice that neglecting the terms $m \neq p$ in the matrix E in Eq. (16) leads finally to the analytical expressions of the frequency shifts and half-widths determined in previous works.[2,6]

The effect of the acoustic perturbations in the cavity and of the resulting modal coupling on the resonance properties of radial modes from (0,2) to (0,6) as function of the static pressure is shown in Figs. 6 and 7. The major perturbation in the cavity being the viscous and thermal effects, the absolute values of the frequency shifts and the half-widths both increase when the static pressure and the quantum number of the modes decrease.

The comparison between the results obtained with this unified model and the model used previously at the NIST to determine the current value of the Boltzmann constant[2] (Figs. 8 and 9) shows the discrepancies due to the only modal coupling, the models used for each source of perturbation being the same in both methods.

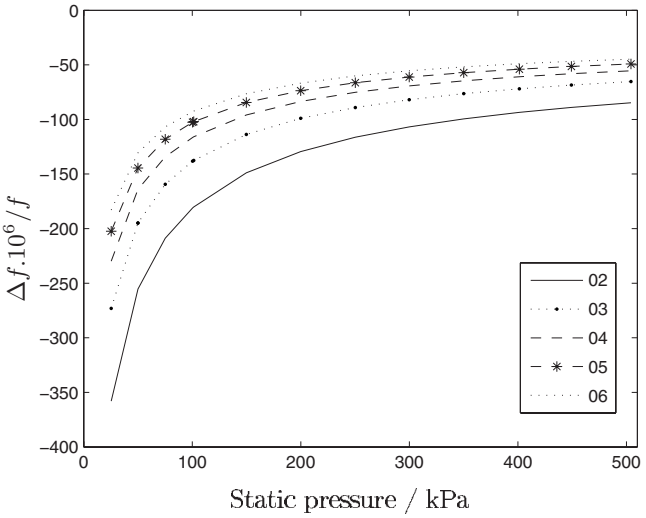


FIG. 6. Frequency shifts for radial modes from (0,2) to (0,6) as function of the static pressure.

According to the curves in Fig. 8, the contribution of modal coupling in the resonance frequencies of the radial modes considered here is significant in the higher static pressure range, but lower than 0.2×10^{-6} in relative values in the lower static pressure range, and then should not take a larger part in the extrapolation at zero static pressure of the speed of sound. However, this contribution should not be neglected regarding the final uncertainty required on the value of the Boltzmann constant. Here, the effect of modal coupling on the values of the resonance frequencies is nearly a linear function of the static pressure and does not depend significantly on the frequency. To interpret this, a few calculation on a simplified expression of Eq. (16) (neglecting the viscous and thermal effects) leads to expressing directly the effect of modal coupling on the resonance frequencies as a term proportional to $\rho_0 c_0^2 \chi \Delta S_r$ (χ being the constant compliance of the receiving microphone), which is a linear function of the static pressure and does not depend on the frequency.

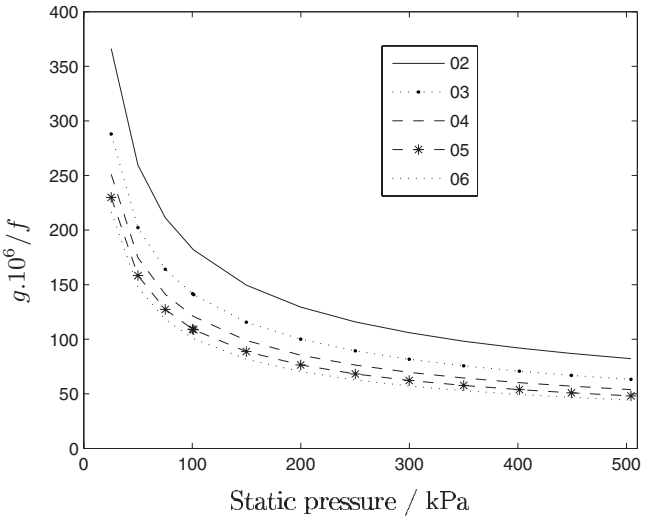


FIG. 7. Theoretical half-widths for radial modes from (0,2) to (0,6) as function of the static pressure.

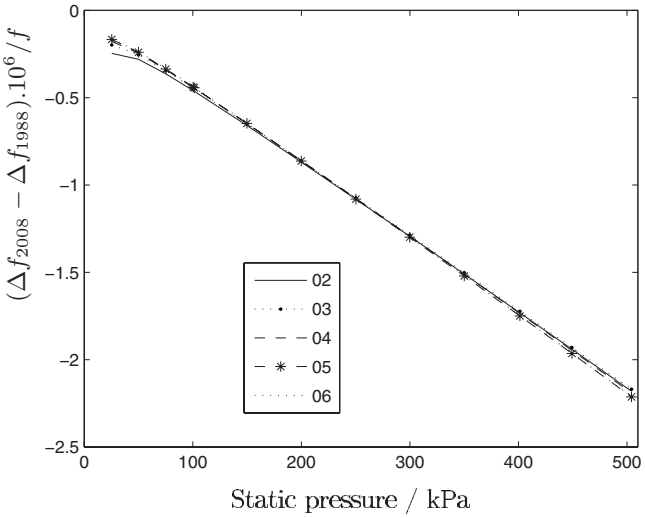


FIG. 8. Differences between the frequency shifts calculated according to actual and previous theories for radial modes from (0,2) to (0,6) as function of the static pressure.

The differences between the half-widths (Fig. 9) calculated from the two models confirm that the effect of the modal coupling in the resonance properties of radial modes should not be neglected in this spherical resonator.

Finally, the comparison between the theoretical half-widths calculated from the unified modeling of the acoustic field and the half-widths measured 20 years ago at the NIST (Ref. 2) is shown in Fig. 10 with the respective experimental deviations as function of the static pressure for acoustic radial modes from (0,2) to (0,6). For static pressures higher than 50 kPa, the discrepancies between theory and experiment are larger than the experimental deviations, showing that this model is not sufficient yet to take into account all the phenomena that occur in this spherical cavity. In the lower static pressure range, the experimental deviations are too large to conclude on the agreement between the analytical and the experimental results.

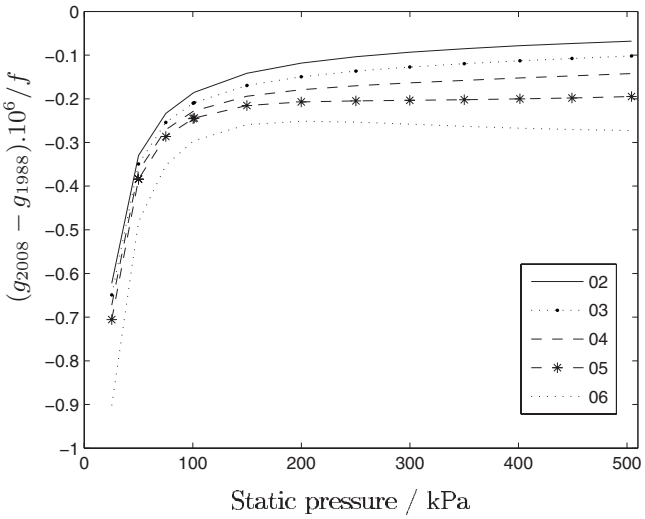


FIG. 9. Differences between the half-widths calculated according to actual and previous theories for radial modes from (0,2) to (0,6) as function of the static pressure.

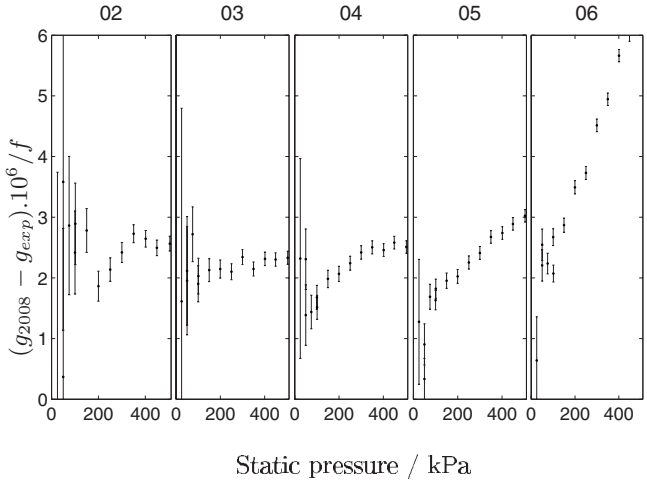


FIG. 10. Excess half-widths and experimental deviations for radial modes from (0,2) to (0,6).

V. CONCLUSION

To sum up, the present paper provides some important improvements to the modeling of the acoustic field in a quasi-spherical cavity with irregular wall (Sec. III) because the model given here allows us to take into account all the different perturbing elements together: the viscous and thermal effects in the boundary layers and in the bulk of the gas, the shape of the cavity (quasi-spherical shape), small elements located on the wall of the cavity (ducts, acoustic transducers, and small geometrical irregularities), the roughness of the wall, and the coupling between the fluid and the shell. This unified model includes the coupling between the different perturbing elements depending on the relative position of these elements on the wall for the small local components, with respect to the real shape of the cavity, and the resulting modal coupling in a consistent manner.

Regarding the first results obtained with this model on the particular case of the spherical cavity used at the NIST 20 years ago,[2] it seems that the contribution of modal coupling in the resonance frequency is small but should not be neglected considering the accuracy required here. Moreover, the discrepancies between the theoretical and experimental half-widths show that several phenomena are still missing in this unified model and should be taken into account in the future.

This work has to be carried out with the accurate modeling of each small element located on the wall of the cavity, especially the acoustic transducers, thin slits (undesirable but unavoidable in practice), which could be modeled as local roughness, and the coupling between the fluid and the shell. All these perturbing elements would be taken into account in the global modeling suggested here in terms of source or specific admittance on the wall.

Another improvement would be the modeling of the viscous and thermal effects at a higher order, which would avoid separating the effects in the boundary layer and in the bulk, taking into account low density gas effects (thermal jump and slip conditions), and integrating it in the unified modeling of the acoustic field in the cavity.

APPENDIX A: ACOUSTIC PROPAGATION IN DISSIPATIVE FLUID

1. Propagation equations

The particle velocity field $\boldsymbol{v}$ can be written as the sum of a vortical velocity $\boldsymbol{v}_v$ and a laminar velocity $\boldsymbol{v}_l$:

$$\boldsymbol{v} = \boldsymbol{v}_v + \boldsymbol{v}_l, \tag{A1a}$$

$$\mathbf{rot}\, \boldsymbol{v}_l = \mathbf{0} \quad \text{and} \quad \operatorname{div} \boldsymbol{v}_l \neq 0, \tag{A1b}$$

$$\operatorname{div} \boldsymbol{v}_v = 0 \quad \text{and} \quad \mathbf{rot}\, \boldsymbol{v}_v \neq \mathbf{0}. \tag{A1c}$$

Neglecting the coupling between the vortical and laminar movements in the bulk of the fluid, Navier–Stokes equation (1) is written as two equations, and then the fundamental equations (1)–(3) lead straightforwardly to the propagation equations for p, τ, $\boldsymbol{v}_l$, and $\boldsymbol{v}_v$.

The calculation of the propagation equation for the temperature variation τ is given in Ref. 11 (Sec. 2.5.1), neglecting the terms of order greater than the first order of the viscous and thermal lengths. The temperature variation τ can be written as the sum of an acoustic temperature τ_a and an entropic temperature τ_h, respectively, solutions of the two following equations:

$$\left[\frac{1}{c^2}\frac{\partial^2}{\partial t^2} - \left(1 + \ell_{vh}\frac{1}{c_0}\frac{\partial}{\partial t}\right)\Delta\right]\tau_a \approx 0, \tag{A2}$$

where $\ell_{vh} = \ell_v + (\gamma - 1)\ell_h$,

$$\left[\frac{1}{c_0}\frac{\partial}{\partial t} - \ell_h\Delta\left(1 + (\gamma - 1)(\ell_v - \ell_h)\frac{1}{c_0}\frac{\partial}{\partial t}\right)\right]\tau_h \approx 0. \tag{A3}$$

The pressure variation $p = p_a + p_h$ and the laminar particle velocity $\boldsymbol{v}_l = \boldsymbol{v}_a + \boldsymbol{v}_h$ satisfy the same equations (A2) and (A3) than the temperature variation τ.

The propagation equation for the vortical particle velocity is obtained, making use of Eqs. (A1a)–(A1c) in the Navier–Stokes equation (1):

$$\left[\frac{1}{c_0}\frac{\partial}{\partial t} - \ell'_v\Delta\right]\boldsymbol{v}_v = \mathbf{0}. \tag{A4}$$

2. General solutions

The solutions for the pressure variation $p = p_a + p_h$ and the particle velocity $\boldsymbol{v}_l = \boldsymbol{v}_a + \boldsymbol{v}_h$ can be derived from the solution for the temperature variation τ. Writing out the solution $\tau = \tau_a + \tau_h$ in Eq. (3), making use of Eqs. (A2) and (A3) leads to

$$p_a \approx \frac{\gamma\hat{\beta}}{\gamma - 1}\left(1 - \ell_h\frac{1}{c_0}\frac{\partial}{\partial t}\right)\tau_a, \tag{A5a}$$

$$p_h \approx \gamma\hat{\beta}(\ell_v - \ell_h)\frac{1}{c_0}\frac{\partial}{\partial t}\tau_h \ll p_a. \tag{A5b}$$

Writing out Eqs. (A5a) and (A5b) in Eq. (1) and making use of Eqs. (A1)–(A3) gives for the particle velocities $\boldsymbol{v}_a$ and $\boldsymbol{v}_h$:

$$\boldsymbol{v}_a \approx -\frac{1}{\rho_0 c_0}\frac{\gamma\hat{\beta}}{\gamma - 1}\left[\left(\frac{1}{c_0}\frac{\partial}{\partial t}\right)^{-1} + (\ell_v - \ell_h)\right]\mathbf{grad}\, \tau_a, \tag{A6a}$$

$$\boldsymbol{v}_h \approx \frac{\gamma\hat{\beta}}{\gamma - 1}\ell_h\, \mathbf{grad}\, \tau_h. \tag{A6b}$$

Equations (A2) and (A6b) can also be written for harmonic motion $e^{j\omega t}$ making use of the three following wavenumbers:

$$k_a^2 \approx \frac{k_0^2}{1 + jk_0\ell_{vh}} \approx k_0^2[1 - jk_0\ell_{vh}], \tag{A7}$$

$$k_h^2 \approx -\frac{jk_0}{\ell_h}[1 - jk_0(\gamma - 1)(\ell_v - \ell_h)], \tag{A8}$$

$$k_v^2 = -\frac{jk_0}{\ell'_v}. \tag{A9}$$

Assuming that $p_a \gg p_h$, the acoustic pressure p_a, denoted by p in the present paper, is governed by the same propagation equation (A2) as the acoustic temperature variation τ_a. This propagation equation should be associated to adapted boundary conditions to describe the acoustic field in the cavity.

APPENDIX B: NORMALIZATION COEFFICIENTS

The normalization coefficients $\mathcal{J}_{\nu\eta}$ of the spherical Bessel functions $j_\nu(k_{\eta\nu}r)$ [Eq. (21)] are given by the orthonormality condition on the Bessel functions:

$$\mathcal{J}_{\nu\eta}^2 = -\int_0^R [j_\nu(k_{\eta\nu}r)]^2 r^2 dr,$$

yielding

$$\mathcal{J}_{\nu\eta}^2 = \frac{R^3}{2}, \tag{B1a}$$

for $\nu = 0$ and $\eta = 1$,

$$\mathcal{J}_{\nu\eta}^2 = \frac{R^3}{2}[j_\nu(\gamma_{\eta\nu})]^2\left[1 - \frac{\nu(\nu + 1)}{\gamma_{\nu\eta}^2}\right], \tag{B1b}$$

otherwise.

The normalization coefficients $\mathcal{J}_{\nu\eta}$ of the spherical Bessel functions $j_\nu(k_{\eta\nu}r)$ [Eq. (6)] are given by the orthonormality condition on the eigenfunctions $\psi_m^t(\boldsymbol{r})$ in the regular domain D_r:

$$a_m^2 = \int\int\int_{D_r} [\psi_m^t(\boldsymbol{r})]^2 d\boldsymbol{r},$$

leading to

$$a_m^2 = \tfrac{4}{3}\pi R^3, \tag{B2a}$$

for $\nu = 0$ and $\eta = 1$,

$$a_m^2 = (1 + \delta_{\mu 0})\pi R^3 [j_\nu(\gamma_{\eta\nu})]^2 \times \left[1 - \frac{\nu(\nu+1)}{\gamma_{\nu\eta}^2}\right] \frac{1}{2\nu+1} \frac{(\nu+\mu)!}{(\nu-\mu)!}, \quad \text{(B2b)}$$

otherwise.

[1] I. M. Mills, P. J. Mohr, T. J. Quinn, B. N. Taylor, and E. R. Williams, "Redefinition of the kilogram, ampere, kelvin and mole: A proposed approach to implementing CIPM recommendation 1 (CI-2005)," Metrologia **43**, 227–246 (2006).

[2] M. R. Moldover, J. P. M. Trusler, T. J. Edwards, J. B. Mehl, and R. S. Davis, "Measurement of the universal gas constant R using a spherical acoustic resonator," J. Res. Natl. Bur. Stand. **93**, 85–144 (1988).

[3] J. B. Mehl, "Acoustic resonance frequencies of deformed spherical resonators," J. Acoust. Soc. Am. **71**, 1109–1113 (1982).

[4] J. B. Mehl, "Acoustic resonance frequencies of deformed spherical resonators 2," J. Acoust. Soc. Am. **79**, 278–285 (1986).

[5] M. R. Moldover, J. B. Mehl, and M. Greenspan, "Gas-filled spherical resonators: Theory and experiment," J. Acoust. Soc. Am. **79**, 253–272 (1986).

[6] J. B. Mehl, M. R. Moldover, and L. Pitre, "Designing quasi-spherical resonators for acoustic thermometry," Metrologia **41**, 295–304 (2004).

[7] J. P. M. Trusler, *Physical Acoustics and Metrology of Fluids* (Higler, Bristol, 1991).

[8] J. W. S. Rayleigh, *The Theory of Sound*, 2nd ed. (Dover, New York, 1896).

[9] P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (McGraw-Hill, New York, 1968).

[10] A. D. Pierce, *Acoustics: An Introduction to Its Physical Principles and Applications* (McGraw-Hill, New York, 1981).

[11] M. Bruneau and T. Scelo, *Fundamentals of Acoustics* (ISTE, London, 2006).

[12] K. S. Sum and J. Pan, "Effects of the inclination of a rigid wall on the free vibration characteristics of acoustic modes in a trapezoidal cavity," J. Acoust. Soc. Am. **119**, 2201–2210 (2006).

[13] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953).

[14] J. B. Mehl, "Spherical acoustic resonators: Effect of shell motion," J. Acoust. Soc. Am. **78**, 782–788 (1985).

Photoacoustic tomography with a single detector in a reverberant cavity

B. T. Cox[a] and P. C. Beard
Department of Medical Physics and Bioengineering, University College London, Gower Street, London WC1E 6BT, United Kingdom

(Received 27 August 2008; revised 3 December 2008; accepted 11 December 2008)

In conventional biomedical photoacoustic tomography (PAT), ultrasonic pulses generated through the absorption of nanosecond pulses of near-infrared light are recorded over an array of detectors and used to recover an image of the initial acoustic pressure distribution within soft tissue. This image is related to the tissue optical coefficients and therefore carries information about the tissue physiology. For high resolution imaging, a large-area detector array with a high density of small, sensitive elements is required. Such arrays can be expensive, so reverberant-field PAT has been suggested as a means of obtaining PAT images using arrays with a smaller number of detectors. By recording the reflections from an acoustically reverberant cavity surrounding the sample, in addition to the primary acoustic pulse, sufficient information may be captured to allow an image to be reconstructed without the need for a large-area array. An initial study using two-dimensional simulations was performed to assess the feasibility of using a *single* detector for PAT. It is shown that reverberant-field data recorded at a single detector are sufficient to reconstruct the initial pressure distribution accurately, so long as the shape of the reverberant cavity makes it ray-chaotic. The practicalities of such an approach to photoacoustic imaging are discussed.
 [DOI: 10.1121/1.3068445]

PACS number(s): 43.35.Ud, 43.20.Ks, 43.60.Pt, 43.28.We [TDM] Pages: 1426–1436

I. INTRODUCTION

In the past decade or so, the potential of photoacoustic tomography[1] (PAT) to become a useful biomedical imaging tool has been clearly demonstrated.[2–6] For imaging soft-tissue to depths where the effect of optical scattering becomes too great for "ballistic" optical imaging modalities, e.g., optical microscopy or optical coherence tomography, PAT provides a means of obtaining an image based on optical contrast but still with good resolution (typically <100 μm at depths of 5–10 mm). It is a hybrid technique in the sense that *optical* pulses are used to generate *ultrasonic* pulses within the tissue. The absorption of short (nanosecond) laser pulses by tissue chromophores results in a small but rapid increase in the local pressure, which then propagates as a high frequency ultrasonic pulse (tens of megahertz). These ultrasonic pulses are recorded by a detector array and, via a numerical reconstruction algorithm, used to form an image.

The spatially varying pressure rise caused when the light energy is absorbed is called the "initial pressure distribution." It is this that a PAT image represents, and it is proportional to the optical energy per unit volume absorbed by the tissue. As this is closely related to the optical absorption coefficient, PAT images are related to the distribution of the chromophores contained in the tissue and carry information about the tissue structure and function. Because of this reliance on optical absorption, PAT has the potential to be used spectroscopically, so could be used in applications such as molecular imaging, in which an image of the distribution of endogenous or exogeneous chromophores with known optical absorption spectra is obtained.[7–10]

Current experimental PAT systems use arrays of detectors to collect data over a sufficiently large measurement surface to form an image. This becomes expensive as the demand for better resolution, and therefore larger arrays, increases. In addition, for a real time imaging system, each of the detector array elements requires dedicated electronics, further increasing the cost. For targets that can be enclosed, or partially enclosed, by an acoustically reverberant cavity, it has been shown that by using the information stored in the reverberant field, the size of the measurement surface (and therefore the number of detection points) can be reduced without losing image quality.[11] This paper suggests taking this approach one step further: using a single point measurement of the reverberation of a photoacoustically generated signal within an acoustically reverberant cavity to form a PAT image. This is applicable only to targets that can be enclosed by the reverberant cavity, but avoids both the use of a large and expensive array and point-by-point scanning. Two-dimensional numerical examples are used to demonstrate how good quality PAT images could be obtained from measurements made using just a single ultrasound detector in a chaotic cavity, and the advantages, practicalities, and limitations of such a scheme are discussed.

II. PHOTOACOUSTIC IMAGE RECONSTRUCTION

Consider a short laser pulse incident on some optically heterogeneous biological tissue (such as skin). A region of slightly increased pressure, $p_0(\mathbf{x})$, is generated locally wher-

[a]Electronic mail: bencox@mpb.ucl.ac.uk

ever the optical energy is absorbed (such as in a blood vessel). When the laser pulse duration is much shorter than the acoustic travel time across a characteristic distance of p_0 (such as the vessel diameter), the subsequent acoustic propagation can be modeled as an initial value problem for the wave equation with the initial conditions $p(\mathbf{x})|_{t=0}=p_0(\mathbf{x})$ and $dp/dt|_{t=0}=0$. In conventional PAT, an image of $p_0(\mathbf{x})$ is reconstructed from measurements of the acoustic pressure waves $p(\mathbf{x}_s,t)$ made by detectors placed at points $\mathbf{x}_s$ on a measurement surface S enclosing V, where $\text{supp}(p_0) \in V$. (Usually the detectors are assumed to be much smaller than the shortest acoustic wavelength in the pulse, and omnidirectional, which is never quite the case in practice.)

In the past decade or so, numerous algorithms to recover $p_0(\mathbf{x})$ from $p(\mathbf{x}_s,t)$ have been proposed. For recent reviews see Refs. 5 and 12. The algorithms differ in the measurement geometries assumed (e.g., spherical, cylindrical, planar, and arbitrary), in the types of solution (e.g., closed form, infinite series, and numerical), and in their computational efficiency. However, all the algorithms make the simplifying assumption that the measurement surface is acoustically transparent and does not affect the free-field propagation of the photoacoustically generated waves. (Experimentally this assumption has been accommodated by truncating the measurements before any reflections, e.g., from other parts of the measurement apparatus, are received, i.e., by time-gating out any reflections.) The image reconstruction described in this paper is quite different in the sense that the free-field assumption is not made, and the reflected or reverberant field is included in the formulation.

A. Limited-view problem

Experimental imaging studies have been carried out using circular, cylindrical, spherical, and planar measurement surfaces. With a spherical measurement surface enclosing V it is possible to record the photoacoustic waves traveling away from the source in all directions. With a planar measurement surface at most half of the waves can be recorded, as a plane cannot surround the source region. Nevertheless, if the plane is infinite in extent this is still sufficient information to recover the source exactly, as shown by Anastasio *et al.*[13] In practice, an actual planar measurement surface will have a finite size, and not even half of the emitted waves can be recorded. This loss of information means that there are insufficient data to reconstruct the image exactly. The approximation of an infinite or closed measurement surface by one of limited extents—the "limited view," "partial scan," or "finite aperture" problem—therefore leads to artifacts and distortion in the reconstructed image. [It has been shown[14–16] that the edge of a source region (an abrupt change in absorbed energy density) can be reconstructed stably only if the normal to the edge crosses the measurement surface. For a spherical measurement surface enclosing V this is readily achievable, but, as hinted at above, for a single planar measurement surface it is necessary for it to extend to infinity to fulfill this requirement.] In practice, image reconstruction algorithms designed with a complete measurement surface in mind, e.g., a sphere or an infinite plane, are applied to partial data sets with the missing information implicitly replaced by zeros, even though such an extension of the data is clearly unphysical. The use of range conditions to constrain the extension of the data to be more realistic has been proposed,[17] but this does not overcome the fundamental lack of data.

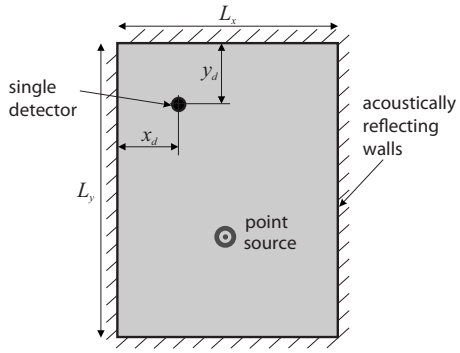


FIG. 1. A rectangular reverberant cavity with a single detector at point (x_d, y_d) and a point source.

B. Reverberant-field PAT

When approximating an infinite planar measurement surface by one of finite extent, one way to ameliorate the limited-view problem is to reflect the sound back onto the measurement surface by placing acoustically reflecting walls perpendicular to, and at the ends of, the planar measurement aperture. The effect of this is to introduce an infinite, periodically repeating array of acoustic image sources. The data can now be extended periodically so that $p(\mathbf{x}_s,t)$ is known over the entire infinite measurement plane, even though it was only recorded over a finite region. An efficient reconstruction algorithm that exploits the periodicity exists for this case.[11,18] In effect, the information that would have been lost is retained in the reverberant field set up between the reflecting walls.

This use of a reverberant acoustic field for imaging is in contrast to most image reconstruction algorithms, which, as mentioned above, suppose that the measurement surface sits within an acoustic free-field, i.e., the waves emanating from the photoacoustic source p_0 travel outward through the measurement surface, unimpeded and unaffected by either the measurement surface itself or any other obstacles. The use of acoustic reflectors, in addition to being one way to tackle the limited-view problem, suggests that smaller detector arrays could replace larger ones without sacrificing the quality of the images, if reverberation can be used to replace the missing data. The question then naturally arises as to how far this idea can be taken. For instance, if p_0 is restricted to a box with reflecting walls, as shown in Fig. 1, is it possible to reconstruct p_0 from measurements of the reverberant field at a single point?[19]

III. RECTANGULAR CAVITY BACKPROJECTION

As a first step, consider the simple scenario: can the position of a single point source (which emits one impulse) be reconstructed from the times at which the pulse and its reflections arrive at a single detector? For a rectangular cav-

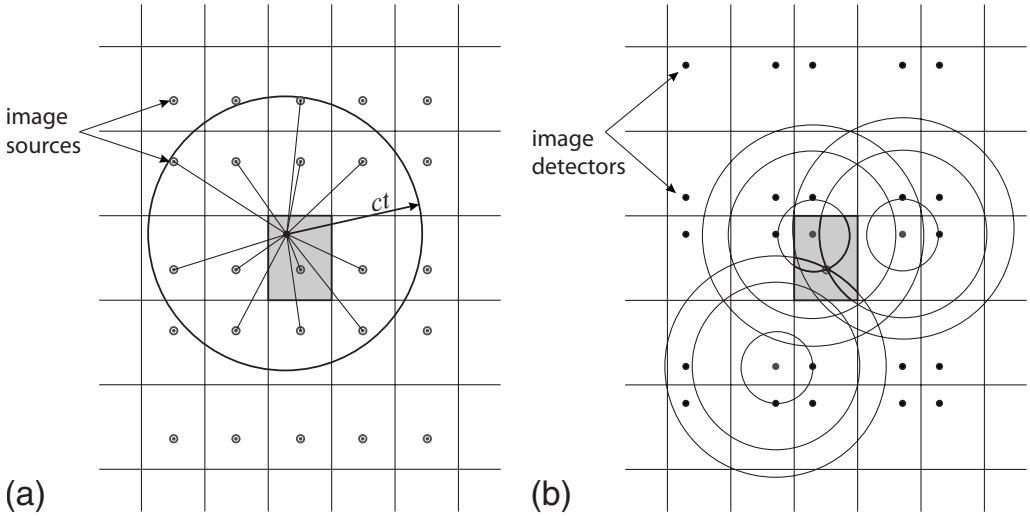


FIG. 2. Image sources and detectors. The shaded region is the actual rectangular cavity; the other rectangles are images. (a) Image sources. Those within a circle of radius ct have contributed to the signal measured at the detector by time t, where c is the sound speed. Rays are shown connecting the image sources to the omnidirectional detector, which records the times of arrivals of the pulses but not their directions. (b) Image detectors. These can be useful when reconstructing the position of a point source. The circles illustrate the backprojection of three impulses received at the detector for three of the image sources, showing that they coincide at the source point.

ity, the reflections from the walls can be represented by image sources, which make the calculations of the arrival times straightforward and provide a simple geometrical way to understand the problem. [Two-dimensional examples are used throughout this paper for ease of presentation and in order to limit the calculation times, but the same principles apply in three dimensions (3D).]

Figure 1 shows a 15 mm $\times$ 20 mm cavity with a single point source and a single point omnidirectional detector, and Fig. 2(a) illustrates the idea of image sources due to the single point source within the reverberant cavity. The image sources within a circle of radius ct, where c is the speed of sound, contribute to the signal measured at the detector up to time t. If the source emits a pulse at time $t=0$, then the times of arrival of the pulse and its reflections can be straightforwardly calculated by using this notion of image sources: each image source supplies one pulse at time (r/c), where r is the distance between the detector and the image source. The effect of geometrical spreading on the waves was included by reducing the amplitude by a factor of $1/\sqrt{r}$ (in 3D this would be $1/r$). At this initial stage no attempt has been made to model the photoacoustic wave propagation more realistically than this, although it is fully modeled in the other examples given below and in Sec. IV. Figure 3 shows the impulse train $p(t)$ that would be recorded by a detector positioned in the corner, $(x_d, y_d)=(0\ \text{mm}, 20\ \text{mm})$, due to a point source at (8 mm, 7 mm). (There are three times at which two reflections have coincided exactly and therefore have twice the amplitude of the neighboring pulses.)

In Fig. 2(a), rays connecting the image sources to the detector have been drawn. Of course, the image sources emit rays in all directions, but it is clear from Fig. 2(a) that the direction in which the *detected* ray leaves its image source can, and usually will, be different for different image sources. This implies that many of the normals to the edge of the source will reach a single measurement point sooner or later. Given that the normals to the edges of a source are required to cross the measurement surface for that edge to be stably reconstructed, as mentioned in Sec. II A, this suggests that it might be possible to reconstruct the edges of the source accurately from a measurement at a single detector.

By considering the symmetries of the cavity in Fig. 1, it is clear that for some detector positions it is not possible to determine the point source uniquely, as point sources at two different positions can give identical pulse trains. For example, if the detector lies on any axis of reflection or rotation symmetry then it is not possible to associate the pulse train with a unique source. However, if the detector is placed away from these symmetry axes, then it is possible to locate the point source uniquely. To demonstrate this, the concept of image detectors—the detector equivalents of image sources—will be introduced.

The pulse train in Fig. 3 was calculated by determining the distance (and therefore propagation time) between the detector and a set of image sources, as described above. Each image source represents one reflection, and therefore contributes one pulse to the pulse train. It could equally well have been generated by calculating the distance between the actual

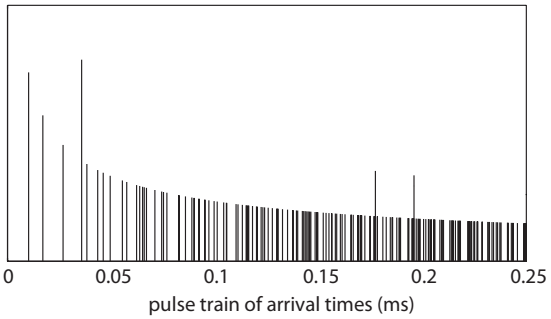


FIG. 3. The first 0.25 ms of the train of pulses received by the detector as calculated using image sources. The amplitude shows cylindrical spreading $\sqrt{ct}$ dependence (sound speed $c=1500$ m/s). On three occasions two pulses arrive at the detector simultaneously, hence the three pulses with larger amplitude.

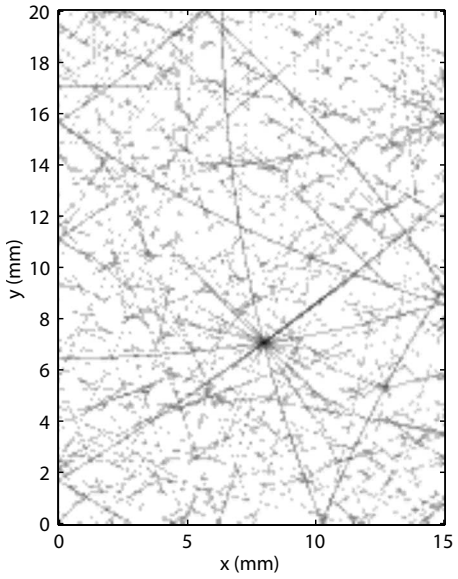


FIG. 4. Backprojecting the train of pulses from the nearest 500 image detectors shows that the point source can be located. The detector is located at the corner (0 mm, 20 mm) and the source is at (8 mm, 7 mm).

source position and a set of image *detectors*. The measured signal at each hypothetical image detector would consist of a single pulse, indicating the distance from the source. If these measured signals are denoted $p_i(t)$ for image detectors $i=1,\ldots,$ then the pulse train for the actual detector would be the summation of the individual signals, $p(t)=\Sigma_i p_i(t)$. Each image detector represents one reflection and, as with the image sources, contributes one pulse to the measured pulse train. If it were possible to measure the signals $p_i(t)$ individually, then the position and amplitude of the point source could be determined by summing the backprojections of each $p_i(t)$

$$p_0(\mathbf{x}) \approx \sum_{i=1}^{N} \left(\frac{\sqrt{|\mathbf{x}-\mathbf{d}_i|}}{N} \right) p_i(|\mathbf{x}-\mathbf{d}_i|/c), \tag{1}$$

where $\mathbf{d}_i$ is the position of the image detector i, and the term $\sqrt{|\mathbf{x}-\mathbf{d}_i|}$ was included to account for the cylindrical spreading. This will give the position and amplitude of the point source exactly in the limit $N\to\infty$. In practice, the set of signals $\{p_i(t);i=1,\ldots\}$ is not measured, but only its summation $p(t)$. Replacing $p_i(t)$ with $p(t)$ in Eq. (1) gives an approximate backprojection formula

$$p_0(\mathbf{x}) \approx \sum_{i=1}^{N} \left(\frac{\sqrt{|\mathbf{x}-\mathbf{d}_i|}}{N} \right) p(|\mathbf{x}-\mathbf{d}_i|/c). \tag{2}$$

Figure 2(b) illustrates the principle by showing the backprojection of three pulse arrival times from the actual detector position and two image detectors, and the three coincide only at the source point. The backprojection of the time series in Fig. 3, from the 25 image detectors nearest to the actual detector, gives the image in Fig. 4. It is clear that with only this single measurement, the source position can nevertheless be located accurately and uniquely. This is similar to single channel time-reversal imaging to locate point sources, which is discussed further in Sec. V.

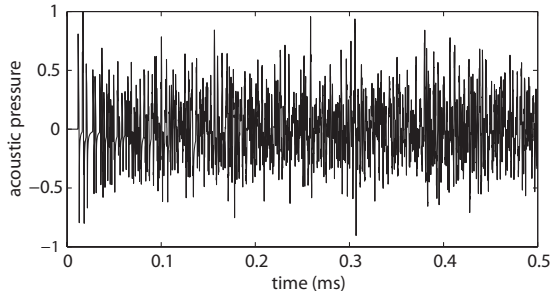


FIG. 5. The first part of a photoacoustic time series simulated from the initial pressure distribution in Fig. 6(A) for the point (0.03 mm, 19.96 mm).

Next, the same backprojection was used to reconstruct an estimate of the initial pressure distribution from a simulated *photoacoustic* time series "measurement" rather than just pulse arrival times. The "measured" signal in Fig. 5 was simulated using a *k*-space method[20] from the initial pressure distribution shown in Fig. 6(A). The detector was positioned close to the corner at (0.03 mm,19.96 mm), and 1% Gaussian noise was added. Figure 6(B) shows the results of backprojecting the 0.5 ms duration time series from the 500 image detector positions nearest to the actual detector. It is apparent that the main features of the initial pressure distributions have been recovered, but also that the image is contaminated by artifacts and noise. A second example, which will be compared to the modal reconstruction approach described in Sec. IV, is also shown in Fig. 7. (These images are placed later in the text to allow the comparison with the images generated using other techniques to be made easily.) The initial pressure distribution, Fig. 7(A), consists of nine small circular sources with Gaussian profiles which are just about located accurately in the image, Fig. 7(B), but the errors due to artifacts are so significant that the image quality is poor.

IV. MODAL RECONSTRUCTION

The backprojection reconstruction described in Sec. III can give at best a crude approximation to the true initial pressure distribution. Furthermore, it is only simple to implement when the reverberant cavity is a regular shape, such as a rectangle, because the positions of the image detectors can then be calculated straightforwardly. For a more accurate reconstruction, and one that could be applied to cavities of more general shape, a reconstruction based on the acoustic modes of a cavity was used. This has the advantage that it is based more closely on the physics of the system, it is exact (within the limits imposed by ill-posedness) and is generally applicable, in that mode shapes and eigenvalues can be calculated for a cavity of any shape.

A. General case

A photoacoustically generated acoustic pressure field at a point (x,y) in a reverberant cavity, at a time t, can be written as a sum of normal modes

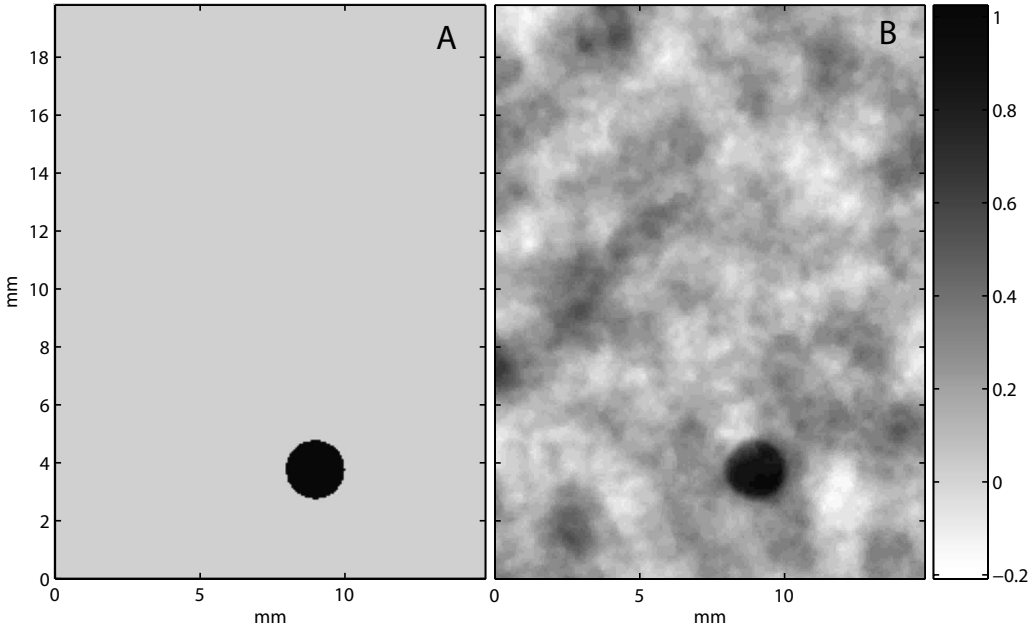


FIG. 6. (A) Circular initial pressure distribution used to generate the time series in Fig. 5. (B) Image formed by backprojecting the reverberant time series from the 500 image detector positions nearest to the actual detector. The circular initial pressure distribution has been recovered, but the image suffers from artifacts and noise.

$$p(x,y,t) = \sum_m A_m \phi_m(x,y)\cos(\omega_m t), \tag{3}$$

where subscript m indicates the mode with amplitude A_m, modal frequency ω_m, and normalized mode shape $\phi_m(x,y)$. In this notation, the initial pressure distribution is

$$p_0(x,y) = \sum_m A_m \phi_m(x,y), \tag{4}$$

and it is clear that $\partial p/\partial t|_{t=0}=0$, i.e., the particle velocity is initially zero, which is true for photoacoustic waves. By discretizing the continuous variables x, y, and t as x_n, y_n, and t_i, where $n=0,\ldots,N$ and $i=0,\ldots,T$, Eqs. (3) and (4) can be written in matrix-vector form

$$\mathbf{p} = M_t\mathbf{A}, \quad \mathbf{p}_0 = M_0\mathbf{A}, \tag{5}$$

where $\mathbf{p}=[p(x_d,y_d,t_0),\ldots,p(x_d,y_d,t_T)]^T$, $\mathbf{p}_0=[p_0(x_0,y_0),\ldots,p_0(x_N,y_N)]^T$, and $\mathbf{A}=[A_0,\ldots,A_M]^T$. M is the total number of modes that are included in the reconstruction. Each column of the $N\times M$ matrix M_0 contains the shape of a single mode m at all the points (x_n,y_n), $n=0,\ldots,N$

$$M_0(n,m) = \phi_m(x_n,y_n), \tag{6}$$

and each column of the $T\times M$ matrix M_t contains the values of a single mode m at the detector position (x_d,y_d) at all the time points t_i, $i=0,\ldots,T$

$$M_t(i,m) = \phi_m(x_d,y_d)\cos(\omega_m t_i). \tag{7}$$

The matrices M_0 and M_t can be generated in advance from knowledge of the mode shapes and frequencies of the cavity. For simple geometries these may be known analytically; otherwise it might be necessary to calculate them numerically using the finite element (FE) method, for instance. (In which case it would also be possible to take into account known variations in the acoustic parameters within the cavity, such as sound speed heterogeneities.)

From Eq. (5) it is clear that an approximation to $\mathbf{p}_0$ can be found from the measured time series, $\mathbf{p}$, in two steps: first, estimate the modal amplitudes $\mathbf{A}$ by inverting the first equation,

$$\mathbf{A} = M_t^{-1}\mathbf{p}, \tag{8}$$

and then use the second equation to obtain an estimate for $\mathbf{p}_0$. Written as one equation these two operations are

$$\mathbf{p}_0 = M_0 M_t^{-1}\mathbf{p}, \tag{9}$$

which could also be written as

$$\mathbf{p}_0 = G_{\text{reverb}}^{-1}\mathbf{p}, \tag{10}$$

where the columns of the Green's function matrix G_{reverb} contain the time series (impulse response functions) that would be recorded at (x_d,y_d) due to impulses leaving the points (x_n,y_n), $n=0,\ldots,N$, at time $t=0$. For simple geometries G_{reverb} can be found analytically; when the boundary is more complex, G_{reverb} could in principle be constructed column by column from time domain calculations of the impulse response functions. Clearly, $G_{\text{reverb}}^{-1}=M_0M_t^{-1}$.

Equations (9) and (10) show that the success of this imaging approach depends on the invertibility of the matrix M_t (or G_{reverb}). This in turn depends on the geometry of the cavity, as will be shown below by comparing the reconstruction using a rectangular cavity with that in a chaotic cavity (a quarter Sinai billiard).

B. Rectangular cavity

For a rectangular cavity, the mode shapes are separable in x and y and take on simple forms, so Eq. (3) becomes

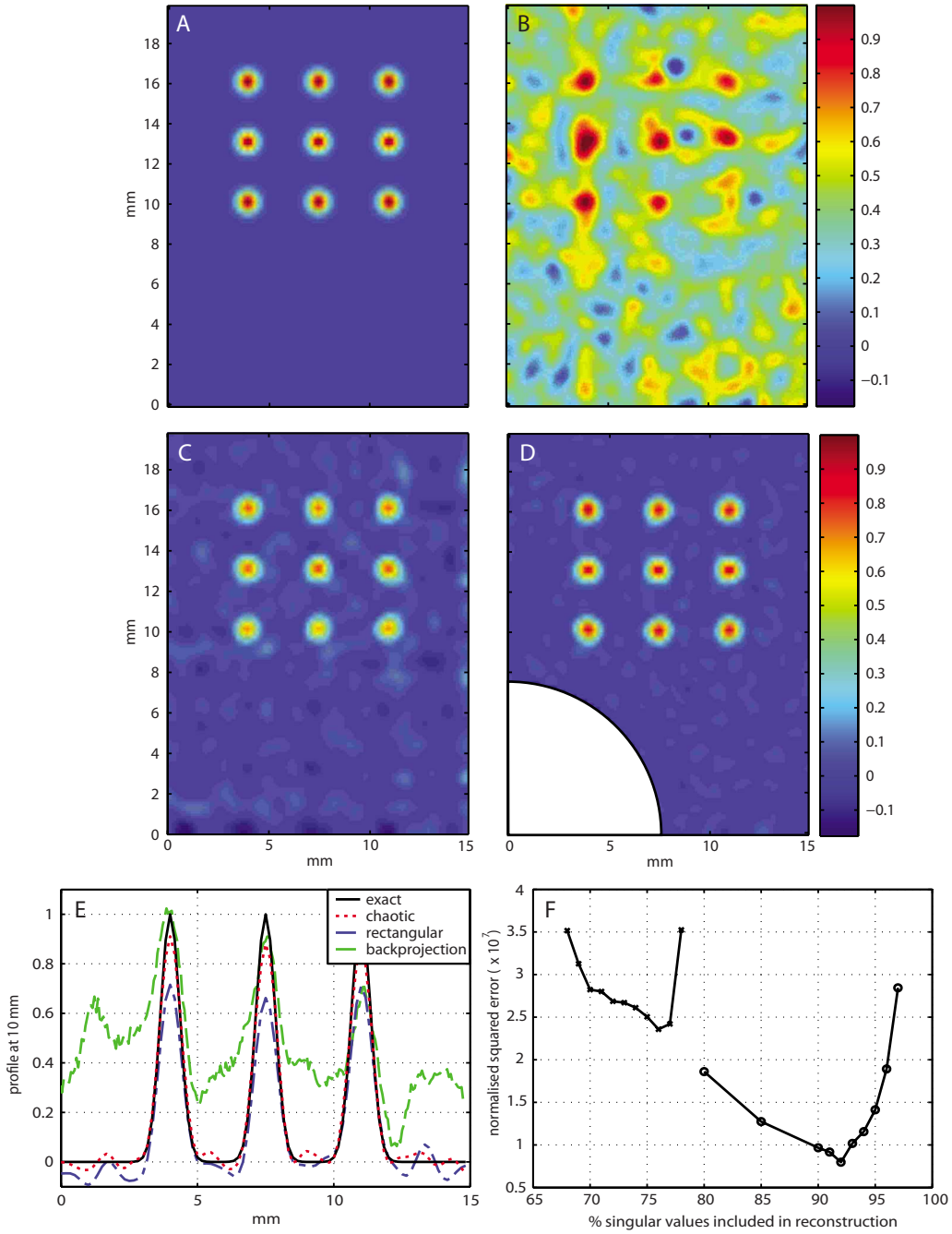


FIG. 7. (Color online) (A) Initial pressure distribution, p_0, consisting of nine small circles with Gaussian profiles. (B) Image of p_0 obtained by backprojection using Eq. (2). (C) Image of p_0 obtained by a modal reconstruction in a rectangular cavity. (D) Image of p_0 obtained by a modal reconstruction in a chaotic cavity. (E) Profiles at $y=10$ mm through (A), (B), (C), and (D), corresponding to the exact p_0 (solid line), backprojection (dashed), modal reconstruction with rectangular cavity (dot-dashed), and chaotic cavity (dotted). (F) Mean squared error as a function of truncation order for the images from rectangular (+) and chaotic (○) cavities.

$$p(x,y,t) = \sum_{k,l} A_{kl} \cos(k\pi x/L_x)\cos(l\pi y/L_y)\cos(\omega_{kl}t), \quad (11)$$

where $k=0,\ldots,K$, $l=0,\ldots,L$, and each pair $\{k,l\}$ corresponds to a single mode with amplitude A_{kl}, modal frequency

$$\omega_{kl} = c\sqrt{(k\pi/L_x)^2 + (l\pi/L_y)^2}, \quad (12)$$

and mode shape $\cos(k\pi x/L_x)\cos(l\pi y/L_y)$. Similarly, Eq. (4) becomes

$$p_0(x,y) = \sum_{k,l} A_{kl} \cos(k\pi x/L_x)\cos(l\pi y/L_y). \quad (13)$$

Each column of the $N\times KL$ matrix M_0 contains, therefore, the shape of a single mode $\{k,l\}$ at all the points (x_n,y_n):

$$M_0(n,\{k,l\}) = \cos\left(\frac{k\pi x_n}{L_x}\right)\cos\left(\frac{l\pi y_n}{L_y}\right), \quad (14)$$

and each column of the $T\times KL$ matrix M_t contains the values of a single mode $\{k,l\}$ at the detector position (x_d,y_d) at the time points $[t_0,\ldots,t_T]$:

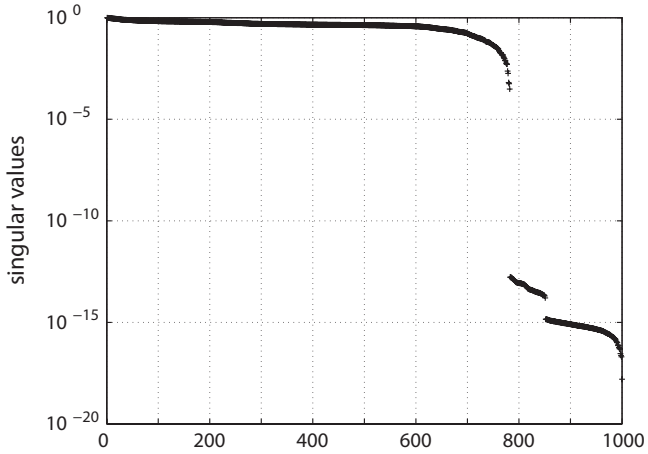


FIG. 8. Singular values of matrix M_t for the rectangular cavity.

$$M_t(i,\{k,l\}) = \cos\left(\frac{k\pi x_d}{L_x}\right)\cos\left(\frac{l\pi y_d}{L_y}\right)\cos(\omega_{kl}t_i). \qquad (15)$$

This modal reconstruction approach was used to estimate the initial pressure distribution shown in Fig. 7(A) from the same single time series measurement (not shown) that was used to reconstruct Fig. 7(B). The $T\times 1$ vector of measured data, **p**, included 1% Gaussian noise and consisted of $T=1600$ equally spaced samples with $t_{\max}=0.16$ ms, and the 15 mm $\times$ 20 mm reconstruction mesh consisted of $N=90\times 90=8100$ rectangular pixels. Matrix M_t was constructed with the 1000 lowest frequency modes, and its singular value decomposition (SVD) was calculated using MATLAB's SVD function, which uses LAPACK's DGESVD routine. The singular value spectrum is shown in Fig. 8. The physical significance of the very small singular values will be discussed below, but numerically they make M_t ill-conditioned, and therefore its (pseudo)inverse is very sensitive to noise in the data. Regularization must be included to suppress the noise, and here a truncated-SVD inversion was used. The mean squared error between the estimated and true images is shown, again with + symbols, in Fig. 7(F) as a function of the truncation order. Figure 7(C) shows the image with the truncation order set to 770 singular values. This reconstruction is much more accurate than the backprojection, Fig. 7(B), but it still used less than 80% of the available modes, as many had to be discarded in order to regularize the solution. If a way could be found to improve the conditioning of the matrix M_t, so that more modes can be included in the inversion, it should be possible to improve the accuracy of the image further.

C. Modal degeneracy

The columns of matrix M_t are given by Eq. (7), and each one is a scaled cosine oscillating with modal frequency ω_m. For M_t to be invertible it is necessary that the columns are linearly independent. This can only be the case when no two modal frequencies are the same, i.e., $\omega_m \neq \omega_n \ \forall m,n$. In the case of a rectangular cavity, the modal frequencies are given by Eq. (12), so many modes have the same modal frequency, and the condition does not hold. This modal degeneracy gives rise to the practically zero singular values in the spectrum shown in Fig. 8 and thus to the noninvertibility of M_t. In order to improve the conditioning of M_t, therefore, it is necessary to choose a cavity shape that minimizes the number of modes with the same modal frequency. Cavity shapes with this property are already known from the study of waves in reverberant cavities known as quantum or wave chaos.[21–23] Indeed, the eigenvalues for cavities of different shapes have been studied in some depth, and it has been shown that the distributions of the spacings between the eigenvalues (or modes) take on quite precise and universal forms which depend on the shape of the cavity. Figure 9(a) is a histogram showing the distribution of the spaces between the lowest 4000 modal frequencies in a 20 mm $\times$ 15.1 mm rectangular cavity as a function of normalized modal spacing, s (modal spacing/number of modes included). This distribution agrees closely with Poisson's distribution $\exp(-s)$, also shown, as is expected for integrable cavities such as this. (An integrable cavity is one in which Hamilton's equations describing the cavity's ray dynamics form an integrable system.) It is notable that this distribution, which applies universally to any cavity of this type, has its peak at a modal spacing of zero. This is of practical significance because it means that there will be many degenerate modal frequencies, and therefore imperfect image reconstruction, for any inte-

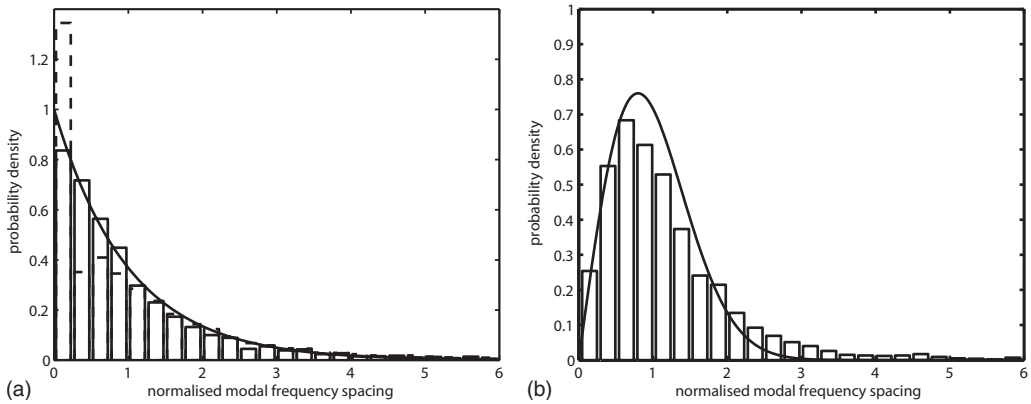


FIG. 9. Histograms showing the distributions of the spaces between the lowest 4000 modal frequencies as a function of normalized modal spacing, s (modal spacing/number of modes included). (a) For a 20×15.1 mm^2 rectangular cavity, the distribution agrees closely with Poisson's distribution $\exp(-s)$ (solid line). (b) For a chaotic cavity (a quarter Sinai billiard), the distribution agrees closely with Wigner's distribution $(\pi/2)s\exp(-(\pi s^2/4))$ (solid line).

grable cavity. (In fact, it is even worse than this for the case investigated in Sec. IV B. When the sides of a rectangular cavity are in simple ratios there is an even larger peak close to zero spacing.[23] The cavity used in Sec. IV B was 20 mm × 15 mm, not 20 mm × 15.1 mm, and the resulting increase in degeneracy is shown in Fig. 9(a) with dotted lines.) If the modal frequencies were spaced so that no two modes share the same modal frequency, and degeneracy is thereby reduced, then the columns of M_t would be linearly independent and more of the modes can be used in the reconstruction, leading to more accurate imaging. That this is true for chaotic cavities is shown in Fig. 9(b) and discussed below.

D. Chaotic cavity

Although the description given for the acoustic field in Sec. IV A was in terms of modes, ray descriptions can also be helpful and intuitive. According to this model, rays are emitted from a source, reflect specularly from the walls of the reverberant cavity, and contribute to the detected signal when they pass through the measurement point. Consider rays emitted in all directions from a source. For some cavity geometries some of these rays will never reach the detector. They may, for instance, follow a periodic orbit around the cavity never encountering the measurement point if it is not on their orbit. The rectangular cavity used here is one example of a type of cavity in which this can happen. It seems intuitive that if some of the rays from a point do not encounter the detector, then that point cannot be reconstructed exactly from the measured data. It seems clear, when thinking in terms of rays, that what is required is a cavity in which every ray travels through every point in the cavity, without getting trapped into periodic orbits. Such a cavity is termed a chaotic (or ray-chaotic) cavity.

To convert the 20 mm × 15 mm cavity used so far into a chaotic cavity, a quadrant of a circle of radius 7 mm, centered at (0,0), was removed to form a "quarter Sinai billiard." (A Sinai billiard is a rectangular cavity with a circle removed from the center.) The modal spacing distribution for all chaotic cavities follows Wigner's distribution, $(\pi/2)s\exp(-(\pi s^2/4))$, shown as the solid line in Fig. 9(b), and the distribution of the spacings of the first 4000 modes of the quarter Sinai billiard clearly follows this trend. The principal point to note is that in a Wigner distribution, there are no modal spacings of zero, i.e., no two modes have the same modal frequency. This is exactly what was required to improve the conditioning of M_t and to lead to better images.

The modal frequencies of the chaotic cavity cannot be calculated using simple algebraic expressions, as they can for the rectangular case, and so a FE model was used. A FE simulation was constructed in MATLAB and the function `sptarn` was used to calculate the eigenvalues (Arnoldi estimates) and eigenvectors of the cavity, and thereby construct matrix M_t. A different FE mesh and the MATLAB function `hyperbolic` were used to calculate the measured photoacoustic time series for a detector at (0.03 mm, 19.96 mm), and 1% Gaussian noise was added. The reconstruction was performed, as above, using a truncated SVD. The singular value spectrum is shown in Fig. 10 (dots) alongside the SVD spectrum from the rectangular case (+), and it shows that the modal degeneracy has been removed so almost all the modes can now be used in the reconstruction without causing ill-conditioning.

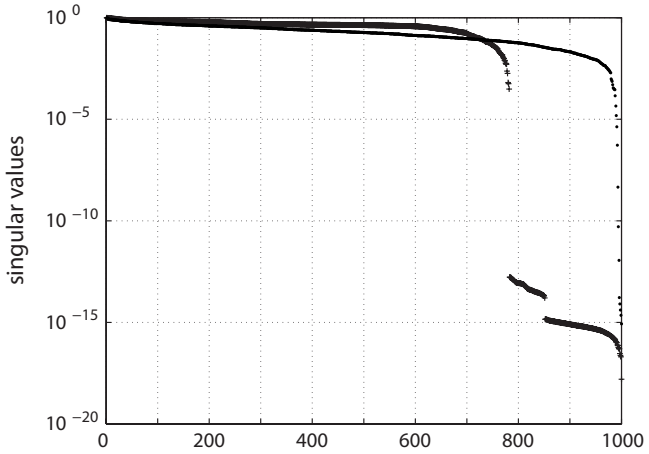


FIG. 10. Singular values of matrix M_t for rectangular (+) and chaotic (.) cavities.

Figure 7(A) shows an initial pressure distribution used to compare the three reconstructions: simple backprojection [Fig. 7(B)], modal inversion in a rectangular cavity [Fig. 7(C)], and modal inversion with the chaotic cavity [Fig. 7(D)].

Figure 7(D) shows the image recovered when 920 modes were included, which when compared to Fig. 7(C) shows better recovery of both the amplitudes and the shapes of the circles. The improvement in the reconstruction of the amplitudes is confirmed by Fig. 7(E), which shows profiles at $y=10$ mm through the images in Figs. 7(A)–7(D), corresponding to the exact p_0 (solid line), p_0 reconstructed using backprojection (dashed), modal reconstruction with rectangular cavity (dot-dashed), and reconstruction with chaotic cavity (dotted). Figure 7(F) shows the mean squared error as a function of the truncation order for the rectangular (cross) and chaotic (open circle) cavities. A much larger percentage of the modes can be used in the chaotic case, and the error is consequently significantly lower than in the rectangular case.

A second example comparing reverberant PAT imaging in the rectangular and chaotic cavities was also calculated. The inversions were performed exactly as before but for the initial pressure distribution shown in Fig. 11(A). The images for the rectangular and chaotic cavities are given in Figs. 11(B) and 11(C), and profiles through these images at $y=15$ mm are shown in Fig. 11(D). Again it is clear that the chaotic cavity gives better results in terms of both amplitude and shape.

V. DISCUSSION

A. Matrix inversion

The aim of this paper is to suggest and try to demonstrate that it might be possible to use a single point measurement of reverberation to do photoacoustic imaging. To this end, simple two-dimensional examples have been given, in

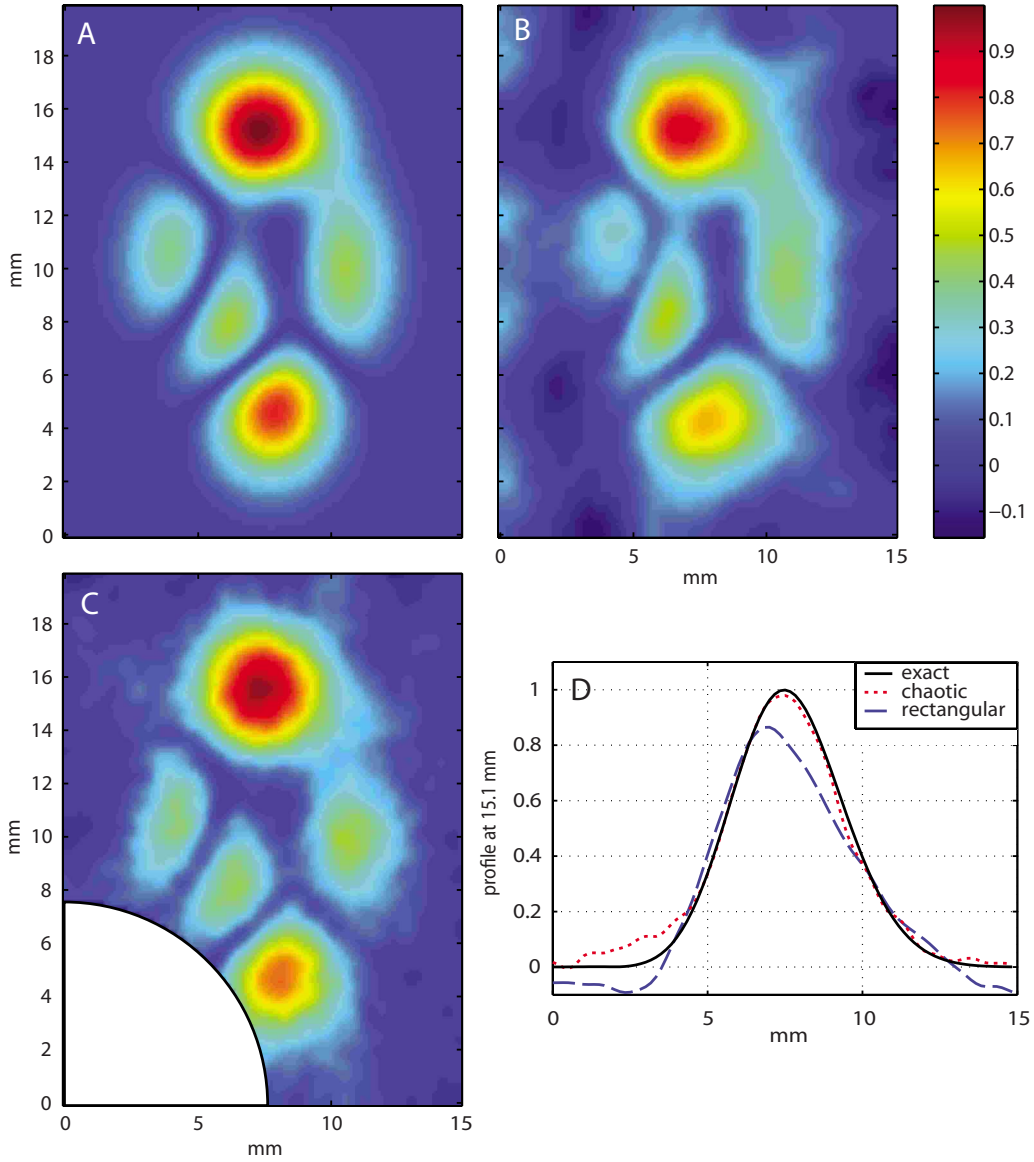


FIG. 11. (Color online) (A) Example initial pressure distribution, p_0. (B) Image of p_0 obtained using a rectangular cavity. (C) Image of p_0 obtained using a chaotic cavity. (D) Profiles at $y=15$ mm through (A), (B), and (C), corresponding to the exact p_0 (solid line), p_0 reconstructed using modal reconstruction with rectangular cavity (dashed), and reconstruction with chaotic cavity (dotted).

which a SVD was used to calculate the matrix inversions. The SVD provides a fundamental decomposition of the matrices, and as such helps to provide insight into the reasons why a chaotic cavity is preferable to an integrable one. (It is worth noting that the inversion of M_t only needs to be performed once for each geometry, i.e., cavity shape and detector position, so for a particular experimental arrangement M_t^{-1} could be precalculated, and each image formed with a single matrix-vector multiplication. It does not need to be calculated for every new initial pressure distribution.) The SVD is only one way to calculate the inversions, and the truncated SVD the only one way to regularize the solution. The literature on algorithms to estimate matrix inversions and regularization is substantial, and there is insufficient space here for a study of which algorithms are best suited to this particular problem. In some cases, perhaps for large-scale problems, a different method altogether may be advantageous. For instance, an optimization approach, in which a norm of the error between the measured data and a model is minimized, may have computational benefits as well as the advantage that a constraint could be added to the cost function to include the prior information that p_0 must be non-negative everywhere.

B. Model assumptions

In the analysis above, a number of idealizations were made to simplify the situation. For instance, it was assumed that the sound speed is the same everywhere within the cavity, that there is no bulk acoustic absorption, and that the walls are perfectly reflecting (have a reflection coefficient of 1). In practice these assumptions will not quite be true, and the sound field will differ from the model, especially at higher frequencies where the absorption may be significant. The time series used in the examples was 0.5 ms duration, over which time the sound waves will have traveled about

0.75 m, or almost 40 times the length of the cavity. The high frequency part of the signal will suffer in two ways over this distance: it will be susceptible to being scattered by localized sound speed heterogeneities and it may be strongly absorbed through the bulk absorption of the medium. Both of these effects will act to reduce the spatial resolution achievable in the image. To counter these effects, both the spatial variation in sound speed and the absorption could be included in the model, if they are known beforehand, and the mode shapes, frequencies, and amplitudes could then be calculated taking these variations into account. When the high frequency absorption is significant, however, the acoustic pressure signal may fall well below the noise floor, and regularization will be required in the inversion to ensure a stable solution. The frequency dependence of absorption varies, so the frequency at which this will start to be a problem will depend on the type of tissue.

The assumption of perfectly reflecting walls will be difficult to achieve in practice, but materials with a large acoustic impedance can achieve high pressure reflection coefficients for sound traveling from water, e.g., for steel $R \approx 0.94$ and for glass $R \approx 0.8$. As above, if the reflection coefficients are known, they can be included in the model. (This will add to the damping of the modes.) It may be possible to do calibration experiments in the cavity containing a known medium, such as distilled water, to ensure that the model of the geometry and boundary conditions are well modeled. Another possibility is to replace one of the walls with an air (pressure-release) boundary and take into account the $R \approx -1$ reflection coefficient when calculating the modal information. This sort of open-topped cavity would have the practical advantage of making the cavity more accessible for inserting and removing samples.

C. Detectors

The question studied in this paper has been whether an image could be obtained from reverberant measurements using *one* detector. Intuitively there would be advantages in using data from more than one measurement point. First, if a single detector happens to be positioned close to the node of a mode, then the contribution of that mode to the measured pressure will be below the noise floor, and it will not be included in the image reconstruction. With two detectors, it is less likely that a mode will be missed in this way, so in some circumstances could improve the image. Second, if the reverberant signal is attenuated through absorption, a particular ray may be too weak to detect by the time it reaches the detector, but a second detector, some distance from the first, may encounter the ray earlier in its "lifetime" and thus with a higher signal-to-noise ratio. However, the conditioning of the matrix M_t, which has been the chief concern of much of this paper, would not be improved, as its conditioning depends on the columns having different modal frequencies, and two different detector points would still see the same modal frequencies. Measurements recorded in different parts of the cavity will also be affected by absorption differently, so it may be that when the absorption is significant, multiple measurements can be used to improve the image.

The detectors in this analysis have been assumed to be omnidirectional, broadband, and located at a single point. "Real" ultrasound detectors are often directional, bandlimited, and of finite size. There is a class of ultrasound detectors, however, which may be suitable for this application. Optical interferometric ultrasound detectors[6,24] can have a very wide, flat, sensitivity range (from dc to >50 MHz), with small element sizes (<50 μm) and could be incorporated into the wall of the cavity, so not affecting the reverberant field.

D. Time-reversal imaging

When using time-reversal to refocus a signal through a heterogeneous or reflecting medium, the measurements are made at detector points and then time-reversed versions are transmitted back into the medium from the same positions. When time-reversal is used for imaging, the retransmission into the medium is replaced by a simulated retransmission in a numerical model. Experiments on time-reversing array measurements of photoacoustic signals back to their origin have been performed[25] and time-reversal has been suggested as a way to reconstruct a photoacoustic image[26,27] when measurements have been made over an array of detectors.

The time-reversal of a single photoacoustic measurement in a reverberant cavity in order to form an image has yet to be investigated, although the idea of using single channel time-reversal in a reverberant cavity to detect the location of a point source has been studied and applied to communication[28] and touch sensitive interactive surfaces.[29] In three interesting papers, Draeger and co-workers[30–32] described and analyzed experimental and numerical studies of the time-reversal of a single point measurement of the reverberation generated by a point source back to its origin. These studies were conducted using flexural waves on a chaotically shaped silicon plate, but are analogous to the acoustic case, and demonstrated that the signal could indeed be refocused. A finite difference model was used to show the waves converging on the point source,[30] perhaps the first example of time-reversal imaging using single channel reverberation data. As an imaging technique, single point time-reversal differs in a number of ways to the modal inversion given here, e.g., it requires the chaoticity of the cavity and so does not work in a rectangular cavity,[21,32] and it is inherently approximate, so can never give an exact image.

It is instructive to understand where the approximation in single channel time-reversal arises, and why it does not apply to the modal inversion described here. This can be explained both in terms of modes or rays. The ray explanation is that while each pulse arrives at the detector from one specific direction, it is retransmitted in all directions, as the omnidirectional detector does not record the direction of arrival.[32] This is similar to the approximation made in the backprojection in Sec. III, which introduced significant artifacts while still roughly locating the regions of large initial pressure. In terms of modes, the measured reverberation signal consists of a contribution from each mode weighted by the modal amplitudes at the detector point, $A_m \phi_m(x_d, y_d)$. The time-reversed signal that is retransmitted into the medium (or

the model in the case of imaging) will only excite the modes at the detector (now transmitter) point with the same amplitude. Therefore, unless the detector is at a maximum of a mode, its contribution to the overall acoustic field will be underestimated by $(A_m\phi_m(x_d,y_d))^2$. Draeger and Fink[31] suggested that the solution to this problem would be to know the amplitude of each eigenmode at the detector point and compensate for it, which is precisely the role of $\phi_m(x_d,y_d)$ in Eq. (7). This amplitude effect only now remains a problem for modes that fall below the noise floor at the detection point. To avoid this, the detector was placed close to the corner of the cavity where the boundary conditions enforce a maximum for all modes.

VI. SUMMARY

Photoacoustic tomography using reverberation recorded by a single detector was proposed, and two-dimensional simulations using a simple imaging algorithm based on a modal inversion were described. The technique was analyzed using a SVD, and it was shown that better quality images can be obtained when the reverberant cavity is chaotic. The practicalities of such an imaging technique, and its advantages over single channel time-reversal imaging, were discussed.

ACKNOWLEDGMENTS

The authors would like to thank Professor Simon Arridge for useful and interesting discussions. This work was supported by the Engineering and Physical Sciences Research Council (UK).

[1]Photoacoustic tomography (PAT) is also known as photoacoustic imaging or optoacoustic tomography. When the optical excitation pulse is replaced by microwave or rf radiation, the technique is called thermoacoustic tomography. The image reconstruction procedure is the same in both cases.

[2]R. A. Kruger, P. Liu, Y. R. Fang, and C. R. Appledorn, "Photoacoustic ultrasound (PAUS)—Reconstruction tomography," Med. Phys. **22**, 1605–1609 (1995).
[3]C. G. A. Hoelen, F. F. M. de Mul, R. Pongers, and A. Dekker, "Three-dimensional photoacoustic imaging of blood vessels in tissue," Opt. Lett. **23**, 648–650 (1998).
[4]X. Wang, Y. Pang, G. Ku, X. Xie, G. Stoica, and L. V. Wang, "Noninvasive laser-induced photoacoustic tomography for structural and functional in vivo imaging of the brain," Nat. Biotechnol. **21**, 803–806 (2003).
[5]M. Xu and L. V. Wang, "Photoacoustic imaging in biomedicine," Rev. Sci. Instrum. **77**, 041101 (2006).
[6]E. Z. Zhang, J. G. Laufer, and P. C. Beard, "Backward-mode multiwavelength photoacoustic scanner using a planar Fabry-Perot polymer film ultrasound sensor for high-resolution three-dimensional imaging of biological tissues," Appl. Opt. **47**, 561–577 (2008).
[7]R. Kruger, W. Kiser, D. Reinecke, G. Kruger, and K. Miller, "Thermoacoustic molecular imaging of small animals," Mol. Imaging **2**, 113–123 (2003).
[8]J. G. Laufer, C. Elwell, D. Delpy, and P. Beard, "In vitro measurements of absolute blood oxygen saturation using pulsed near-infrared photoacoustic spectroscopy: Accuracy and resolution," Phys. Med. Biol. **50**, 4409–4428 (2005).
[9]J. G. Laufer, C. Elwell, D. Delpy, and P. Beard, "Quantitative spatially resolved measurement of tissue chromophore concentrations using photoacoustic spectroscopy: Application to the measurement of blood oxygenation and haemoglobin concentration," Phys. Med. Biol. **52**, 141–168 (2007).
[10]B. T. Cox, S. R. Arridge, and P. C. Beard, "Simultaneous estimation of chromophore concentration and scattering distributions from multiwavelength photoacoustic images," Proc. SPIE **6856**, 68560Y (2008).
[11]B. T. Cox, S. R. Arridge, and P. C. Beard, "Photoacoustic tomography with a limited-aperture planar sensor and a reverberant cavity," Inverse Probl. **23**, S95–S112 (2007).
[12]P. Kuchment and L. Kunyansky, "Mathematics of thermoacoustic tomography," Eur. J. Appl. Math. **19**, 191–224 (2008).
[13]M. A. Anastasio, J. Zhang, X. Pan, Y. Zou, G. Ku, and L. V. Wang, "Half-time image reconstruction in thermoacoustic tomography," IEEE Trans. Med. Imaging **24**, 199–210 (2005).
[14]A. K. Louis and E. T. Quinto, *Surveys on Solution Methods for Inverse Problems* (Springer-Verlag, Vienna, 2000), pp. 147–154.
[15]Y. Xu, L. V. Wang, G. Ambartsoumian, and P. Kuchment, "Reconstructions in limited-view thermoacoustic tomography," Med. Phys. **31**, 724–733 (2004).
[16]M. A. Anastasio and J. Zhang, "Image reconstruction in photoacoustic tomography with truncated cylindrical measurement apertures," Proc. SPIE **6086**, 608610 (2006).
[17]S. K. Patch, "Thermoacoustic tomography—consistency conditions and the partial scan problem," Phys. Med. Biol. **49**, 2305–2315 (2004).
[18]B. T. Cox and P. C. Beard, "Exact photoacoustic image reconstruction using a planar sensor array and image sources," Proc. SPIE **6437**, 64371H (2007).
[19]B. T. Cox, S. R. Arridge, and P. C. Beard, "Photoacoustic tomography using reverberant field data from a single detector," Proc. SPIE **6856**, 68561X (2008).
[20]B. T. Cox, S. Kara, S. R. Arridge, and P. C. Beard, "k-space propagation models for acoustically heterogeneous media: Application to biomedical photoacoustics," J. Acoust. Soc. Am. **121**, 3453–3464 (2007).
[21]G. Tanner and N. Søndergaard, "Wave chaos in acoustics and elasticity," J. Phys. A: Math. Theor. **40**, R443–R509 (2007).
[22]M. C. Gutzwiller, *Chaos in Classical and Quantum Mechanics*, Interdisciplinary Applied Mathematics (Springer-Verlag, New York, 1990).
[23]H.-J. Stöckmann, *Quantum Chaos: An Introduction* (Cambridge University Press, Cambridge, 1999).
[24]Y. Hou, J. S. Kim, H. Sheng-Wen, S. Ashkenazi, L. J. Guo, and M. O'Donnell, "Characterization of a broadband all-optical ultrasound transducer—From optical and acoustical properties to imaging," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **55**, 1867–1877 (2008).
[25]E. Bossy, K. Daoudi, A.-C. Boccara, M. Tanter, J.-F. Aubry, G. Montaldo, and M. Fink, "Time reversal of photoacoustic waves," Appl. Phys. Lett. **89**, 184108 (2006).
[26]Y. Xu and L. V. Wang, "Time reversal and its application to tomography with diffracting sources," Phys. Rev. Lett. **92**, 033902 (2004).
[27]P. Burgholzer, G. J. Matt, M. Haltmeier, and G. Paltauf, "Exact and approximate imaging methods for photoacoustic tomography using an arbitrary detection surface," Phys. Rev. E **75**, 046706 (2007).
[28]M. G. Heinemann, A. Larrazza, and K. B. Smith, "Acoustic communications in an enclosure using single-channel time-reversed acoustics," Appl. Phys. Lett. **80**, 694–696 (2002).
[29]R. K. Ing, N. Quieffin, S. Catheline, and M. Fink, "In solid localization of finger impacts using acoustic time-reversal process," Appl. Phys. Lett. **87**, 204104 (2005).
[30]C. Draeger and M. Fink, "One-channel time reversal of elastic waves in a chaotic 2-D cavity," Phys. Rev. Lett. **79**, 407–410 (1997).
[31]C. Draeger and M. Fink, "One-channel time reversal in chaotic cavities: Theoretical limits," J. Acoust. Soc. Am. **105**, 611–617 (1999).
[32]C. Draeger, J.-C. Aime, and M. Fink, "One-channel time reversal in chaotic cavities: Experimental results," J. Acoust. Soc. Am. **105**, 618–625 (1999).

Laser ultrasonic inspection of the microstructural state of thin metal foils

O. Balogun
Lawrence Livermore National Laboratory, University of California, Livermore, California 94550 and Department of Materials Science and Engineering, The Johns Hopkins University, Baltimore, Maryland 21218

R. Huber and D. Chinn
Lawrence Livermore National Laboratory, University of California, Livermore, California 94550

J. B. Spicer
Department of Materials Science and Engineering, The Johns Hopkins University, Baltimore, Maryland 21218

(Received 14 June 2008; revised 11 December 2008; accepted 12 December 2008)

A laser-based ultrasonic technique suitable for characterization of the microstructural state of metal foils is presented. The technique relies on the measurement of the intrinsic attenuation of laser-generated longitudinal waves at frequencies reaching 1 GHz resulting from ultrasonic interaction with the sample microstructure. In order to facilitate accurate measurement of the attenuation, a theoretical model-based signal analysis approach is used. The signal analysis approach isolates aspects of the measured attenuation that depend strictly on the microstructure from geometrical effects. Experimental results obtained in commercially cold worked tungsten foils show excellent agreement with theoretical predictions. Furthermore, the experimental results show that the longitudinal wave attenuation at gigahertz frequencies is strongly influenced by the dislocation content of the foils and may find potential application in the characterization of the microstructure of micron thick metal foils. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3068447]

PACS number(s): 43.35.Zc [YHB] Pages: 1437–1443

I. INTRODUCTION

The microstructural state of a material is critical to predicting the mechanical behavior and assessing the damage state of the material. Ultrasound-based inspection techniques have emerged as powerful tools for the characterization of the microstructure of metals and metal alloys.[1–14] The ultrasonic wave velocity depends on several bulk mechanical properties including the elastic modulus and Poisson's ratio, which depend on the compliance and crystalline orientation of the constituent grains in the microstructure.[1,2] The ultrasonic wave attenuation, which relates to the decrease in amplitude of ultrasound with propagation distance, depends on several characteristics of the microstructure including grain orientation, grain size, and the population of defects including dislocations, voids, vacancies, and secondary phases.[3–9] It is also well known that certain physical aspects of a material, such as anharmonicity of the crystal lattice bonding and the dislocation content, can lead to nonlinear elastic behavior that produces harmonic generation of ultrasound.[10–14] Consequently, by measuring the nonlinear-ultrasonic response, these aspects of the microstructure can be characterized.

Several experimental ultrasound-based techniques have been reported for the characterization of the microstructure of metals and metal alloys. The traditional ultrasonic techniques use piezoelectric transducers for the generation and detection of ultrasound. A potential limitation of these techniques arises from the coupling medium that facilitates the transmission of ultrasound into the sample. The coupling medium leads to loss of ultrasonic energy, particularly at high frequencies. Laser-based ultrasonic (LBU) techniques provide a suitable alternative to the traditional ultrasonic techniques. These techniques use pulsed or intensity-modulated lasers for the generation of ultrasound. The generation laser is delivered to a specimen surface, where the laser energy is absorbed leading to localized specimen heating and thermal stress production. The thermal stresses relax by launching ultrasound in the specimen by transient thermoelastic expansion. The generated ultrasound is detected at the specimen surface by monitoring the surface displacement or strain using an optical interferometer or a diffraction-based probe.[15,16] LBU techniques are noncontacting, remote, and can provide optical diffraction-limited spatial resolution. Pulsed lasers allow for the generation of broadband ultrasound with frequency components ranging from a few kilohertz up to hundreds of gigahertz, and intensity-modulated lasers can be used to generate narrowband ultrasound.

The majority of the reported LBU work on microstructural characterization use broadband linear-ultrasonic approaches where the velocity and attenuation of bulk or surface acoustic waves are measured and analyzed to obtain the pertinent microstructural information.[17–21] In the general case, the ultrasonic attenuation is more strongly affected by the microstructure than is the velocity, and the sensitivity increases with frequency. While LBU measurements of attenuation at megahertz frequencies in metals and metal alloys are widespread, there are limited experimental reports

on attenuation measurements in the gigahertz range. The most notable reports include the works of Moreau and Lord[22] and Blanchard *et al.*[23] in which the attenuation coefficients of laser-generated longitudinal waves were measured in polycrystalline aluminum plates at frequencies reaching 500 MHz. Relating the measured attenuation to the microstructural state of a material can be challenging because the attenuation is influenced by multiple physical processes that may or may not depend directly on the microstructure. Furthermore, the precision needed to isolate the contribution of the microstructure to the attenuation is also limited at gigahertz frequencies due to low measurement signal-to-noise ratio resulting from the small ultrasonic amplitude obtained using a thermoelastic source along with the broad detection bandwidth needed for certain types of frequency-resolved material property measurements.

In this paper, a LBU technique is used for the nondestructive characterization of the microstructure of commercially cold worked tungsten foils with micrometer scale thickness. The tungsten specimen is chosen specifically to monitor the interaction of the dislocations in the microstructure with pulsed-laser-generated longitudinal waves. The attenuation coefficient of the longitudinal waves resulting from dislocation induced scattering and absorption is measured at frequencies reaching 1 GHz. Owing to the cold working process used for fabrication, a large network of dislocations accumulates in the specimen, producing a dominant effect on the attenuation of the laser-generated ultrasound compared to other competing sources of attenuation that may result from the specimen microstructure. In order to facilitate the measurement, a model-based signal analysis approach is used, that allows one to isolate aspects of the attenuation that are directly related to the microstructure. The experimental results show excellent agreement with theoretical predictions, and the measured longitudinal wave attenuation is found to be highly sensitive to changes in the dislocation content of the specimen that occurred during annealing. The LBU technique may be suitable for assessing the extent of damage in the microstructure of thin metal foils, which is a critical aspect in evaluating their structural integrity.

II. THEORY

The theory of laser-generated ultrasound has been discussed extensively in the literature for various laser source configurations and sample geometries.[24–27] In this work, the appropriate model pertaining to the experimental configuration used is applied as an inverse signal analysis tool for the measurement of laser ultrasonic attenuation in cold worked tungsten foils. The results obtained are used to characterize the microstructural state of the foils. As a guide to the signal analysis approach, a brief description of various modes of ultrasonic attenuation in metals is presented, and the pertinent features of the thermoelastic model used to characterize the attenuation are detailed.

Various workers have investigated the attenuation of ultrasonic waves in metals. In these investigations, the test materials are assumed to be polycrystalline, and ultrasound is transmitted through the materials over relatively long distances compared to the average grain size. In order to analyze the measured attenuation, mean field theories are used to estimate the required microstructural information. In the general case, ultrasonic attenuation in polycrystalline materials at room temperature results from a combination of effects, namely, grain scattering, absorption, and ultrasonic diffraction. The contribution of grain scattering is typically large in most metals and is due to microscopic anisotropic effects resulting from the variation of the elastic constants and orientation from grain to grain. These grain effects lead to ultrasonic refraction at the grain boundaries and a reduction in the ultrasonic amplitude with propagation distance. Depending on the ratio of the wavelength to the average grain size, the ultrasonic attenuation due to grain scattering has characteristic frequency dependences.[1,4,28] Ultrasonic wave absorption in polycrystalline materials is associated with thermoacoustic damping, dislocation damping, and other anelastic relaxation mechanisms.[29–32] Thermoacoustic damping results from the irreversible conversion of ultrasonic energy to heat as a material undergoes compression or expansion following the propagation of an ultrasonic wave.[31] This attenuation mechanism is limited to longitudinal waves and is typically very small compared to grain scattering effects. Dislocation damping is considered by various workers as being hysteretic in nature and occurs as a result of dislocation motion being out of phase with the incident ultrasonic wave leading to energy loss.[32] Dislocation damping depends on the dislocation loop length and is appreciable as the loop length becomes comparable to the ultrasonic wavelength. Thermoacoustic and dislocation damping produce viscous attenuation effects, leading to a decrease in ultrasonic energy with the square of the frequency.

Attenuation resulting from diffraction effects in most laser ultrasonic measurements occurs primarily because of the finite size of the ultrasonic source and probe. The effect of ultrasonic diffraction is understood through examination of the displacement fields produced by a point excitation force in a solid, as calculated by Aki and Richards.[33] The expression for the displacement u_i along an arbitrary axis i produced by a point excitation source $X_0(t)$ acting along a direction x_j is represented by

$$\begin{aligned} u_i &= X_0 * G_{ij} \\ &= \frac{1}{4\pi\rho}(3\upsilon_i\upsilon_j - \delta_{ij})\frac{1}{r^3}\int_{r/c_l}^{r/c_s} \tau X_0(t-\tau)d\tau \\ &\quad + \frac{1}{4\pi\rho {c_l}^2}\upsilon_i\upsilon_j\frac{1}{r}X_0\left(t-\frac{r}{c_l}\right) + \frac{1}{4\pi\rho {c_s}^2}(\upsilon_i\upsilon_j - \delta_{ij}) \\ &\quad \times\frac{1}{r}X_0\left(t-\frac{r}{c_s}\right), \end{aligned} \tag{1}$$

where G_{ij} is the Green's function, c_l and c_s are the longitudinal and shear wave velocities in the solid, ν is the direction cosine for the vector x, δ_{ij} is the Kronecker delta function, ρ is the material density, and r is the propagation distance. The first term in the equation shows that the displacement field decays with propagation distance as r^{-2} for the case where $X_0(t)$ is nonzero for times that are short compared to r/c_s

$-r/c_l$ and represents a coupled field comprised of longitudinal and shear waves. The second and third terms represent the displacement fields of the pure longitudinal and shear waves, which decay as r^{-1}. For small r, the first term in Eq. (1) dominates the latter terms, yielding the amplitude decay in the acoustic near field. As r tends to infinity, the second and third terms become dominant, and the amplitudes of the longitudinal and shear waves decay as r^{-1} with distance as a spherical wave. These terms are the far field terms. Note that as r tends to zero, the amplitude decay becomes negligibly small; as such, the ultrasonic field can be approximated by a plane wave. In a practical case where the excitation source and receiver spot sizes are finite, there are two limiting cases where the ultrasonic diffraction loss is well characterized. Assuming that the source and receiver spot sizes are equal, a spherical wave approximation holds for a small spot size, long propagation distance, or low ultrasound frequency, while a plane wave approximation can be used for a large spot size, short propagation distance, or high ultrasonic frequency. These limiting cases are distinguishable by the Fresnel diffraction parameter $S=\lambda r/a^2$, where a is the spot size of the source/receiver and λ is acoustic wavelength; $S<0.2$ for the plane wave approximation, and $S>5$ for the spherical wave approximation. For a general case where the source and receiver spot sizes differ, the ultrasonic diffraction loss depends on S and the ratio of source and receiver spot sizes.[18]

In this work, a theoretical model-based signal analysis approach is used to estimate the attenuation of laser-generated longitudinal waves in laser ultrasonic measurements. This approach allows for attenuation due to the microstructure to be isolated from other contributing attenuation effects. Furthermore, by using model-based approaches, limitations to direct analysis of the experimental results due to signal-to-noise ratio considerations can be overcome. The theoretical model used is based on the solution to the temperature-rate dependent thermoelastic wave equation derived by Spicer and Hurley[34] for elastically isotropic plates. Consider the experimental geometry shown in Fig. 1 in which the excitation and detection lasers are incident on the epicenter on the same side of the sample. Figure 2 shows a calculated waveform obtained in a 50 μm thick tungsten plate. An excitation laser source radius of 25 μm [full-width-at-half-maximum (FWHM) intensity level] and a pulse duration of 900 ps are used in the calculation. The reported bulk values of the optical and thermal properties of tungsten are used,[30] in addition to the bulk longitudinal and shear wave velocities of $c_l=5.22$ mm/μs and $c_t=2.89$ mm/μs, respectively.[30] The waveform shows a rapid thermal expansion of the sample surface close to time zero followed by a slow amplitude increase with time. Several transient echoes, including the bulk longitudinal (P) and shear (S) waves, mode converted waves, and a low frequency oscillation, are also observed. The thermal expansion signal is eliminated by applying a 200 MHz first order high pass filter to Fig. 2(a). The filtered waveform [labeled curve A in Fig. 2(b)] clearly shows the transient echoes. The amplitudes of the longitudinal wave echoes decrease slightly from one arrival to the next due to ultrasonic diffraction.

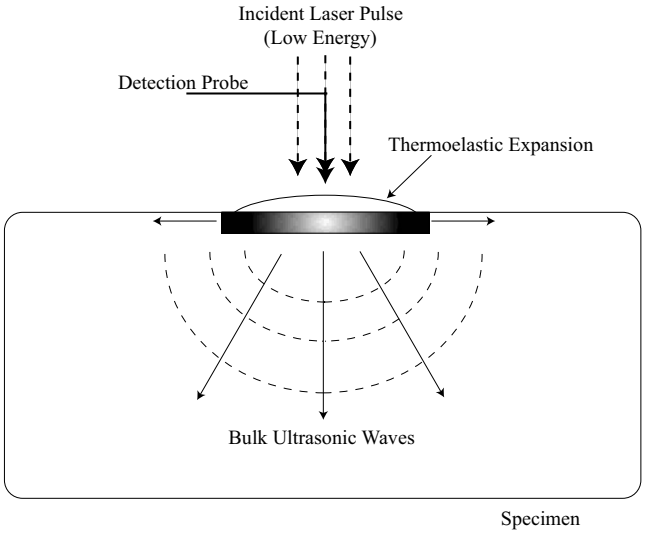


FIG. 1. A schematic of the geometry of a freestanding plate excited by a laser source.

Consider the case in which the microstructure of the tungsten foil produces frequency (or wavelength) dependent attenuation of the longitudinal waves. In order to simulate this effect, a damping mechanism is introduced into the model by assuming that the longitudinal wave speed is complex as expressed below,

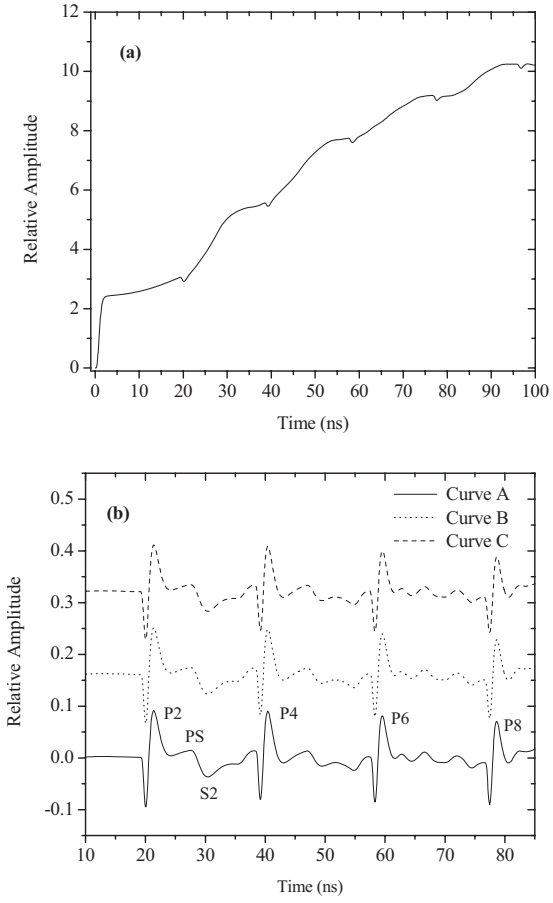


FIG. 2. Theoretical waveforms showing (a) the thermal expansion and transient acoustic waves arriving on epicenter in a tungsten foil and (b) results of retaining high frequency content by numerical filtering.

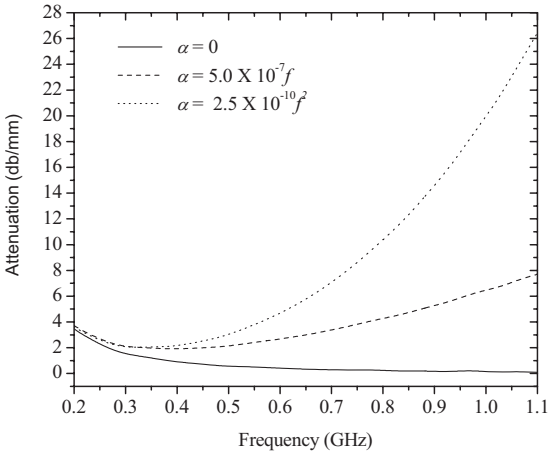


FIG. 3. Calculated longitudinal wave attenuation curves for various power law profiles.

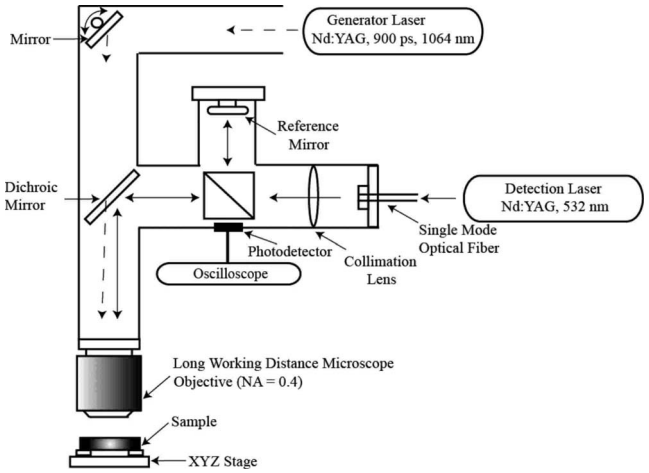


FIG. 4. A schematic of the experimental apparatus used to perform ultrasonic measurements in thin metal foils.

$$c_1 = c_l + c_{ld}, \quad c_{ld} = \alpha f^n. \tag{2}$$

Here c_1 is the complex longitudinal wave velocity expressed in terms of the true wave speed c_l and an imaginary component c_{ld} that has an assumed power law frequency dependence. The coefficient of the power law α is taken to be zero for the case where the microstructure-based attenuation is absent. While this assumption is not justified in detail, ultrasonic wave speeds are complex valued and frequency dependent in a viscoelastic medium, owing to the complex elastic properties, which lead to ultrasonic energy loss.[35] Here, it is assumed that the elastic moduli can also be expressed as complex quantities even though the underlying physical mechanisms are not completely specified. This assumption immediately implies that the ultrasonic wave speed can be represented as a complex quantity.

Figure 2(b) shows the filtered waveforms obtained for the case where α has a linear (curve B) and a quadratic (curve C) dependence on frequency. In both cases, the arrival times of the longitudinal wave echoes are the same as in curve A, indicating that the complex velocity does not change the actual propagation speed. Note that small arbitrary values are chosen for the coefficient of the power laws for curves A and B so that the apparent amplitude decay of the echoes in these curves appears to be similar. In order to calculate the attenuation coefficient of the longitudinal wave echoes in units of dB/mm for each of these cases, the amplitude spectra of the first two echoes (P2 and P4) in curves A and B are compared. These echoes are extracted from the time trace using a rectangular window, and the windowed data are Fourier transformed to obtain the magnitude spectra. The attenuation coefficient is evaluated as a function of frequency using the relationship $(20/\Delta l)\log(A_1/A_2)$, where A_1 and A_2 are the amplitudes of the P2 and P4 echoes in the frequency domain, $\Delta l = 2l$ is the difference in propagation distance between the two echoes, and l is the plate thickness. The effect of the high pass filter applied is removed in taking the ratio of the amplitude spectra A1 and A2 since the filtering operation is linear. Figure 3 shows the attenuation curves for the three cases between 0.2 and 1.1 GHz. For the first case where $\alpha=0$, the calculated attenuation coefficient is small and decreases progressively with frequency as a result of ultrasonic diffraction. The plane wave or spherical wave approximations for the ultrasonic diffraction are not applicable in this case because the ultrasonic fields for the P2 and P4 echoes are in the intermediate region between these extreme cases. Using the excitation laser spot size, the calculated Fresnel diffraction parameter goes from 1.0 to 0.2 for the P2 echo and from 2.0 to 0.4 for the P4 echo between 0.2 and 1.1 GHz. The theoretical model allows one to calculate the frequency dependence of the ultrasonic diffraction loss irrespective of the regime of ultrasonic wave propagation provided that the sample properties and spot sizes of the excitation source and receiver are well known. Note that ultrasonic mode conversions between longitudinal and shear waves produce an energy loss that contributes to the calculated attenuation. For the second case where $\alpha=5.0\times10^{-7}f$, the attenuation curve is approximately linear beyond about 0.5 GHz, and for $\alpha=2.5\times10^{-10}f^2$, the attenuation curve is quadratic. The variations in the attenuation coefficients with frequency for the three cases are clearly different even though the time domain waveforms are similar. By subtracting the attenuation due to ultrasonic diffraction and mode conversions, one can isolate the contribution of the microstructural attenuation in a practical case.

It is important to note that the approach used to introduce microstructure-based attenuation into the model is not rigorous and is only used here owing to the involved nature of physically accurate modeling approaches. Also of importance is the fact that there are secondary effects, such as surface roughness, that can lead to ultrasonic attenuation.[36] Surface roughness effects are assumed to be negligible in this work because, for the samples tested, the scale of the surface roughness is significantly smaller than the ultrasonic wavelength.

III. EXPERIMENTAL SETUP

A schematic of the experimental setup used is shown in Fig. 4. A neodymium doped yttrium aluminum garnet (Nd:YAG) microchip laser is used for ultrasonic wave generation. The laser has a pulse energy of 5 μJ, a pulse dura-

tion of 900 ps, a repetition rate of 50 kHz, and a wavelength of 1064 nm. The generation laser is focused by a 20× long-working-distance microscope objective on the sample surface. The sample is supported on a three-axis translation stage. The circular spot diameter of the excitation laser is expanded to 145 μm at the FWHM intensity level by defocusing the input light into the microscope. Expansion of the generation laser beam diameter is needed in order to generate approximately planar ultrasonic waves. In addition, the total laser energy deposited in the sample is maximized without sample ablation.

The normal surface displacement associated with the interaction of the generated ultrasound with the sample boundary is detected on the excited side of the sample using a path-stabilized Michelson interferometer. The interferometer uses a frequency doubled Nd:YAG laser with a wavelength of 532 nm. The detection laser is delivered into the experimental setup through a single mode optical fiber, collimated, and directed to a beam splitter that splits the light into signal and reference beams. The signal beam is directed to the sample surface through the microscope objective and is redirected on reflection from the sample surface to a high speed photodetector with a rise time of 1 ns. The circular diameter of the signal beam on the sample surface is 1.2 μm at the FWHM. The reference beam is delivered to a reference mirror that redirects the beam to the photodetector where it interferes with the signal beam producing an intensity-modulated signal that is related to the absolute displacement of the sample surface. The interferometer is stabilized at the point of maximum displacement sensitivity through active adjustment to the optical path length difference between the reference and signal beams. This is achieved by electronically controlling the reference mirror position using a piezoelectric actuator. The output signal from the photodetector is recorded using a digital oscilloscope at the rate of 10 gigasamples per second and is sent to a computer for data processing.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

Figure 5 shows the filtered normal surface displacement signal measured in an *as-received* tungsten foil with a nominal thickness of 50 μm. The tungsten sample (99.95% purity) is manufactured using a cold rolling process. The excitation and detection lasers are aligned on the epicenter, and the sample response is averaged for about 50 000 times to minimize the incoherent noise in the measured data. The measured sample response contained the thermoelastic expansion signal, which is removed by applying a 300 MHz first order high pass filter to the measured waveform. A 1.0 GHz first order low pass filter is also applied to the measured waveform to reduce the broadband noise from the optical detection system. Several longitudinal wave echoes are seen in the filtered waveform; the amplitudes of the successive echoes decrease monotonically due to ultrasonic attenuation. Also shown in Fig. 5 are two filtered theoretical waveforms obtained by applying the same filter to the calculated transient response. In curve A, the reported values of the longitudinal and shear wave velocities of tungsten (c_l $=5.22$ mm/μs, $c_t=2.89$ mm/μs) are used. The measured spot size and pulse duration of the excitation laser and the bulk values of the optical and thermal properties of tungsten are used as input parameters to the model. The plate thickness is estimated to be 55 μm based on the autocorrelated time interval (Δt) between successive echoes given by $\Delta t = 2l/c_l$. The damping parameter α is set to zero, allowing for only ultrasonic energy loss from diffraction and mode conversion to be accounted for in the model. The apparent amplitude decay of the echoes is less pronounced in curve A compared to the experimental data. Curve B is in closer agreement with experimental data. A complex longitudinal wave velocity was assumed in this case, and c_{ld} was set equal to $23.5\times10^{-10}f^2$. Several power laws are used, and the quadratic law provided the best agreement between the experimental data and the model when the amplitude decay and the width of the echoes are compared. Figure 5(b) shows the comparison between the variation of the peak-to-peak amplitudes of the echoes with distance in the experimental and modeled data. The propagation distance of each echo is given by $2nl$, where n is the echo number. The amplitude decay of the echoes in curve B closely matches the experi-

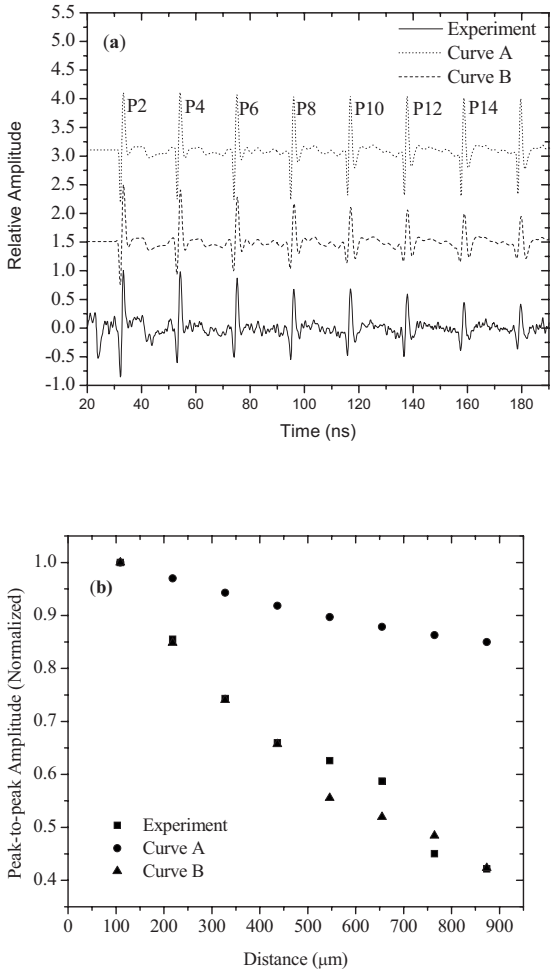


FIG. 5. (a) Time domain waveform measured in a cold worked tungsten foil with a nominal thickness of 50 μm. Curves A and B show the theoretical waveforms calculated with a plate thickness of 55 μm and damping parameters of $c_{ld}=0$ and $c_{ld}=23.5\times10^{-10}f^2$, respectively. (b) Comparison of the peak-to-peak amplitudes of the longitudinal wave echoes in (a).

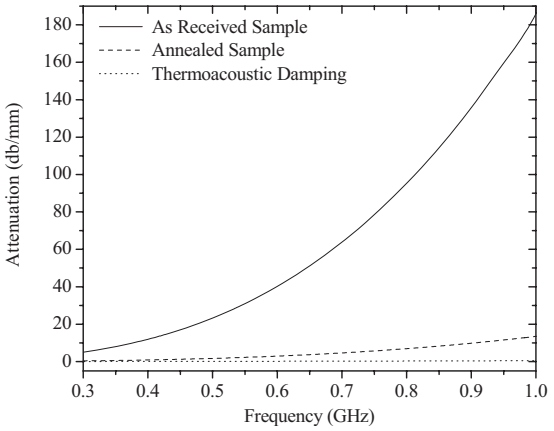


FIG. 6. Longitudinal wave attenuation curves obtained in cold worked tungsten foils.

mental data, suggesting the presence of intrinsic microstructure-based ultrasonic attenuation.

Figure 6 shows the modeled intrinsic attenuation coefficient of the longitudinal wave echoes as a function of frequency. The amplitude spectra of the first two echoes (P2 and P4) in curve B in Fig. 5 were compared, as discussed in Sec. II. Note that the useful frequency range for evaluating the attenuation coefficient is limited to 0.3–1.0 GHz due to the passband of the filters applied to the unfiltered waveform. The contributions of ultrasonic diffraction and mode conversion to the attenuation are calculated by comparing the amplitude spectra of the P2 and P4 echoes in curve A in Fig. 5, and the resulting data are subtracted from the attenuation curve obtained from the spectra of the same echoes in curve B. The calculated ultrasonic diffraction loss is negligibly small (<2 dB/mm) over the frequency range in the figure. Also shown in Fig. 6 is the modeled attenuation coefficient in an annealed tungsten sample of the same thickness. The annealed sample was obtained by heating the as-received sample to approximately 1400 °C and holding the temperature for 3 h in a nitrogen environment to prevent surface oxidation. The attenuation coefficients obtained in the as-received and annealed samples follow quadratic frequency dependent curves. The differences in attenuation between the samples are significant and increase with frequency, suggesting that a change in the microstructure of the sample occurred as a result of the annealing operation. Most likely, the intrinsic attenuation of the longitudinal wave in the tungsten samples is related to absorption and scattering from dislocation networks. This is consistent with the processing history of the material since dislocation densities in cold worked materials are significantly higher than in annealed materials, and the effects of grain scattering are negligible due to the well-known elastic isotropy of tungsten. Note that due to the small length of an individual dislocation in the tungsten sample compared to the wavelength of the longitudinal waves, a negligible scattering signal is expected from the interaction of the longitudinal wave with a single dislocation. However, interaction with a population of dislocations would produce a cumulative loss in ultrasonic energy that is measurable. This idea is also supported by the fact that the longitudinal wave velocity measured in the annealed foil is found to be lower than that in the as-received sample by about 0.9%. The velocity change could be related to the larger intrinsic residual stress resulting from the dislocation networks in the microstructure of the as-received sample. These networks are reduced through annealing, leading to relieving of the residual stress in the rolled foil. Indeed, the longitudinal wave attenuation in the annealed sample approaches the fundamental thermodynamic limit dictated by thermoacoustic damping at frequencies lower than 0.4 GHz. The accepted thermoacoustic damping coefficient for longitudinal waves in tungsten is 4.34×10^{-19} dB/mm GHz^2.[30] The dislocation content in the annealed sample is significantly reduced, and the sensitivity of the attenuation to the residual dislocation content in the sample is only evident at higher frequencies.

V. CONCLUSIONS

In this work, ultrasonic characterization of certain microstructural characteristics of tungsten foils was demonstrated. A LBU technique was used that allowed for measurement of the intrinsic attenuation of high frequency ($\sim$1 GHz) longitudinal waves resulting from ultrasonic interaction with the sample microstructure. Direct mapping of the measured attenuation to the microstructural state is typically very challenging due to the multiple physical effects that lead to ultrasonic attenuation as well as the low signal-to-noise ratio of laser ultrasonic signals. In order to overcome these challenges, a model-based signal analysis approach was used to estimate the intrinsic attenuation from the measured laser ultrasonic data. The model used here incorporates ultrasonic attenuation by assuming a complex wave speed that varies with frequency according to a power law. With this approach, microstructure-based attenuation of the longitudinal wave was estimated by comparison of the experimental measurements and modeled results. Experimental results obtained in cold worked tungsten foils, where several longitudinal wave echoes were observed, show excellent agreement with modeled predictions. Furthermore, the experimental results showed that the longitudinal wave attenuation at gigahertz frequencies is strongly influenced by the dislocation content of the foil and could be used to characterize this aspect of thin metal foil microstructures.

ACKNOWLEDGMENTS

This work was performed under the auspices of the U.S. Department of Energy by the Lawrence Livermore National Laboratory, University of California under Contract No. W-7405-Eng-48 and was based on work supported by, or in part by, the Office of Basic Energy Sciences, U.S. Department of Energy under Grant No. DEFG0203ER46090 and the Air Force Office of Scientific Research under Grant No. FA9550-06-1-0309.

[1]F. E. Stanke and G. S. Kino, "A unified theory of elastic wave propagation in polycrystalline materials," J. Acoust. Soc. Am. **75**, 665–681 (1984).
[2]W. Johnson, F. Mauer, D. Pitchure, S. J. Norton, Y. Grinberg, and F. Bendec, "Temperature and annealing dependence of the longitudinal ultrasonic velocity in aluminum alloys," J. Mater. Res. **8**, 1558–1566 (1993).
[3]A. Briggs, *Acoustic Microscopy* (Clarendon, Oxford, UK, 1992).

[4]E. P., Papadakis, "Revised grain-scattering formulas and tables," J. Acoust. Soc. Am. **37**, 703–710 (1965).
[5]R. L. Roderick and R. Truell, "The measurement of ultrasonic attenuation in solids by pulse technique and some results in steel," J. Appl. Phys. **23**, 267–279 (1952).
[6]W. P. Mason and H. J. McSkimin, "Attenuation and scattering of high frequency sound waves in metals and glasses," J. Acoust. Soc. Am. **19**, 464–473 (1947).
[7]W. Roth, "Scattering of ultrasonic radiation in polycrystalline metals," J. Appl. Phys. **19**, 901–910 (1948).
[8]R. B. Thompson and D. O. Thompson, "Ultrasonics in nondestructive evaluation," Proc. IEEE **73**, 1716–1755 (1985).
[9]A. B. Bhatia, "Scattering of high-frequency sound waves in polycrystalline materials," J. Acoust. Soc. Am. **31**, 16–23 (1959).
[10]T. Suzuki, A. Hikata, and C. Elbaum, "Anharmonicity due to glide motion of dislocations," J. Appl. Phys. **35**, 2761–2766 (1964).
[11]A. Hikata, B. B. Chick, and C. Elbaum, "Dislocation contribution to the second harmonic generation of ultrasonic waves," J. Appl. Phys. **36**, 229–236 (1965).
[12]W. T. Yost, J. H. Cantrell, and M. A. Breazeale, "Ultrasonic nonlinearity parameters and third order elastic constants of copper between 300 and 3° K ," J. Appl. Phys. **52**, 126–128 (1981).
[13]J. H. Cantrell and W. T. Yost, "Acoustic harmonic generation from fatigue-induced dislocation dipoles," Philos. Mag. A **69**, 315–326 (1994).
[14]D. H. Hurley, K. L. Telschow, and D. Cottle, "Probing acoustic nonlinearity on length scales comparable to material grain dimensions," Ultrasonics **40**, 617–620 (2002).
[15]C. B. Scruby and L. E. Drain, *Laser Ultrasonics, Techniques and Applications* (Hilger, New York, 1990).
[16]S. J. Davies, C. Edwards, G. S. Taylor, and S. B. Palmer, "Laser generated ultrasound: Its properties, mechanisms and multifarious applications," J. Phys. D **26**, 329–348 (1993).
[17]R. J. Dewhurst, C. Edwards, A. D. W. Mckie, and S. B. Palmer, "A remote laser system for ultrasonic velocity measurement at high temperatures," J. Appl. Phys. **63**, 1225–1227 (1988).
[18]J. D. Aussel and J. P. Monchalin, "Measurement of ultrasound attenuation by laser ultrasonics," J. Appl. Phys. **65**, 2918–2922 (1989).
[19]J. B. Spicer and S. L. Wallace, "Laser ultrasonic monitoring of aluminum microstructural evolution," J. Mater. Sci. **33**, 3899–3906 (1998).
[20]B. F. Pouet and N. J. P. Rasolofosaon, "Measurement of broadband intrinsic ultrasonic attenuation and dispersion in solids with laser techniques," J. Acoust. Soc. Am. **93**, 1286–1292 (1993).
[21]W. Gao, C. Glorieux, S. E. Kruger, K. Van de Rostyne, V. Gusev, W. Lauriks, and J. Theon, "Investigation of microstructure of cast iron by laser ultrasonic surface wave spectroscopy," Mater. Sci. Eng., A **313**, 170–179 (2001).
[22]A. Moreau and M. Lord, "High frequency laser ultrasonics," in *Nondestructive Characterization of Materials VIII*, edited by R. E. Green, Jr. (Plenum, New York, 1998), pp. 27–32.
[23]F. Blanchard, F. Martin, J. C. Keiffer, F. Vidal, N. Perret, T. W. Johnston, A. Blouin, A. Moreau, J. P. Monchalin, M. Choquet, and B. L. Fontaine, "High frequency ultrasound generation using a femtosecond laser," AIP Conf. Proc. **657**, 319–325 (2003).
[24]L. R. F. Rose, "Point-source representation of laser-generated ultrasound," J. Acoust. Soc. Am. **75**, 723–732 (1984).
[25]J. B. Spicer, A. D. W. Mckie, and J. W. Wagner, "Quantitative theory of laser ultrasonic waves in a thin plate," Appl. Phys. Lett. **57**, 1882–1884 (1990).
[26]J. A. Rogers and K. A. Nelson, "Study of Lamb acoustic waveguide modes in unsupported polymide thin films using real-time impulsive stimulated thermal scattering," J. Appl. Phys. **75**, 1534–1556 (1994).
[27]A. Cheng, T. W. Murray, and J. D. Achenbach, "Simulation of laser-generated ultrasonic waves in layered plates," J. Acoust. Soc. Am. **110**, 848–855 (2001).
[28]D. Nicoletti, N. Bilgutay, and B. Onaral, "Power-law relationships between the dependence of ultrasonic attenuation on wavelength and the grain size distribution," J. Acoust. Soc. Am. **91**, 3278–3284 (1992).
[29]A. S. Nowick and B. S. Berry, *Anelastic Relaxation in Crystalline Solids* (Academic, New York, 1972).
[30]*American Physics Handbook* (McGraw-Hill, New York, 1972).
[31]W. Nowacki, *Dynamic Problems of Thermoelasticity* (Noordhoff International, Leyden, 1976).
[32]A. Granato and K. Lucke, "Theory of mechanical damping due to dislocations," J. Appl. Phys. **27**, 583–593 (1956).
[33]K. Aki and P. G. Richards, *Quantitative Seismology: Theory and Methods* (Freeman, San Francisco, 1980), Vol. **1**.
[34]J. B. Spicer and D. H. Hurley, "Epicentral and near epicenter surface displacements on pulse laser irradiated metallic surfaces," Appl. Phys. Lett. **68**, 3561–3563 (1996).
[35]R. M. Christensen, *Theory of Viscoelasticity* (Academic, New York, 1971).
[36]M. de Billy, G. Quentin, and E. Baron, "Attenuation measurements of an ultrasonic Rayleigh wave propagating along rough surfaces," J. Appl. Phys. **61**, 2140–2145 (1987).

On-axis and far-field sound radiation from resilient flat and dome-shaped radiators

Ronald M. Aarts[a] and Augustus J. E. M. Janssen[b]
Philips Research Europe, HTC 36 (WO-02), NL-5656AE Eindhoven, The Netherlands

(Received 25 July 2008; revised 23 December 2008; accepted 24 December 2008)

On-axis and far-field series expansions are developed for the sound pressure due to an arbitrary, circular symmetric velocity distribution on a flat radiator in an infinite baffle. These expansions are obtained by expanding the velocity distributions in terms of orthogonal polynomials $R_{2n}^0(\sigma/a) = P_n(2(\sigma/a)^2-1)$ with P_n the Legendre polynomials. The terms R_{2n}^0 give rise to a closed-form expression for the pressure on-axis as well as for the far-field pressure. Furthermore, for a large number of velocity profiles, including those associated with the rigid piston, the simply supported radiator, and the clamped radiators as well as Gaussian radiators, there are closed-form expressions for the required expansion coefficients. In particular, for the rigid, simply supported, and clamped radiators, this results in explicit finite-series expressions for both the on-axis and far-field pressures. In the reverse direction, a method of estimating velocity distributions from (measured) on-axis pressures by matching in terms of expansion coefficients is proposed. Together with the forward far-field computation scheme, this yields a method for far-field loudspeaker assessment from on-axis data (generalized Keele scheme). The forward computation scheme is extended to dome-shaped radiators with arbitrary velocity distributions.
[DOI: 10.1121/1.3075594]

PACS number(s): 43.38.Ar, 43.20.Bi, 43.40.At, 43.20.Px [AJZ] Pages: 1444–1455

I. INTRODUCTION

In this paper a new analytic method for the calculation of on-axis and far-field sound pressures is presented. The theory of sound radiation from a flat or dome-shaped radiator in a rigid infinite baffle until 1980 is broadly reviewed by Harris[1] while a whole set of analytic results for the circular symmetric flat piston radiator has been given by Greenspan.[2] Recent analytical and/or numerical efforts have been undertaken by Mast and Yu,[3] Hansen,[4] Mellow,[5] and Kelly and McGough,[6] to name just a few. The point of view taken in the present paper, viz., the applications to sound radiation of the analytical results as developed in the diffraction theory of optical aberrations by Nijboer[7] and Zernike and Nijboer[8] (also see Refs. 9 and 10) is, however, apparently new. Using this new approach, many of the analytic results in Ref. 2, such as those on on-axis pressure and those on reaction on radiator and on radiated power, and the results in textbooks[11] on far-field expressions and directivity can be presented and extended in a systematic fashion. The aim of the present paper is to work out this approach for the results on on-axis pressure and far-field expressions for arbitrary velocity distributions on both flat piston radiators and dome-shaped radiators. The radiated pressure is given in integral form by the Rayleigh integral[11,12] as

$$p(\underline{r},t) = \frac{i\rho_0 ck}{2\pi} e^{i\omega t} \int_S v(\underline{r}_s) \frac{e^{-ikr'}}{r'} dS, \tag{1}$$

where ρ_0 is the density of the medium, c is the speed of sound in the medium, $k=\omega/c$ is the wave number, and ω is the radian frequency of the vibrating surface S. Furthermore t is time, $\underline{r}$ is a field point, $\underline{r}_s$ is a point on the surface S, $r' = |\underline{r}-\underline{r}_s|$ is the distance between $\underline{r}$ and $\underline{r}_s$, and $v(\underline{r}_s)$ is the normal component of a (not necessarily uniform) velocity profile on the surface S. The time variable t in $p(\underline{r},t)$ and the harmonic factor $\exp(i\omega t)$ in front of the integral in Eq. (1) will be omitted in the sequel. For transparency of exposition, the surface S is assumed initially to be a disk of radius a, $|\underline{r}_s| = r_s \leq a$, with average velocity V_s; later on, generalization to the case of dome-shaped radiator surfaces S is done. See Fig. 1 for the geometry and notations used in the case of a flat piston. The volume velocity at the piston is

$$\int_S v(\underline{r}_s) dS = V_s \pi a^2. \tag{2}$$

Frankort[13] showed that loudspeaker cones mainly vibrate in a radially symmetric fashion. Therefore the attention in this paper is restricted to radially symmetric velocity distributions v, which are denoted as $v(\sigma)$, $0 \leq \sigma \leq a$. Under an integrability condition, viz., $\int_S |v(\underline{r}_s)|^2 dS < \infty$, these v's admit a representation

$$v(\sigma) = V_s \sum_{n=0}^{\infty} u_n R_{2n}^0(\sigma/a), \quad 0 \leq \sigma \leq a, \tag{3}$$

in which

[a]Electronic mail: ronald.m.aarts@philips.com. Also at Technical University Eindhoven, Den Dolech 2, PT3.23, P.O Box 513, NL-5600 MB Eindhoven, The Netherlands.
[b]Electronic mail: a.j.e.m.janssen@philips.com

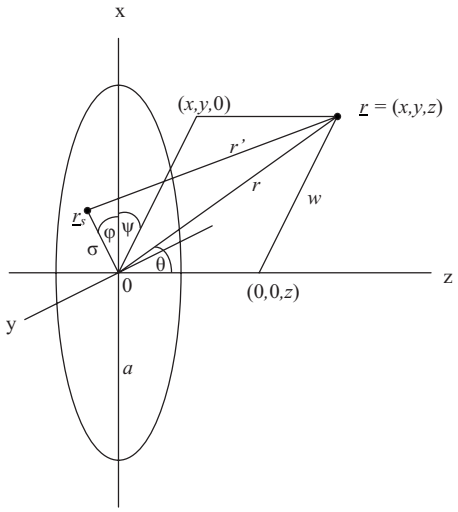


FIG. 1. Setup and notations.

$$\underline{r}_s = (x_s, y_s, 0) = (\sigma \cos\varphi, \sigma \sin\varphi, 0),$$

$$\underline{r} = (x, y, z) = (r \sin\theta \cos\psi, r \sin\theta \sin\psi, r \cos\theta),$$

$$w = r \sin\theta = (x^2 + y^2)^{1/2}, \quad z = r \cos\theta,$$

$$r = |\underline{r}| = (x^2 + y^2 + z^2)^{1/2} = (w^2 + z^2)^{1/2},$$

$$r' = |\underline{r} - \underline{r}_s| = (r^2 + \sigma^2 - 2\sigma w \cos(\psi - \varphi))^{1/2}$$

$$R_{2n}^0(\rho) = P_n(2\rho^2 - 1), \quad 0 \le \rho \le 1, \tag{4}$$

where P_n are the Legendre polynomials.[14] [The notation $R_{2n}^0(\rho)$ stems from the analytical theory of optical aberrations.[7–9]] The main aim of the present paper is to establish analytical results for the on-axis and far-field pressures $p(\underline{x})$ in Eq. (1) related to the coefficients u_n and polynomials R_{2n}^0 occurring in the expansion in Eq. (3).

By orthogonality of the terms $R_{2n}^0(\rho)$, the coefficients u_n in Eq. (3) can be found in integral form as

$$u_n = \frac{2(2n+1)}{V_s} \int_0^1 R_{2n}^0(\rho) v(a\rho) \rho d\rho, \quad n = 0, 1, \ldots . \tag{5}$$

An expansion of the type as in Eq. (3) is usually referred to as a Zernike[0] expansion; in this paper only azimuthal order $m=0$ occurs, and so the superscript 0 is dropped. There is an impressive number of cases where one can explicitly find the u_n in Eq. (5); some of these appear in the present paper and include the rigid, simply supported, and clamped radiators

$$v^{(\ell)}(\sigma) = (\ell + 1) V_s (1 - (\sigma/a)^2)^\ell H(a - \sigma), \quad \ell = 0, 1, \ldots , \tag{6}$$

and the Gaussian velocity profile

$$v(\sigma; \alpha) = \frac{\alpha V_s}{1 - e^{-\alpha}} e^{-\alpha(\sigma/a)^2} H(a - \sigma), \tag{7}$$

where $H(x)$ is the Heaviside function, i.e., $H(x)=0$, $1/2$, or 1 according as x is negative, zero, or positive. Hence, the velocity profiles in Eqs. (6) and (7) vanish for $\sigma > a$.

The relevance of the Zernike terms R_{2n}^0 for the purposes of the present paper is the existence of closed-form formulas, involving spherical Bessel and Hankel functions and Bessel functions of the first kind, respectively, for the on-axis pressure and for the far-field pressure due to a velocity profile described by the term R_{2n}^0. Thus, by linearity in Eq. (3), one can compute the on-axis and far-field pressures due to a velocity profile v, once its expansion coefficients u_n are available. For instance, the radiators in Eq. (6) give rise to an on-axis pressure expansion in the form of a series of $n+1$ terms $u_\ell j_\ell(kr_-) h_\ell^{(2)}(kr_+)$, with $r_\pm$ argument values directly related to the axial position $(0,0,r)$, while the far-field pressure expansion is a similar series involving terms $u_\ell J_{2\ell+1}(ka \sin\theta)/(ka \sin\theta)$. In the reverse direction, the forward computation schemes for the on-axis and far-field pressures can be complemented by an inverse method with potential use in far-field loudspeaker assessment. Here one estimates the expansion coefficients u_n of a velocity profile v by matching with a measured on-axis pressure data set and then one predicts the far-field sound radiation using the far-field forward formula. Because the on-axis pressure data can be collected in the (relative) near-field of a loudspeaker, this avoids the use of anechoic rooms that would be necessary if the far-field were to be assessed directly. The fact that Zernike series are so efficient in representing velocity profiles is very instrumental here: A smooth velocity profile can already be represented adequately by as few as two to five terms (see Appendix A 2 where this is shown for the truncated Gaussian as an example).

An issue that must be addressed is the following: The set of velocity profiles in Eq. (6) is well known and has been studied in considerable detail, and has in principle the same potential for the purposes of this paper as the set of profiles associated with the Zernike terms. Indeed, closed-form expressions for the far-field pressure due to $v^{(n)}$ were found by Stenzel,[15] and analytical expressions for the on-axis pressure due to the first few $v^{(n)}$ are given by Dekker *et al.*[16] and by Greenspan.[2] However, the formulas for the on-axis pressure in the references given become quite complicated, even for values of n as low as 1 or 2. Furthermore, while Greenspan[2] noted that any polynomial or power series in $(\sigma/a)^2$ can be expanded as a linear combination of the functions $(1 - (\sigma/a)^2)^n$, these expansions require relatively large coefficients due to nearly linear dependence of the expansion functions. Because the Zernike terms are orthogonal, it can be expected that the Zernike expansions behave much better with this respect. Indeed, according to Appendix A, the nth coefficient of the expansion into $(1-(\sigma/a)^2)^n$ is roughly a factor 4^n larger than the nth coefficient in the Zernike expansion [note that the functions $(1-(\sigma/a)^2)^n$ and $R_{2n}^0(\sigma/a)$ have the same mean square value]. This point is of particular importance for the inverse problem, formulated above, of estimating velocity profiles from the on-axis pressure data by matching series representation coefficients. For these two reasons, complicated on-axis pressure expressions and nearly linear dependence, the expansion of velocity profiles in terms of Zernike terms is preferred in this paper.

II. PAPER OUTLINE

In Sec. III the definition and basic properties of the Zernike terms R_{2n}^0 are given, and some of the expansion results that are relevant for this paper are presented. Furthermore, the Hankel transform of R_{2n}^0 is presented in closed form. The latter result is of importance both for the forward computation scheme for the far-field and for establishing results on reaction on radiator, radiated power, etc., that are related to King's integral. More information on Zernike expansions is contained in Appendix A.

In Sec. IV the basic formulas for the flat piston are highlighted and discussed. Thus, the closed form involving a spherical Bessel and Hankel function for the on-axis pressure associated with a single term R_{2n}^0 is presented, with comments on both near-field and far-field behaviors and on behavior for small and large values of ka. Also, in Sec. IV, the far-field expression, in terms of a Bessel function of the first kind, of the pressure associated with a single term R_{2n}^0 is presented with particular attention given to the case that n gets large or that ka gets large, etc. In Sec. V the inverse method of estimating the Zernike expansion coefficients of the velocity profile from the on-axis (measured, sampled) pressure data is considered. Together with the forward scheme for computing far-field pressures from Zernike expansions, this yields a loudspeaker assessment method that generalizes a well-known method in audio engineering for estimating the far-field of a loudspeaker from near-field on-axis data in the case of a rigid piston (Keele scheme[17]). In this paper only a few measurements and simulation results will be shown with attention given to more fundamental issues such as the choice of the on-axis sampling scheme and possible ill-posedness of the inversion method.

In Sec. VI the analytic expression for the on-axis pressure developed for the flat piston is extended to dome-shaped radiators. This extension is feasible due to a unique property of the assumed dome profiles, and forms the basis for the characterization of velocity profiles on dome-shaped radiators from on-axis pressure data.

III. THE ZERNIKE TERMS R_{2n}^0

The Zernike terms R_{2n}^0 are polynomials of degree $2n$ given by

$$R_{2n}^0(\sigma/a) = P_n(2(\sigma/a)^2 - 1) = \sum_{s=0}^{n} (-1)^s \binom{2n-s}{n} \binom{n}{s} \times (\sigma/a)^{2n-2s}, \tag{8}$$

where P_n is the Legendre polynomial of degree n, see Ref. 14 (22.3.8 and 22.5.42). The first few R_{2n}^0 are given in Table I, and in Fig. 2 some of them are plotted as a function of $\rho = \sigma/a \in [0,1]$. The R_{2n}^0 cannot be interpreted directly in physical terms, unlike the velocity profiles $v^{(\ell)}$ in Eq. (6) in which ℓ has the interpretation of a smoothness parameter for the transition from the non-zero values on the piston ($\sigma < a$) to 0 outside the piston ($\sigma > a$). Rather, their significance for loudspeaker analysis stems from the following facts.

TABLE I. Zernike polynomials.

n	$R_{2n}^0(\sigma/a)$
0	1
1	$2(\sigma/a)^2-1$
2	$6(\sigma/a)^4-6(\sigma/a)^2+1$
3	$20(\sigma/a)^6-30(\sigma/a)^4+12(\sigma/a)^2-1$

- They are very efficient and convenient in representing a general velocity profile v. This is due to the orthogonality property

$$\int_0^1 R_{2n_1}^0(\rho) R_{2n_2}^0(\rho) \rho d\rho = \frac{\delta_{n_1 n_2}}{2(2n_1+1)} \tag{9}$$

(where δ is Kronecker's delta), as well as the fact that many velocity profiles considered in loudspeaker analysis can be represented as a Zernike series. In Appendix A, a number of cases are listed, such as the expansion

$$(1-(\sigma/a)^2)^\ell = \sum_{n=0}^{\ell} (-1)^n \frac{2n+1}{n+1} \frac{\binom{\ell}{n}}{\binom{\ell+n+1}{\ell}} R_{2n}^0(\sigma/a), \tag{10}$$

which are relevant for the rigid and simply supported ($\ell = 0,1$) and the clamped radiators ($\ell \geqslant 2$) in Eq. (6), and the expansion

$$e^{-\alpha\sigma^2} = e^{-(1/2)\alpha} \sum_{n=0}^{\infty} (-1)^n (2n+1) \sqrt{\frac{\pi}{\alpha}} I_{n+1/2}(\alpha/2) R_{2n}^0(\sigma/a), \tag{11}$$

which is relevant for the truncated Gaussian radiator in Eq. (7).

- An expansion result of direct relevance to the Rayleigh integral for on-axis field points is the formula

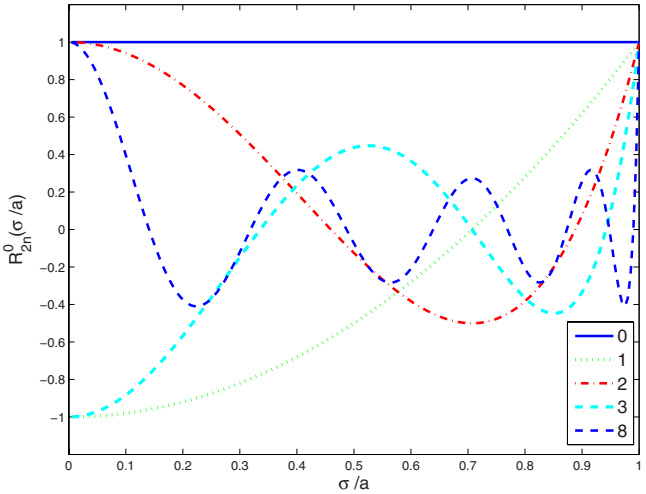


FIG. 2. (Color online) The Zernike terms R_{2n}^0 vs σ/a for various values of n.

$$\frac{e^{-i\lambda(T^2+x\rho^2)^{1/2}}}{-i\lambda(T^2+x\rho^2)^{1/2}} = \sum_{n=0}^{\infty}(2n+1)(-1)^n j_n(\lambda T_-)h_n^{(2)}(\lambda T_+) \times R_{2n}^0(\rho), \tag{12}$$

in which $x \in \mathbb{R}$, $T>0$, $T^2+x>0$, and

$$T_\pm = \tfrac{1}{2}[(T^2+x)^{1/2} \pm T]. \tag{13}$$

This result can be applied to both flat radiators (Sec. IV A) and dome-shaped radiators (Sec. VI). See Appendix A for the proof of Eq. (12) and Sec. IV A for more details on the spherical Bessel functions j_n and spherical Hankel functions $h_n^{(2)}$ occurring in Eq. (12).

- The Hankel transform of zeroth order of the R_{2n}^0 has a closed form, viz.,

$$\int_0^a J_0(u\sigma)R_{2n}^0(\sigma)\sigma d\sigma = (-1)^n \frac{a}{u}J_{2n+1}(ua). \tag{14}$$

This formula has been proved in Ref. 7 as a special case of a formula expressing the mth order Hankel transform of Zernike polynomials of azimuthal order m in terms of Bessel functions of the first kind. This formula is very important for the development of explicit analytic results in the spirit of Ref. 2. For the purposes of the present paper, the result is important since it gives the far-field expression for the pressure due to a single term R_{2n}^0 in the velocity profile, see Sec. IV B and Appendix B.

IV. ON-AXIS AND FAR-FIELD EXPRESSIONS FOR THE FLAT PISTON

The velocity profile $v(\sigma)$ considered in this section (normal component) vanishes outside the disk $\sigma \leqslant a$ and has been developed into a Zernike series as in Eq. (3) with coefficients u_n given in accordance with Eq. (5) or explicitly as in the cases discussed earlier.

A. On-axis expression

There holds for an on-axis point $\underline{r}=(0,0,r)$ with $r \geqslant 0$ by radial symmetry of the integrand in Eq. (1) the formula

$$p(\underline{r}) = i\rho_0 ck \int_0^a v(\sigma)\frac{e^{-ik(r^2+\sigma^2)^{1/2}}}{(r^2+\sigma^2)^{1/2}}\sigma d\sigma. \tag{15}$$

Inserting $v(\sigma)=V_s\Sigma_{n=0}^{\infty}u_nR_{2n}^0(\sigma/a)$ into the integral and setting $\sigma=a\rho$, $0 \leqslant \rho \leqslant 1$, it follows that

$$p(\underline{r}) = i\rho_0 cka^2V_s\sum_{n=0}^{\infty}u_n\int_0^1 \frac{e^{-ik(r^2+a^2\rho^2)^{1/2}}}{(r^2+a^2\rho^2)^{1/2}}R_{2n}^0(\rho)\rho d\rho. \tag{16}$$

Then by Eq. (12) and orthogonality of the Zernike terms,

$$p(\underline{r}) = \frac{1}{2}\rho_0 cV_s(ka)^2\sum_{n=0}^{\infty}\gamma_n(k,r)u_n, \tag{17}$$

in which

$$\gamma_n(k,r) = (-1)^n j_n(kr_-)h_n^{(2)}(kr_+),$$

$$r_\pm = \tfrac{1}{2}(\sqrt{r^2+a^2} \pm r). \tag{18}$$

The j_n and $h_n^{(2)}=j_n-iy_n$ are the spherical Bessel and Hankel functions, respectively, of the order $n=0,1,\ldots$, see Ref. 14 (Sec. 10.1). In particular, $j_0(z)=(\sin z)/z$ and $h_0^{(2)}(z)=(ie^{-iz})/z$.

What follows now is a discussion of the results in Eqs. (17) and (18). The $r_\pm$ of Eq. (18) satisfy

$$0 \leq r_- \leq \tfrac{1}{2}a \leq r_+, \quad r_+r_- = \tfrac{1}{4}a^2, \quad r_+ + r_- = \sqrt{r^2+a^2}. \tag{19}$$

Consider the case of the rigid piston, i.e., $\ell=0$ in Eq. (10). Then $u_0=1$, $u_1=u_2=\cdots=0$, and from Eqs. (17)–(19) it is found, using $j_0(z)=\sin z/z$ and $h_0^{(2)}(z)=ie^{-iz}/z$ and some administration, that

$$p(\underline{r}) = \frac{1}{2}\rho_0 cV_s(ka)^2\frac{\sin kr_-}{kr_-}\frac{ie^{-ikr_+}}{kr_+} = 2i\rho_0 cV_s e^{-(1/2)ik((r^2+a^2)^{1/2}+r)}\sin\frac{1}{2}k((r^2+a^2)^{1/2}-r). \tag{20}$$

This is the classical result on the on-axis pressure for a rigid piston as can be found in the textbooks, see, e.g., Ref. 11 [Eqs. (8.31a) and (8.31b)]. Figure 3 shows a plot of $|\gamma_{\ell=n=0}(k,r)|$ as a function of r/a (rigid piston) and of $|\gamma_n(k,r)|$ for $n=1,2,3$. Some comments on these plots are presented at the end of this subsection.

For the simply supported radiator, case $\ell=1$ in Eq. (6), one has $u_0=1$ and $u_1=-1$ so that

$$p(\underline{r}) = \tfrac{1}{2}\rho_0 cV_s(ka)^2[j_0(kr_-)h_0^{(2)}(kr_+) + j_1(kr_-)h_1^{(2)}(kr_+)], \tag{21}$$

and for the lowest order clamped radiator, case $\ell=2$ in Eq. (6), one has $u_0=1$, $u_1=-3/2$, and $u_2=1/2$, so that

$$p(\underline{r}) = \tfrac{1}{2}\rho_0 cV_s(ka)^2\left[j_0(kr_-)h_0^{(2)}(kr_+) + \tfrac{3}{2}j_1(kr_-)h_1^{(2)}(kr_+) + \tfrac{1}{2}j_2(kr_-)h_2^{(2)}(kr_+)\right]. \tag{22}$$

The results of Eqs. (21) and (22) generalize to higher ℓ immediately since the required coefficients u for the general case are available through Eq. (10). Also see Ref. 6, Eqs. (9)–(11), where the expression for $\ell=1,2$ leads to considerable complications. In Fig. 4 the rigid piston ($\ell=0$), the simply supported radiator ($\ell=1$), and the first two clamped radiators ($\ell=2,3$) are considered ($|p(\underline{r})|$, normalized as a function of r/a).

The next comments concern the behavior of the terms γ_n in Eq. (18). From Eqs. (18) and (19) it follows that

$$r_- \approx \tfrac{1}{2}(a-r) \approx \tfrac{1}{2}a \approx \tfrac{1}{2}(a+r) \approx r_+, \quad r \ll a. \tag{23}$$

Therefore, when ka is large and $r \to 0$ (with n not large), it follows from the results in Ref. 14, Sec. 10.1 that

$$|\gamma_n(k,r)| \approx \frac{|\cos\frac{1}{2}k(a-r)|}{\frac{1}{4}k^2a^2}, \tag{24}$$

confirming the presence of zeros and the largely n-independent envelope of the curves in Fig. 3 near $r=0$.

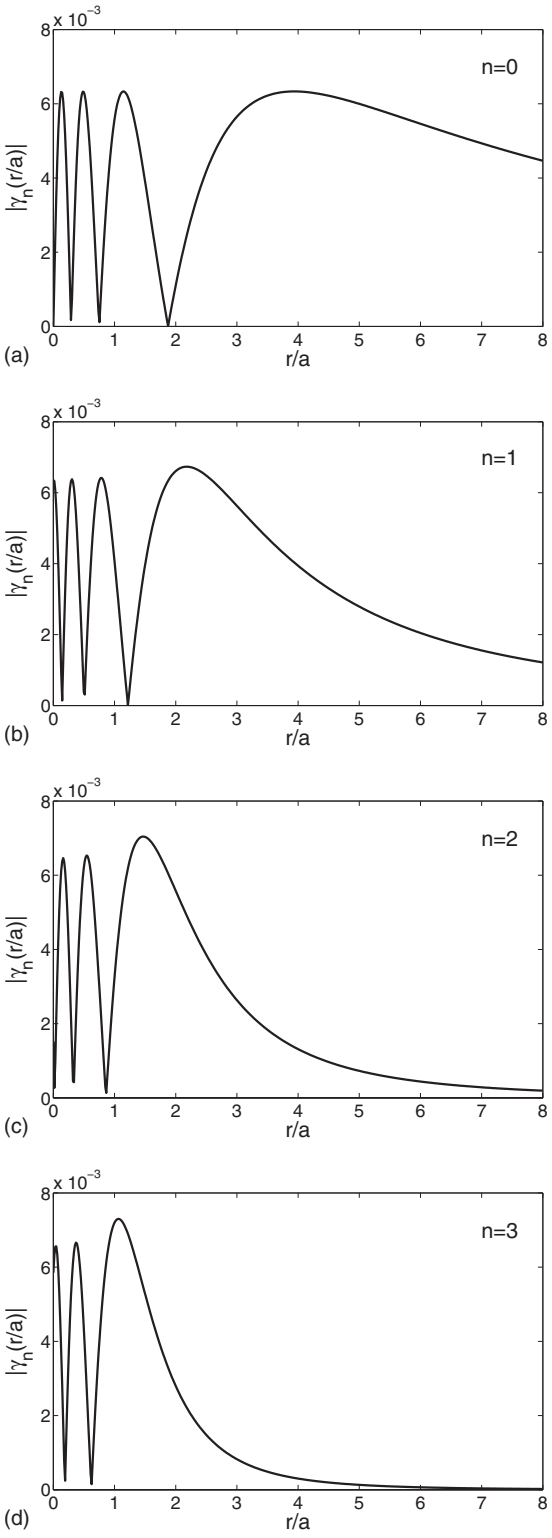


FIG. 3. The product $|j_n(kr_-)h_n^{(2)}(kr_+)|$ from Eq. (18), for $n=0,\ldots,3$, vs r/a, where $a/\lambda=4$ and $a=0.1$ m, which yields $f=13.7$ kHz and $ka=8\pi$. (a) $n=0$, (b) $n=1$, (c) $n=2$, and (d) $n=3$.

Finally, when $r \gg a$, it follows from Eqs. (18) and (19) that

$$r_- \approx \frac{a^2}{4r}, \quad r_+ \approx r. \tag{25}$$

Therefore, from Ref. 14, Sec. 10.1,

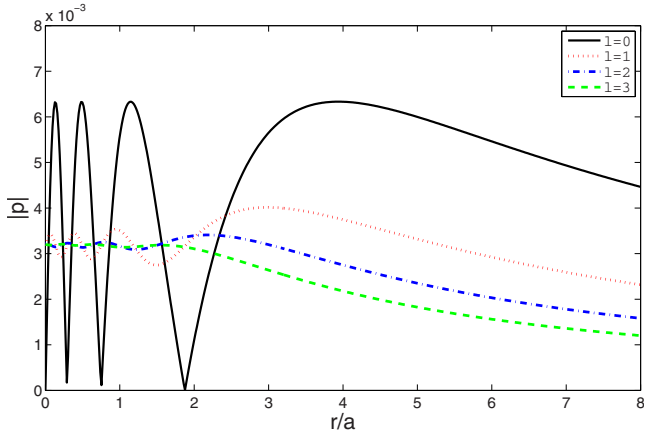


FIG. 4. (Color online) Normalized $|p|$ vs r/a for the rigid piston ($\ell=0$) (solid curve), the simply supported radiator ($\ell=1$) (dotted curve), and the first two clamped radiators ($\ell=2,3$) (dashed-dotted and dashed curves, respectively) using Eq. (17). Here $a/\lambda=4$ and $ka=8\pi$. The normalization is equal to $(\ell+1)/2\rho_0 c V_s (ka)^2$. The factor $\ell+1$ allows an easier comparison of the four curves.

$$\gamma_n(k,r) \approx \frac{\left(\frac{-ika^2}{4r}\right)^n}{1\times 3\cdots(2n+1)} \frac{e^{-ikr_+}}{-ikr_+}, \tag{26}$$

which shows an $O(1/r^{n+1})$-behavior of $\gamma_n(k,r)$ as $n\to\infty$.

B. Far-field expression

Using the Zernike expansion (3) of $v(\sigma)$ it is shown in Appendix B that the following far-field approximation holds: When $\underline{r}=(r\sin\theta, 0, r\cos\theta)$ and $r\to\infty$,

$$p(\underline{r}) \approx i\rho_0 ckV_s \frac{e^{-ikr}}{r} a^2 \sum_{n=0}^{\infty} u_n (-1)^n \frac{J_{2n+1}(ka\sin\theta)}{ka\sin\theta}. \tag{27}$$

The result of Eq. (27) will now be discussed. In the case of a rigid piston, it follows that

$$p(\underline{r}) \approx i\rho_0 ckaV_s \frac{e^{-ikr}}{r} \frac{J_1(ka\sin\theta)}{ka\sin\theta}. \tag{28}$$

This is the familiar result for the far-field pressure of a rigid piston as can be found in the textbooks, see, e.g., Ref. 11 [Eq. (8.35)]. In Fig. 5 a plot can be found of $|J_{2n+1}(ka\sin\theta)/ka\sin\theta|$, $n=0,1,2,3$.

For the simply supported radiator, case $\ell=1$ in Eq. (6), and for the clamped radiators, cases $\ell\geqslant 2$ in Eq. (6), there is the far-field result of Stenzel,[15]

$$p(\underline{r}) \approx i\rho_0 cV_s ka^2(\ell+1)!2^\ell \frac{e^{-ikr}}{r} \frac{J_{\ell+1}(ka\sin\theta)}{(ka\sin\theta)^{\ell+1}}; \tag{29}$$

also see Ref. 2 (Sec. II).

Alternatively, the coefficients u in the Zernike expansion of the $v^{(\ell)}$ in Eq. (6) are available per Eq. (10), and this gives the far-field approximation of $p(\underline{r})$ via Eq. (27) in a form different from Eq. (29). Equating the r-independent parts of the right-hand sides (RHSs) of the two expressions leads to a non-obvious analytic relation between Bessel functions,

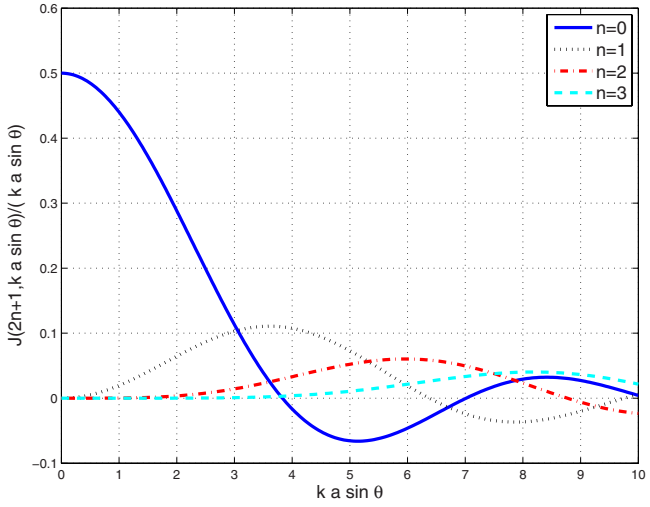


FIG. 5. (Color online) $|J_{2n+1}(ka\sin\theta)/ka\sin\theta|$ vs $ka\sin\theta$.

$$\ell!2^{\ell}\frac{J_{\ell+1}(x)}{x^{\ell+1}}=\sum_{n=0}^{\ell}\frac{2n+1}{n+1}\frac{\binom{\ell}{n}}{\binom{\ell+n+1}{\ell}}\frac{J_{2n+1}(x)}{x}, \tag{30}$$

and either method yields the result shown in Fig. 6.

Some comments on the behavior of the terms $J_{2n+1}(z)/z$, $z=ka\sin\theta$, as they occur in the series in Eq. (27) are presented now. From the asymptotics of the Bessel functions, as given in Ref. 14 [Eq. (9.3.1)], it is seen that in the series in Eq. (27) only those terms contribute significantly for which $2n+1\leqslant\frac{1}{2}e\,ka\sin\theta$. In particular, when $\theta=0$, it is only the term with $n=0$ that is non-vanishing, and this yields

$$p((0,0,r))\approx\frac{1}{2}i\rho_0cV_ska^2\frac{e^{-ikr}}{r},\quad r\to\infty. \tag{31}$$

This is in agreement with what is found from Eq. (17) when only the term with $n=0$ is retained and r_+ is replaced with r, and r_- is replaced with 0. For small values of ka the terms in the series (27) decay very rapidly with n. For large values of ka, however, a significant number of terms may contribute, especially for angles θ far from 0.

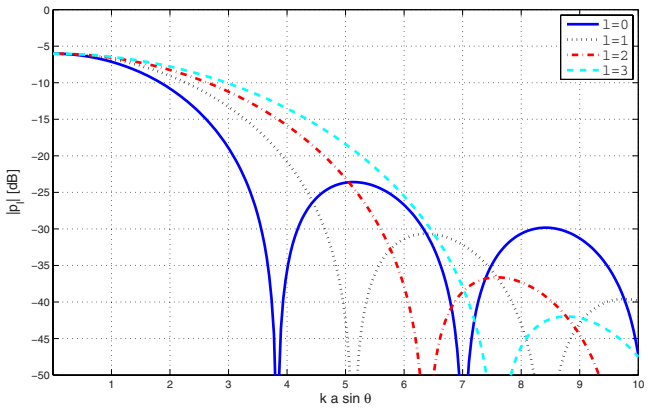


FIG. 6. (Color online) Normalized $|p|$ vs $ka\sin\theta$, using the Zernike expansion of the $v^{(\ell)}$ and Eq. (27) or Eq. (29).

V. ESTIMATING VELOCITY PROFILES FROM ON-AXIS RADIATION DATA FOR FAR-FIELD LOUDSPEAKER ASSESSMENT

A. Estimating velocity profiles from on-axis radiation

The on-axis expressions (17) and (18) for the pressure can, in reverse direction, be used to estimate the velocity profile on the disk from (measured) on-axis data via its expansion coefficients u_n. This can be effectuated by adopting a matching approach in which the coefficients u_n in the "theoretical" expressions (17) and (18) are determined so as to optimize the match with at $M+1$ points measured data. Thus, one has for the pressure $p_m=p((0,0,r_m))$ due to the velocity profile $v(\sigma)=V_s\Sigma_{n=0}^{N}u_nR_{2n}^0(\sigma/a)$ the expression

$$p_m=\frac{1}{2}\rho_0cV_s(ka)^2\sum_{n=0}^{N}(-1)^nj_n(kr_{m,-})h_n^{(2)}(kr_{m,+})u_n, \tag{32}$$

where $r_m\geqslant 0$ and

$$r_{m,\pm}=\tfrac{1}{2}(\sqrt{r_m^2+a^2}\pm r_m), \tag{33}$$

and $m=0,1,\ldots,M$. With

$$A=(A_{mn})_{\substack{m=0,1,\ldots,M,\\ n=0,1,\ldots,N}},$$

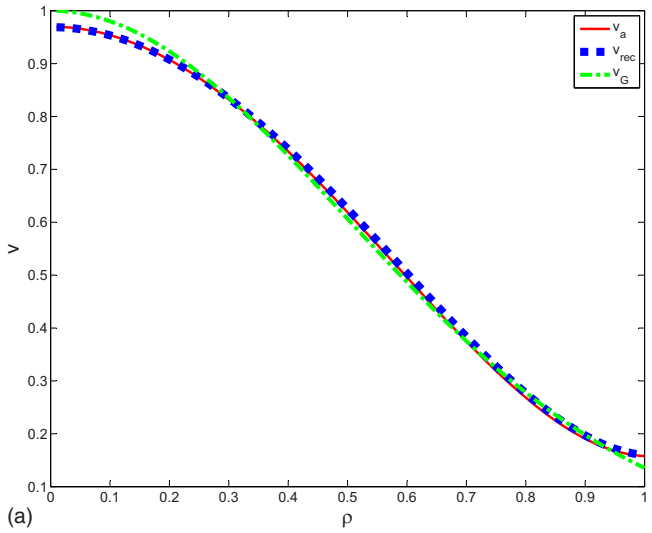
$$A_{mn}=\tfrac{1}{2}\rho_0cV_s(ka)^2j_n(kr_{m,-})h_n^{(2)}(kr_{m,+}), \tag{34}$$

$$\underline{p}=[p_0,\ldots,p_M]^T,\quad \underline{u}=[u_0,\ldots,u_N]^T, \tag{35}$$

the relation between on-axis pressures p_m and coefficients u_n can be concisely written as

$$A\underline{u}=\underline{p}. \tag{36}$$

Now given a (noisy) on-axis data vector $\underline{p}$ one can estimate the coefficients' vector $\underline{u}$ by adopting a least mean-squares approach for the error $A\underline{u}-\underline{p}$. This will be illustrated by a simulated experiment and, subsequently, by a real experiment, below. In the simulated experiment, we assume a loudspeaker with a Gaussian velocity profile ($\alpha=2$), as shown in Fig. 7(a) curve v_G (dashed-dotted) and given by the left-hand side (LHS) of Eq. (A2). This profile is approximated using three Zernike coefficients ($u_0=0.4323$, $u_1=-0.4060$, and $u_2=0.1316$) given by the RHS of Eq. (A2), and this leads to the velocity profile v_a (solid curve) in Fig. 7(a). It can be seen from Fig. 7(a) that including three Zernike terms provides a fair approximation (3×10^{-2} absolute accuracy on the whole range). Using the three coefficients of the approximated velocity profile, the sound pressure was calculated by applying Eq. (17) and plotted in Fig. 7(b) as p_{calc} (solid curve). Then random white noise was added to p_{calc} as shown as p +noise in Fig. 7(b) (dotted curve). Subsequently the inversion procedure was followed by using the noisy pressure data vector $\underline{p}$ to estimate the coefficients' vector $\underline{u}$ by adopting a least mean-squares approach for the error $A\underline{u}-\underline{p}$ [see Eq. (36)]. Using the recovered three Zernike coefficients the velocity profile and pressure data were calculated and plotted in Fig. 7(a) (thick dotted curve) and Fig. 7(b) (thick dotted curve), respectively. It appears that the inversion procedure is rather robust against noise since the calculated and recovered pressure curves in Fig. 7(b) are almost coincident.



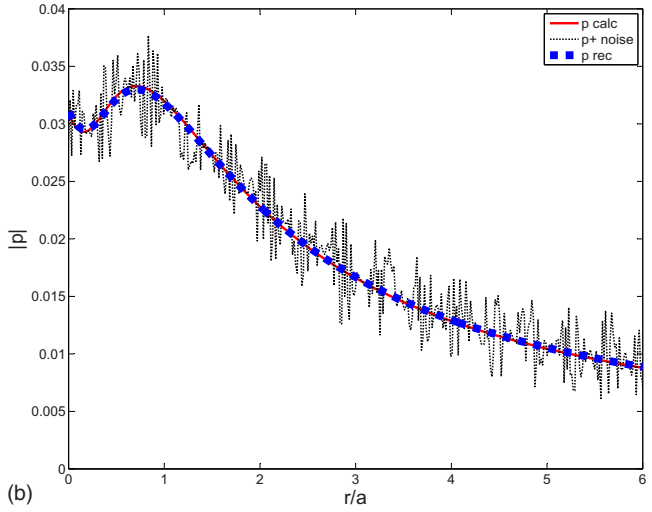


FIG. 7. (Color online) Simulated experiment. (a) Gaussian velocity profile ($\alpha=2$) v_G vs ρ [dashed-dotted curve, given by the LHS of Eq. (A2)]. Approximated velocity profile v_a using the series at the RHS of Eq. (A2), truncated at $n=2$ (solid curve). From noisy pressure data recovered velocity profile v_{rec} (thick dotted curve). (b) Sound pressure using Eq. (17) and $ka=8$ (p_{calc}, solid curve). Pressure with added noise ($p+$noise, dotted curve). Recovered pressure data (p_{rec}, thick dotted curve).

For the second experiment we measured a loudspeaker (vifa MG10SD09-08, $a=3.2$ cm) in an IEC-baffle,[18] at ten near-field positions ($\underline{r}_m=0.00,0.01,0.02,0.03,0.04,0.05,0.07,0.10,0.13,0.19$ m), and finally in the far-field at 1 m distance at 13.72 kHz ($ka=8.0423$). The magnitude of the sound pressure is plotted in Fig. 8 (solid curve, "p meas"). Using the same procedure as described above for the first experiment, the inverse process was followed by using the ten measured near-field pressure data points to estimate the coefficients' vector $\underline{u}$. Using four Zernike coefficients the pressure data were recovered and plotted in Fig. 8 (dotted curve, "p rec"). It appears that the two curves show good resemblance to each other and that only four coefficients are needed to provide a very good description of the near-field at rather high frequencies (13.72 kHz). Furthermore, it appears that using these four coefficients, the calculated sound pressure level at 1-m distance yields −42 dB. The measured value at that far-field point is −44 dB. These values match

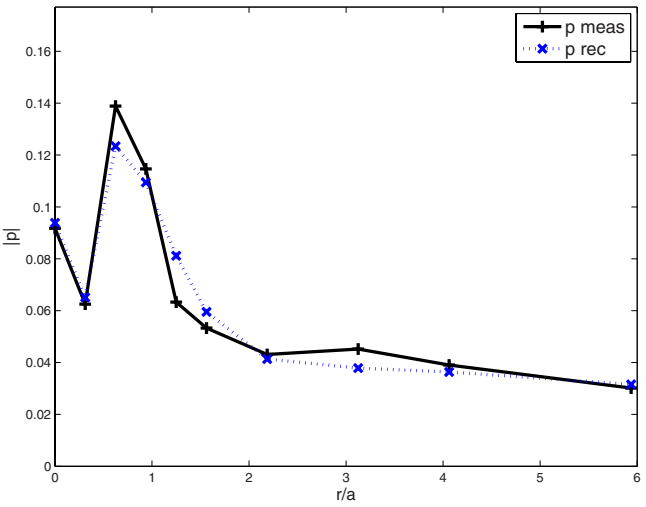


FIG. 8. (Color online) Measured loudspeaker at 13.72 kHz ($ka=8.0423$, p meas, solid curve) vs r/a. Recovered pressure data (p rec, dotted curve).

rather closely, even though at the used frequency of 13.72 kHz the cone vibrates not fully circularly symmetric anymore due to break-up behavior. This match provides a proof of principle as the far-field measurement point was not used to determine the Zernike coefficients, also see Sec. V B below.

In the experiments just described, no particular effort was spent in forming and handling the linear systems so as to have small condition numbers. The condition number, the ratio of the largest and smallest non-zero singular value of the matrix A in Eq. (34), equals 50 in the case of the loudspeaker experiment leading to Fig. 8. In Fig. 9 the condition number is plotted as a function of the number $N+1$ of columns for the case of the simulation experiment leading to Fig. 7. It is seen that the condition number grows rapidly with N. It is expected that considerably improved condition numbers result when the structure of the linear systems and their constituent functions j and h is employed; this is, however, outside the scope of the present paper. The influence of the condition number on noise-sensitivity and the like of the

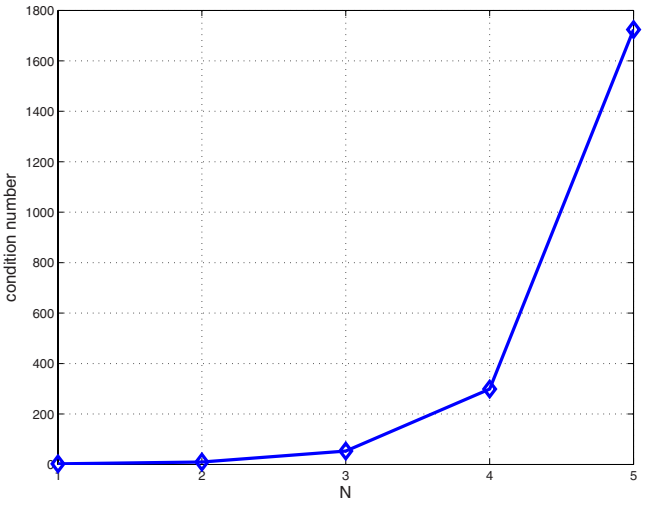


FIG. 9. (Color online) The condition numbers of the first experiment (diamonds, for $ka=8$, $\alpha=2$, and $M=500$).

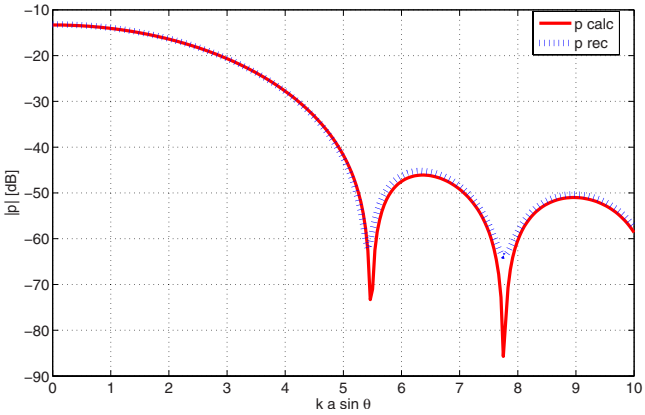


FIG. 10. (Color online) Simulated experiment Gaussian radiator ($\alpha=2$). Normalized sound pressure in the far-field using Eq. (27) and $ka=8$ (p_{calc}, solid curve). Recovered normalized far-field pressure data (p_{rec}, dotted curve).

solution of least-squares problems is discussed in Ref. 19. In practical cases the number of required Zernike coefficients will be less than, say, six. This will not cause numerical difficulties. Furthermore, such a modest number of coefficients already allow a large set of velocity profiles.

B. Far-field assessment from on-axis measurements

In Ref. 17 a method is described to assess low-frequency loudspeaker performance in the on-axis far-field from an on-axis near-field measurement. In the case of the rigid piston, the on-axis pressure $p(\underline{r})=p((0,0,r))$ is given by Eq. (20). Now assume $ka \ll 1$. When $r \ll a$ it holds that

$$\sin\left(\tfrac{1}{2}k\sqrt{r^2+a^2}-r\right) \approx \sin\left(\tfrac{1}{2}ka\right) \approx \tfrac{1}{2}ka, \tag{37}$$

and, when $r \gg a$ it holds that

$$\sin\frac{1}{2}k(\sqrt{r^2+a^2}-r) \approx \sin\left(\frac{ka^2}{4r}\right) \approx \frac{ka^2}{4r}. \tag{38}$$

Therefore, the ratio of the moduli of near-field and far-field on-axis pressures is given by $2r/a$. This is the basis of Keele's method;[17] it allows far-field loudspeaker assessment without having to use an anechoic room.

With the inversion procedure to estimate velocity profiles from on-axis data (which are taken in the relative near-field) as described in Sec. V A together with the forward calculation scheme for the far-field as described in Sec. IV B, it is now possible to generalize Keele's scheme.[17] This is illustrated by comparing the far-field responses pertaining to the two sets of Zernike coefficients occurring in the Gaussian simulated experiment, see Fig. 7 in Sec. V A. Using Eq. (27) the normalized far-field pressure is plotted in Fig. 10 as p_{calc} (solid curve) and p_{rec} (dotted) curves, respectively ($\alpha=2$ and $ka=8$), where the normalization is such that the factor in front of the series at the RHS of Eq. (27) equals unity. It appears that the two curves are very similar. This confirms that the u_n obtained from the noisy near-field measured pressure data yield a good estimate of the far-field spatial pressure response.

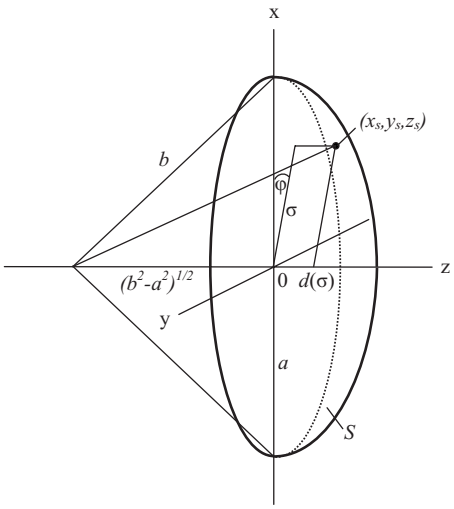


FIG. 11. Geometry of a (convex) dome-shaped radiator, with radiating surface S given as the set of points $(x_s,y_s,z_s)=(\sigma\cos\varphi,\sigma\sin c,d(\sigma))$ in accordance with Eqs. (39) and (40).

VI. GENERALIZATION OF THE ON-AXIS RESULT TO DOME-SHAPED RADIATORS

The on-axis result in Eq. (17) for the flat-piston radiators is now generalized to dome-shaped and inverted-dome-shaped radiators. The formal treatment of the dome-shaped radiators requires solving the Helmholtz equation. However, in this paper, a geometrical approach is used, where the Rayleigh integral in Eq. (1) is maintained, with distance function r' defined in accordance with the function d describing the dome surface. This geometrical approach is similar to the one used by Bordoni[20] and Kates.[21] From Suzuki and Tichy[22] (Sec. 2 A) it appears that the geometrical approach yields a fairly good approximation, at least, in the high-frequency region.

Thus the considered radiating surface S is now assumed to consist of the points

$$\underline{r}_s = (x_x,y_s,z_s) = (\sigma\cos\varphi,\sigma\sin\varphi,d(\sigma)),$$

$$0 \le \sigma \le a, \quad 0 \le \varphi \le 2\pi, \tag{39}$$

in which $d(\sigma)$ is the function

$$d(\sigma) = \pm[(b^2-\sigma^2)^{1/2}-(b^2-a^2)^{1/2}], \quad 0 \le \sigma \le a, \tag{40}$$

with $b \ge a$. The $+$-sign in Eq. (40) corresponds to a convex dome while the $-$-sign corresponds to a concave (or inverted) dome. In the $+$-case, S is the cap of a sphere of radius b bounded by the parallel of latitude of angle arccos (a/b). In Fig. 11 the situation is depicted for a convex dome. Thus the dome surface S consists of a spherical cap whose boundary coincides with the boundary of the infinite baffle.

The integral expression for the pressure is still given by the Rayleigh integral in Eq. (1), with $v(\underline{r}_s)$ the normal component of the velocity distribution on S. The normal on the surface S is given by

$$\underline{n}(\sigma) = \frac{1}{b}(\pm\sigma\cos\varphi,\ \pm\sigma\sin\varphi,(b^2-\sigma^2)^{1/2}). \tag{41}$$

The function r', being the distance between a field point $\underline{r} = (r\sin\theta, 0, r\cos\theta)$ and the point $\underline{r}_s$ in Eq. (39), is given by

$$r' = (r^2+\sigma^2-2r\sigma\sin\theta\cos\varphi - 2rd(\sigma)\cos\theta + d^2(\sigma))^{1/2}, \tag{42}$$

and is to be considered in the set (r,θ) with $r\cos\theta \geqslant d(\sigma)$.

When transforming the Rayleigh integral (1) from an integral over the disk $\sigma \leqslant a$, one should account for the surface element in the usual manner according to

$$dS = \sqrt{1+\left(\frac{\partial f}{\partial x}\right)^2+\left(\frac{\partial f}{\partial y}\right)^2}\,dxdy = \frac{b\sigma d\sigma d\varphi}{\sqrt{b^2-\sigma^2}}, \tag{43}$$

where $f(x,y) = d(\sqrt{x^2+y^2})$. Then writing

$$v(\sigma) = v((\sigma\cos\varphi, \sigma\sin\varphi, d(\sigma))), \quad 0 \leq \sigma \leq a, \tag{44}$$

one obtains

$$p(\underline{r}) = \frac{i\rho_0 ck}{2\pi}\int_0^a\int_0^{2\pi} v(\sigma)\frac{e^{-ikr'}}{r'}\frac{b}{\sqrt{b^2-\sigma^2}}\sigma d\sigma d\varphi. \tag{45}$$

In Appendix C the following is shown. Let

$$g(\tau) = \tau\left(\frac{2-(1-c_0)\tau^2}{1+c_0}\right)^{1/2}, \tag{46}$$

in which $c_0 = \sqrt{1-(a/b)^2}$. Then g is a mapping from $[0, 1]$ onto $[0, 1]$. Now when

$$W(\tau) = v(ag(\tau)), \quad W(\tau) = W_s\sum_{n=0}^{\infty} w_n R_{2n}^0(\tau), \tag{47}$$

with W_s such that $w_0 = 1$, then

$$p(\underline{r}) = i\rho_0 cW_s(ka)^2\frac{1}{1+c_0}\sum_{n=0}^{\infty}\delta_n(k,r)w_n, \tag{48}$$

where

$$\delta(k,r) = (-1)^n j_n\left(k\left(r_- + \tfrac{1}{2}d(0)\right)\right)h_n^{(2)}\left(k\left(r_+ + \tfrac{1}{2}d(0)\right)\right). \tag{49}$$

Here $r_\pm$ is as in Eqs. (17) and (18). The validity of this result is largely due to the special form of the assumed dome shape and does not seem to admit a generalization to other dome shapes.

A discussion of the result of Eqs. (47)–(49) follows now. As Eq. (47) shows, a simple warping operation on v is required so as to obtain the series representation (48) with coefficients per Eq. (49) that is strikingly similar to the result of Eqs. (17) and (18). In Fig. 12 the warping function $g(\tau)$, $0 \leqslant \tau \leqslant 1$, for $s_0 = a/b = 1, 0.9, 0.5, 0$ is shown, the case $s_0 = 0$ yielding $g(\tau) = \tau$ [i.e., the result of Eqs. (17) and (18)]. The plots show that for moderate values of s_0, say, $s_0 \leqslant 1/2$, the influence of the warping operation is quite modest. Consequently, in that case, $w_n \approx u_n$, where u_n are the coefficients in the Zernike expansion (3) of v itself.

A particularly interesting case occurs when $b = a$ so that S is a hemisphere. Then $c_0 = 0$ and

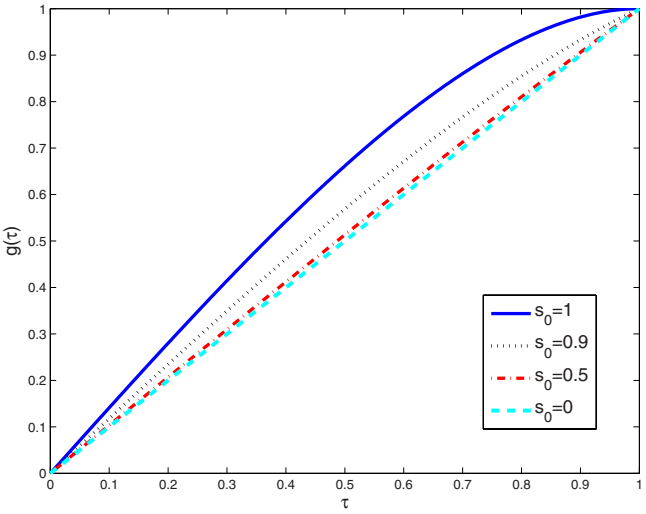


FIG. 12. (Color online) The warping function [Eq. (46)] $g(\tau)$ vs τ.

$$g(\tau) = \tau(2-\tau^2)^{1/2} = (1-(1-\tau^2)^2)^{1/2}. \tag{50}$$

When $v(\sigma) = v^{(\ell)}(\sigma) = (\ell+1)V_s(1+(\sigma/a)^2)^\ell$, see Eq. (6), one gets

$$v^{(\ell)}(ag(\tau)) = (\ell+1)V_s(1-\tau^2)^{2\ell} = \frac{\ell+1}{2\ell+1}v^{(2\ell)}(\tau), \tag{51}$$

and so the required Zernike coefficients w_n in Eq. (47) are available per Appendix A 1, case $k = 2\ell$. For the Gaussian in Eq. (7) one computes, using Eq. (50) and $\exp(x) = \Sigma_{\ell=0}^{\infty} x^\ell/l!$,

$$v(ag(\tau);\alpha) = \frac{\alpha V_s}{e^\alpha - 1}\sum_{\ell=0}^{\infty}\frac{\alpha^\ell}{\ell!}(1-\tau^2)^{2\ell}. \tag{52}$$

Hence the required Zernike coefficients w_n in Eq. (47) can be obtained explicitly in the form of infinite series by combining Eq. (7) and Appendix A 1.

VII. DISCUSSION AND OUTLOOK

In this paper the foundation is laid for a method to perform forward and inverse sound pressure computations for circular radiators with a non-uniform velocity profile. In the forward problem, the velocity profile is assumed to be known and the on-axis and far-field sound pressures are expressed analytically in terms of Zernike expansion coefficients of the velocity profile and (spherical) Bessel (and Hankel) functions. In the inverse problem, the velocity profile is unknown and is estimated in terms of Zernike expansion coefficients from on-axis pressure data by adopting a matching approach based on the analytic result for the on-axis pressure. Well-behaved velocity profiles are already adequately represented by only a few terms of their Zernike expansion. Therefore, the Zernike series approach is more convenient for both the forward problems and the inverse problem than, for instance, an approach based on expansions involving the family of rigid, simply supported, and clamped radiators. The forward and inverse methods are proposed for use in assessment of the far-field of a loudspeaker without the need for an anechoic room. Here, the Zernike coefficients of the velocity profile are estimated from the on-axis (rela-

tively near-field) data, and these coefficients are used in the forward scheme to compute the far-field. This assessment procedure has not been fully worked out in the present paper due to a variety of practical issues that need to be addressed. Among these practical issues are

- choice of the on-axis measurement points,
- condition of the linear systems that arise,
- influence of ka,
- influence of noise,
- influence of misalignment of the measurement points,
- influence of inclination of the measurement axis, and
- incorrect setting of the radius of the radiator,

while various combinations of these issues should also be considered. The authors intend to work out the method for the loudspeaker assessment with attention for the above mentioned points.

In this paper, the theory has been developed mainly for flat radiators. However, the basic result for the on-axis pressure as a series expansion, in terms of Zernike coefficients and spherical Bessel and Hankel functions, has been generalized to the case of dome-shaped radiators. It is, therefore, to be expected that both the forward and the inverse methods can also be generalized to the case of dome-shaped radiators. The forward problem and the inverse problem concerning the on-axis pressure are not much more difficult than in the case of flat radiators: Only a simple warping operation performed on the velocity profile is required. The computation of the far-field is, however, more complicated for the dome-shaped radiators. In the first place, the estimated Zernike coefficients require a transformation due to the warping operation. Secondly, the basic integrals corresponding to a single Zernike term, which occur in the far-field expression for the sound pressure, are not readily expressed in Bessel functions as was the case for flat radiators. The authors intend to develop the method both on a theoretical and practical level for dome-shaped radiators as well.

VIII. CONCLUSIONS

Zernike polynomials are an efficient and robust method to describe velocity profiles of resilient sound radiators. Only a few coefficients are necessary to approximate a wide variety of velocity profiles including the rigid piston, the simply supported radiator, the clamped radiators, Gaussian radiators, as well as real loudspeaker drivers. This method enables one to solve the inverse problem of calculating the actual velocity profile of the radiator using (measured) on-axis sound pressure data. This computed velocity profile allows the extrapolation to far-field loudspeaker pressure data, including off-axis behavior. The forward computation scheme can be extended to dome-shaped radiators with arbitrary velocity distributions.

APPENDIX A: ZERNIKE EXPANSIONS

In this appendix a number of explicit Zernike expansions for the radially symmetric case are collected, some general results on Zernike expansions are given, and some examples are given to demonstrate that the nth coefficient in a Zernike expansion is roughly 4^n as small as the nth coefficient in an expansion based on the functions $(1-(\sigma/a)^2)^n$ in Eq. (6). The results are presented in terms of the normalized variable $\rho=\sigma/a$, $0\leqslant\rho\leqslant 1$.

1. Zernike expansion of $\rho^{2\ell}$, $(1-\rho^2)^\ell$

From Ref. 23, Appendix A,

$$\rho^{2\ell}=\sum_{n=0}^{\ell}\frac{2n+1}{n+1}\frac{\binom{\ell}{n}}{\binom{\ell+n+1}{\ell}}R_{2n}^0(\rho). \tag{A1}$$

Using $P_n(-x)=(-1)^nP_n(x)$, so that $R_{2n}^0((1-\rho)^2)^{1/2})=(-1)^nR_{2n}^0(\rho)$, one obtains Eq. (10).

2. Expansion of $e^{-\alpha\rho^2}$

It holds that

$$e^{-\alpha\rho^2}=e^{-(1/2)\alpha}\sum_{n=0}^{\infty}(-1)^n(2n+1)\sqrt{\frac{\pi}{\alpha}}I_{n+1/2}(\alpha/2)R_{2n}^0(\rho), \tag{A2}$$

where $(\pi/2z)^{1/2}I_{n+1/2}(z)$ is the modified spherical Bessel function of order n. See Ref. 14, Secs. 10.2 and 10.2.37, with $z=\alpha/2$ and $\cos\theta=2\rho^2-1$ in Sec. 10.2.37. From Ref. 14, Secs. 22.14.7 and 10.25,

$$|R_{2n}^0(\rho)|\leq 1,\quad \sqrt{\frac{\pi}{\alpha}}I_{n+1/2}(\alpha/2)\approx\frac{(\alpha/2)^n}{1\times 3\cdots(2n+1)},\quad n\to\infty. \tag{A3}$$

For instance, when $\alpha=1$, including the terms in the series in Eq. (A2) with $n=0,\ldots,5$ yields absolute accuracy 10^{-6} on the whole range $0\leqslant\rho\leqslant 1$. Alternatively, there is an expansion in terms of the velocity profiles $(1-(\sigma/a)^2)^n$ of Eq. (6), viz.,

$$e^{-\alpha\rho^2}=e^{-\alpha}\sum_{n=0}^{\infty}\frac{\alpha^n}{n!}(1-\rho^2)^n. \tag{A4}$$

Denoting the expansion coefficients in Eq. (A2) by u_n and those in Eq. (A4) by w_n, it follows from the asymptotics of the Γ-function that

$$\frac{u_n}{w_n}\approx e^{(1/2)\alpha}\frac{\sqrt{\pi n}}{4^n},\quad n\to\infty. \tag{A5}$$

Thus the nth coefficient required in the clamped radiator expansion in Eq. (A4) is roughly 4^n as large as the one needed in the Zernike expansion in Eq. (A2). This phenomenon has been noted and verified for a large set of other analytic velocity profiles such as (spherical) Bessel functions.

3. Expansion of $\exp(-i\lambda(T^2+x\rho^2)^{1/2})/-i\lambda(T^2+x\rho^2)^{1/2}$

To show Eq. (12), consider Ref. 14, Secs. 10.1.45 and 10.1.46,

$$\frac{e^{i\lambda V}}{i\lambda V} = \sum_{n=0}^{\infty} (2n+1) j_n(\lambda r_1) h_n^{(1)}(\lambda r_2) P_n(\cos\gamma), \tag{A6}$$

where $V=(r_1^2+r_2^2-2r_1r_2\cos\gamma)^{1/2}$ with $\gamma\in\mathbb{R}$, and $0\leqslant|r_1|\leqslant r_2$. With the choice

$$r_1 = T_-, \quad r_2 = T_+, \quad \cos\gamma = 1-2\rho^2, \tag{A7}$$

one has $V=(T^2+x\rho^2)^{1/2}$ and Eq. (12) follows using $P_n(-x)=(-1)^nP_n(x)$, $R_{2n}^0(\rho)=P_n(2\rho^2-1)$ and Ref. 14, Secs. 10.1.1, 10.1.3,

$$j_n(-z) = (-1)^n j_n(z), \quad h_n^{(1)}(-z) = (-1)^n h_n^{(2)}(z). \tag{A8}$$

APPENDIX B: FAR-FIELD APPROXIMATION

Letting $\underline{r}=(r\sin\theta,0,r\cos\theta)$, it holds

$$p(\underline{r}) = \frac{i\rho_0 ck}{2\pi}\int_S v(\underline{r}_s)\frac{e^{-ikr'}}{r'}dS = i\rho_0 ck\int_0^a v(\sigma) \times\left(\frac{1}{2\pi}\int_0^{2\pi}\frac{e^{-ik\sqrt{r^2+\sigma^2-2r\sigma\sin\theta\cos\varphi}}}{\sqrt{r^2+\sigma^2-2r\sigma\sin\theta\cos\varphi}}d\varphi\right)d\sigma. \tag{B1}$$

Now, when $r\to\infty$,

$$\sqrt{r^2+\sigma^2-2r\sigma\sin\theta\cos\varphi} \approx r-\sigma\sin\theta\cos\varphi \approx r. \tag{B2}$$

Replacing r' in $\exp(-ikr')$ with $r-\sigma\sin\theta\cos\varphi$ and r' in the denominator in the integrand in Eq. (B1) with r while using

$$\int_0^{2\pi} e^{iz\cos\varphi}d\varphi = 2\pi J_0(z), \tag{B3}$$

the conventional approximation

$$p(\underline{r}) \approx i\rho_0 ck\frac{e^{-ikr}}{r}\int_0^a v(\sigma)J_0(k\sigma\sin\theta)\sigma d\sigma \tag{B4}$$

follows. Then Eq. (27) follows upon inserting in Eq. (B4) the Zernike expansion (3) of v and using the result of Eq. (14).

APPENDIX C: ON-AXIS FIELD FOR A DOME-SHAPED RADIATOR

In this appendix the result of Eqs. (46)–(49) is proved. The on-axis pressure at $\underline{r}=(0,0,r)$ is given by, see Eq. (45),

$$p(\underline{r}) = i\rho_0 ck\int_0^a \frac{bv(\sigma)}{\sqrt{b^2-\sigma^2}}\frac{e^{-ikr'}}{r'}\sigma d\sigma, \tag{C1}$$

in which

$$r' = (r^2+\sigma^2-2rd(\sigma)+d^2(\sigma))^{1/2} \tag{C2}$$

and $d(\sigma)$ is given by Eq. (40). Working this out, while noting cancellation of terms σ^2 due to the special form of d, it follows that

$$(r')^2 = c(r)+d(r)\psi(\sigma), \quad \psi(\sigma) = -(b^2-\sigma^2)^{1/2}, \tag{C3}$$

where $c(r)$ and $d(r)$ only depend on r. In normalized coordinates

$$\rho=\sigma/a, \quad s_0=a/b, \tag{C4}$$

Eq. (C3) assumes the form

$$(r'/a)^2 = C(r)+D(r)\Psi(\rho) =: F(\rho;r), \tag{C5}$$

where $C(r)=a^{-2}c(r)$, $D(r)=a^{-2}d(r)$, and

$$\Psi(\rho) = \frac{1}{b}\psi(\sigma) = -(1-s_0^2\rho^2)^{1/2}, \quad 0\leq\rho\leq 1. \tag{C6}$$

This $F(\rho;r)$ can be written as

$$F(\rho;r) = F(0;r)+(F(1;r)-F(0;r))f^2(\rho), \tag{C7}$$

where

$$f(\rho) = \sqrt{\frac{\Psi(\rho)-\Psi(0)}{\Psi(1)-\Psi(0)}} = \sqrt{\frac{1-(1-s_0^2\rho^2)^{1/2}}{1-c_0}} \tag{C8}$$

and $c_0=(1-s_0^2)^{1/2}$. Letting

$$T^2 = F(0;r), \quad x = F(1;r)-F(0;r), \tag{C9}$$

it follows that $r'=a(T^2+xf^2(\rho))^{1/2}$. Hence, from Eqs. (C1) and (12) with $\lambda=ka$ and $f(\rho)$ instead of ρ,

$$p(\underline{r}) = i\rho_0 cka\int_0^1 \frac{v(a\rho)}{\sqrt{1-s_0^2\rho^2}}\frac{e^{-ikr'}}{r'}\rho d\rho = \rho_0 c(ka)^2\sum_{n=0}^{\infty}\delta_n(k,r)t_n, \tag{C10}$$

in which, see below Eq. (12),

$$\delta_n(k,r) = (-1)^n j_n(kaT_-)h_n^{(2)}(kaT_+) \tag{C11}$$

and

$$t_n = (2n+1)\int_0^1 \frac{v(a\rho)}{\sqrt{1-s_0^2\rho^2}}R_{2n}^0(f(\rho))\rho d\rho. \tag{C12}$$

From Eqs. (C5) and (C9) it follows that

$$aT = r-d(0), \quad a\sqrt{T^2+x} = \sqrt{r^2+a^2}; \tag{C13}$$

hence, $\delta_n(k,r)$ is given by the RHS of Eq. (49). Next, substitute $\tau=f(\rho)$, $\rho=g(\tau)$ in the integral defining t_n, where the inverse function g of f is given by

$$g(\tau) = \tau\left(\frac{2-(1-c_0)\tau^2}{1+c_0}\right)^{1/2}, \quad 0\leq\tau\leq 1. \tag{C14}$$

Then, using

$$\frac{g'(\tau)g(\tau)}{\sqrt{1-s_0^2g^2(\tau)}} = \frac{2}{1+c_0}, \tag{C15}$$

it follows that

$$t_n = \frac{2(2n+1)}{1+c_0}\int_0^1 v(ag(\tau))R_{2n}^0(\rho)\rho d\rho. \tag{C16}$$

Noting that $W(\tau)=v(ag(\tau))$, see Eq. (47), the proof is complete.

[1]G. R. Harris, "Review of transient field theory for a baffled planar piston," J. Acoust. Soc. Am. **70**, 10–20 (1981).
[2]M. Greenspan, "Piston radiator: Some extensions of the theory," J. Acoust. Soc. Am. **65**, 608–621 (1979).
[3]T. D. Mast and F. Yu, "Simplified expansions for radiation from a baffled circular piston," J. Acoust. Soc. Am. **118**, 3457–3464 (2005).
[4]T. B. Hansen, "Probe-corrected near-field measurements on a truncated cylinder," J. Acoust. Soc. Am. **119**, 792–807 (2006).
[5]T. J. Mellow, "On the sound field of a resilient disk in an infinite baffle," J. Acoust. Soc. Am. **120**, 90–101 (2006).
[6]J. F. Kelly and R. J. McGough, "An annular superposition integral for axisymmetric radiators," J. Acoust. Soc. Am. **121**, 759–765 (2007).
[7]B. R. A. Nijboer, "The diffraction theory of aberrations," Ph.D. dissertation, University of Groningen, The Netherlands, 1942.
[8]F. Zernike and B. R. A. Nijboer, "The Theory of Optical Images (La théorie des images optiques), Revue d'Optique, Paris, 227–235 (1949).
[9]M. Born and E. Wolf, *Principles of Optics*, 7th ed. (Cambridge University Press, Cambridge, 2002), Chap. 9.
[10]J. J. M. Braat, S. van Haver, A. J. E. M. Janssen, and P. Dirksen, "Assessment of optical systems by means of point-spread functions," in *Progress in Optics*, edited by E. Wolf (Elsevier, Amsterdam, 2008), Vol. **51**, Chap. 6.
[11]L. E. Kinsler, A. R. Frey, A. B. Coppens, and J. V. Sanders, *Fundamentals of Acoustics* (Wiley, New York, 1982).
[12]J. W. S. Rayleigh, *The Theory of Sound* (The Macmillan Co., London, 1896), Vol. **2**; *ibid.* (Dover, New York, 1945).
[13]F. J. M. Frankort, "Vibration and sound radiation of loudspeaker cones," Ph.D. dissertation, Delft University of Technology, Delft, The Netherlands, 1975.
[14]M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1972).
[15]H. Stenzel, "On the acoustical radiation of membranes (Über die akustische strahlung von membranen)," Ann. Phys. **7**, 947–982 (1930).
[16]D. L. Dekker, R. L. Piziali, and E. Dong, "Effect of boundary conditions on the ultrasonic-beam characteristics of circular disks," J. Acoust. Soc. Am. **56**, 87–93 (1974).
[17]D. B. (Don) Keele, Jr., "Low-frequency loudspeaker assessment by nearfield sound-pressure measurement," J. Audio Eng. Soc. **22**, 154–162 (1974).
[18]International Electrotechnical Commission (IEC), Geneva, Switzerland, IEC 60268-5 Sound System Equipment—Part 5: Loudspeakers, 2007.
[19]G. H. Golub and C. F. Van Loan, *Matrix Computations*, 2nd ed. (The Johns Hopkins University Press, Baltimore, MD, 1989).
[20]P. G. Bordoni, "The conical sound source," J. Acoust. Soc. Am. **17**, 123–126 (1945).
[21]J. M. Kates, "Radiation from a dome," J. Audio Eng. Soc. **24**, 735–737 (1976).
[22]H. Suzuki and J. Tichy, "Sound radiation from convex and concave domes in an infinite baffle," J. Acoust. Soc. Am. **69**, 41–49 (1981).
[23]J. J. M. Braat, P. Dirksen, A. J. E. M. Janssen, and A. S. van de Nes, "Extended Nijboer-Zernike representation of the vector field in the focal region of an aberrated high-aperture optical system," J. Opt. Soc. Am. A **20**, 2281–2292 (2003).

Effect of electrical circuits on duration of an acoustic pulse radiated by a piezoplate

S. I. Konovalov and A. G. Kuz'menko
St. Petersburg State Electrotechnical University, Professor Popov Street 5, St. Petersburg 197376, Russia

(Received 11 February 2008; revised 3 November 2008; accepted 19 December 2008)

By means of numerical computation, pulse mode of a piezoelectric radiator in the form of a plate, loaded with a liquid and connected to *R-L* circuits in various combinations, is investigated. The computations were carried out having taken into account the internal resistance of the electric generator. Numerical results, based on the theory, are provided for examples of parallel and series circuits connected to a piezoelectric plate. The optimal values of parameters are determined, providing minimal duration of radiated acoustic pulses.
 [DOI: 10.1121/1.3075582]

PACS number(s): 43.38.Yn [AJZ] Pages: 1456–1460

I. INTRODUCTION

There are many problems in applied acoustics, which require a short pulse, duration of which is about some cycles (2–3) of high-frequency oscillations, because this allows to enhance resolution and to decrease the length of the dead space (for instance, in ultrasonic nondestructive testing and medical diagnostic apparatuses).

One of the methods for obtaining a short acoustic pulse at the output of the transducer is to connect an electric *R-L* circuit in different combinations to the radiating element. However, the problem of choice of optimal parameters of electric circuits, which provide minimal duration of radiated pulses, was not elucidated duly in scientific literature.

In the papers[1–5] the authors investigated pulse mode of a piezoelectric plate connected in series to *R-L* circuits, which elements *R* and *L* may be connected both in series and in parallel to each other. The electric generator exciting the transducer was assumed to be an ideal source of electric voltage, i.e., its internal resistance is equal to zero.

II. ANALYTICAL CONSIDERATION

This paper relates to pulse mode of a piezoelectric plate for the case when the generator has the internal resistance that is not equal to zero. It should be noted that in this case the corrective *R-L* circuit may be connected to the piezoelectric plate not only in series but in parallel as well, and the elements within the *R-L* circuit may be connected in series or in parallel to each other.

Figure 1(a) illustrates two methods for connecting electric load Z_{el} to the piezoelectric plate *P*. Herein the following notations are used: $U(t)$ is electric voltage, produced by a generator for exciting of the acoustic radiator; R_i is internal resistance of the generator; S_1 and S_2 are switches designed for connecting electric load either in parallel (position 1) or in series (position 2) to the piezoelectric plate.

In Fig. 1(b), two methods for realizing the electric load Z_{el} are shown (series or paralleling connection *R* and *L*).

The problem is to determine parameters of electric circuits, when R_i is given or being selected, so as to achieve at the output of the transducer the acoustic pulse of minimal duration. These values of parameters will be referred to as optimal values.

The acoustic load of the transducer will be water, at one side and air at the other side.

Pulse duration will be defined as the time interval from the beginning of the pulse until the time of decreasing pulse level by 20 dB from its relative maximum value of particle velocity during the pulse.

To resolve the above-mentioned problem, we shall use the method of analogous electric circuits and also the method based on the use of Fourier transform pair.

Having used wide known analogous electric circuit for a piezoelectric plate,[6] we obtain the equivalent scheme for the radiating system, presented in Fig. 2.

In this figure, Z_{el} is connected in parallel to the transducer, and elements *R* and *L* are connected in series one to another. In this case switches S_1 and S_2 in Fig. 1 are in position 1. The mechanical impedance loading the piezoelectric plate is designated as Z_l.

In the Fig. 2, the impedances are correspondingly

$$Z_1 = j(\rho c)_{\text{cer}} S \tan\frac{x_{\text{cer}}}{2}, \quad Z_2 = \frac{(\rho c)_{\text{cer}} S}{j \sin x_{\text{cer}}}, \quad x_{\text{cer}} = \frac{\omega}{c} l,$$

index "cer" relates to piezoelectric ceramics, x_{cer} is wave thickness of the active element, ω is current angular frequency, l is thickness of the plate, $(\rho c)_{\text{cer}}$ is specific acoustic impedance of piezoceramics, $c = (c_{33}^D/\rho)^{1/2}$ is the phase velocity of the longitudinal wave in the clamped plate, C_0 is the electrical capacity of the clamped plate, and S is the area of the plate.

The total mechanical impedance measured at the ends of the secondary winding of the electromechanical transformer we shall refer to as Z_M. This is impedance of a part of circuit connected to the secondary winding of the electromechanical transformer. It may be expressed through impedances Z_1, Z_2, and Z_l by means of conventional methods known in the theory of electrical circuits. This impedance may be transferred into the part of the circuit including the primary winding, according to the rules known for transformers: $Z_{\text{ins}} = Z_M/K_U^2$, where K_U is coefficient of electromechanical trans-

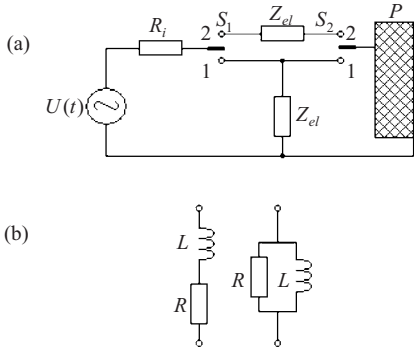


FIG. 1. The schematic model of the assigned task (a) and the types of electric load Z_{el} (b).

formation. In the theory of circuits impedance, Z_{ins} is referred to as "reflected secondary impedance."

In Fig. 2, the elements R and L are connected in series. Similarly the case of parallel connection may be considered.

This scheme may be made simpler, if two parallel branches R-L and C_0 will be replaced by their resulting impedance Z. Then the scheme may be calculated by means of well known in electrical engineering conventional methods.

The current flowing through the inserted impedance Z_{ins} may be determined as

$$I = \frac{UZ\omega C_0}{[R_i\omega C_0 + Z\omega C_0]\dfrac{Z_M}{K_U^2} + Z(\omega C_0 R_i + j) + jR_i}.$$

The total mechanical "current" (in fact, the particle velocity) flowing through the secondary winding of the transformer is $I_M = I/K_U$.

Further, we introduce a number of new variables according to

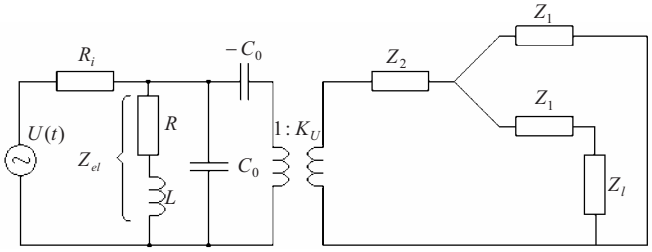


FIG. 2. The analogous circuit for a transducer.

$$LC_0 = \frac{1}{\omega_{el}^2}, \quad \frac{\omega_{el}}{\omega_0} = n, \quad \frac{\omega}{\omega_0} = \gamma, \quad \frac{\omega_{el}L}{R} = Q, \quad \frac{1}{\omega_0 C_0 R_i} = Q_1.$$

New variables (dimensionless) are introduced for obtaining the general solution of the problem, suitable for the entire class of the transducers in the form of a plate.

Now we explain the sense of these variables: ω_{el} is the resonant frequency of the oscillating circuit formed by inductance L and capacity C_0, ω_0 is the antiresonant frequency of the plate, parameter n characterizes "tuning" of the oscillating circuit (LC_0) relative ω_0, γ is dimensionless (relative) frequency, Q is the quality factor for electric load, Q_1 is the quality factor that characterizes losses of energy due to internal resistance of the electric generator.

We shall use also the expression for coefficient of electromechanical transformation: $K_U = e_{33}S/l$, where e_{33} is piezoelectric constant.

Yet one equation will be useful: $K_U^2 1/\omega_0 C_0 = (\beta^2/\pi)Z_{cer}$, where β^2 is the square of the coefficient of electromechanical coupling for the plate under longitudinal oscillations through the thickness.

Taking into account the variables introduced, we obtain

$$I_M = (K_U U)\frac{\dfrac{A(\gamma)}{B(\gamma)}}{\left(\dfrac{\gamma}{Q_1} + \dfrac{A(\gamma)}{B(\gamma)}\right)Z_M(\gamma\pi) + \left[\left(\dfrac{\gamma}{Q_1} + j\right)\dfrac{A(\gamma)}{B(\gamma)} + j\dfrac{\gamma}{Q_1}\right]Z_{cer}\dfrac{\beta^2}{\gamma\pi}}.$$

For the case of R and L connected in series,

$$A(\gamma) = \frac{\gamma}{nQ} + j\frac{\gamma^2}{n^2}, \quad B(\gamma) = jA(\gamma) + 1.$$

For the case of R and L connected in parallel,

$$A(\gamma) = j\frac{\gamma^2}{n^2 Q}, \quad B(\gamma) = j\frac{\gamma}{n} + \frac{1}{Q} + jA(\gamma).$$

The formula for I_M comprises the value of Z_M. It is not very difficult to find the expression for it:

$$Z_M = Z_{cer}\frac{1 + \cos x_{cer}}{\sin x_{cer}}\frac{Z_{cer}\sin x_{cer} - jZ_l\cos x_{cer}}{(1 + \cos x_{cer})Z_l + j2Z_{cer}\sin x_{cer}}.$$

The current I_l flowing through the acoustic load Z_l may be determined through the current I_M, using the transfer coefficient: $I_l = I_M K(\gamma)$, where for coefficient $K(\gamma)$ the expression can be obtained:

$$K(\gamma) = \frac{\sin(\gamma\pi)}{2\sin(\gamma\pi) - j\dfrac{Z_l}{Z_{cer}}(1 + \cos(\gamma\pi))}.$$

Similarly the case may be considered, when Z_{el} is connected in series to the piezoelectric plate, the switches S_1 and S_2 being this time in position 2.

As well as in the preceding case, the expressions for the currents I and $I_M = I/K_U$ can be obtained.

For the case of Z_{el} in series,

$$I_M = \frac{UK_U}{\left[j\gamma \frac{1}{n}\left(\frac{1}{Q} + \frac{n}{Q_1}\right) - \frac{\gamma^2}{n^2} + 1 \right] Z_M + j\beta^2 \frac{Z_{cer}}{\gamma\pi}}.$$

For the case of Z_{el} in parallel,

$$I_M = \frac{UK_U}{\left[j\gamma \frac{1}{Q_1} + \frac{1}{Q^2 + \frac{n^2}{\gamma^2}}\left(-1 + j\frac{\gamma Q}{n}\right) + 1 \right] Z_M + j\beta^2 \frac{Z_{cer}}{\gamma\pi}}.$$

Further the computation of I_l, i.e., vibrational speed of the radiating surface of the piezoelectric plate can be accomplished analogously to the preceding case.

Having obtained the frequency response $I_l(\gamma\pi)$, we can turn to computation of the pulse mode.

The dependence of the vibrational speed on time in acoustic pulse is determined by the inverse Fourier transform:

$$v(T) = \frac{\omega_0}{2\pi} \int_{-\infty}^{\infty} U(\gamma) I_l(\gamma\pi) e^{j\gamma\pi T} d\gamma.$$

Here $U(\gamma)$ is the spectral function of electric pulse exciting the piezoplate; $I_l(\gamma\pi)$ is the spectral function of vibrational speed; $T = t/T_0/2$ is dimensionless time; T_0 is period of natural oscillations of the plate at the angular frequency ω_0.

Taking into account that values of $v(T)$ are to be real and the property of symmetry of the integral at the intervals $(-\infty, 0)$ and $(0, \infty)$, the formula for $v(T)$ can be written in the form

$$v(T) = \left(\frac{\omega_0}{\pi}\right) \mathrm{Re} \int_0^{\infty} U(\gamma) I_l(\gamma\pi) e^{j\gamma\pi T} d\gamma.$$

Let the exciting pulse be one-half period of the sinusoid at the angular frequency ω_0 that corresponds to the resonance exciting. The spectral function of this pulse may be written as

$$U(\gamma) = \frac{\cos\left(\frac{\gamma\pi}{2}\right)}{1 - \gamma^2} e^{-j\gamma\pi/2}.$$

Now, if the expression for $U(\gamma)$ and $I_l(\gamma\pi)$ would be substituted in the formula for $v(T)$, one can determine the form of pulse of vibrational speed.

Piezoelectric material used in the paper is of PZT-type. Its parameters used for calculations have the following values: $\beta^2 = 0.23$; $Z_{cer} = 30 \times 10^6$ Pa s/m; $c = 4100$ m/s; for acoustical load (water) $Z_l = 1.5 \times 10^6$ Pa s/m.

The results of calculations are considered and discussed below.

III. THE NUMERICAL RESULTS

Connection Z_{el} in parallel to the piezoelectric plate under condition that switches S_1 and S_2 in Fig. 1 are in position 1.

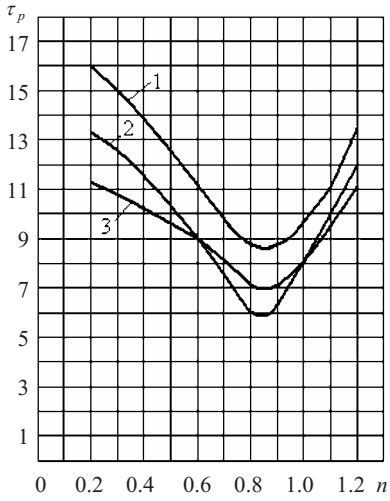


FIG. 3. The dependences τ_p versus n for constant value Q and different values Q_1 under R-L connections being in series. $1-Q_1=0.4$, $2-Q_1=0.6$, and $3-Q_1=0.8$.

The significant number of systematically accomplished calculations allowed determining the approximate areas for values of parameters n, Q_1, and Q, which provide the noticeable decrease in acoustic pulses durations. In particular, the computation has shown that when the elements R and L are connected to one another in series, the positive effect may be found, if $n=0.7,\ldots,1.0$, $Q_1=0.4,\ldots,0.8$, and $Q \approx 10$. These notations mean that n may take any value from 0.7 to 1.0, respectively, and Q_1 may take any value from 0.4 to 0.8.

When the connection is a parallel one, $n=0.7,\ldots,1.0$, $Q_1=0.2,\ldots,0.7$, and $Q=0.15,\ldots,0.80$.

The process of searching more suitable values of parameters n, Q, and Q_1 is presented in Fig. 3 (for connection R and L in series) and in Fig. 4 (for connection R and L in parallel). In both figures along the axes of abscissas the values of relative resonance frequency n of the oscillating LC_0 circuit are laid. The ordinate axes contain the values of duration τ_p of radiated acoustic pulse, presented by the number of half-period at the natural frequency of the piezoelectric plate.

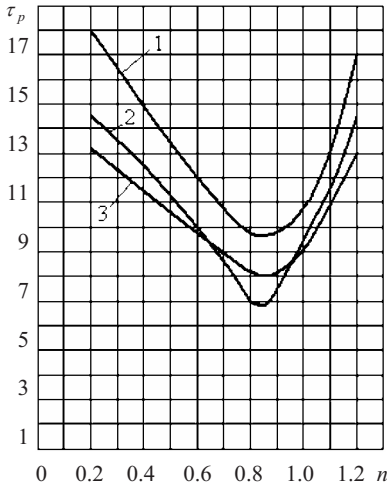


FIG. 4. The dependences analogous to those in Fig. 3, but for R-L connections being in parallel. $1-Q_1=0.3$, $2-Q_1=0.5$, and $3-Q_1=0.7$.

The data presented in Fig. 3 correspond to the value $Q=10$. The numbers of curves in dependence on Q_1 are $1-Q_1=0.4$, $2-Q_1=0.6$, and $3-Q_1=0.8$.

Comparing these curves we can make the conclusion that minimal duration of the pulse $\tau_p \approx 6$ is achieved for the next set of parameter values: $Q=10$, $Q_1=0.6$, and $n=0.85$. We can take these values of parameters as the optimal values.

The results of computation, presented in Fig. 4, correspond to the value $Q=0.2$. The numbers of curves in dependence on Q_1 are $1-Q_1=0.3$, $2-Q_1=0.5$, and $3-Q_1=0.7$.

It can be seen that minimal duration of the pulse is achieved for $Q=0.2$, $Q_1=0.5$, and $n=0.85$. Thus these values are close to the optimal values.

Note that in both cases considered, corresponding connection of the $R-L$ circuit, having configuration in series or in parallel, to the piezoplate in parallel, the duration of the radiated acoustic pulses, as well as their amplitudes are very close to each other.

Connection Z_{el} in series with the piezoelectric plate under condition that switches S_1 and S_2 in Fig. 1 are in position 2.

In both cases, parallel and series connections R and L, as the computation has shown, the durations of radiated pulses are $\tau_p=5.5-6$, and their amplitudes are very close.

The forms of acoustic signals, radiating by the system under optimal values of parameters for all combinations of connection $R-L$ circuit both to each other and to the piezoelectric plate, are presented in the Fig. 5. In particular, Figs. 5(a) and 5(b) correspond to the switches S_1 and S_2 in Fig. 1 being in position 1.

Figure 5(a) relates to the case of connections R and L in series, and Fig. 5(b) corresponds to connection R and L in parallel.

Two other pulses [Fig. 5(c) and 5(d)] correspond to the case, when the switches S_1 and S_2 in Fig. 1 are in position 2. The pulse presented in Fig. 5(c) relates to $R-L$ connection in series, and the pulse presented in Fig. 5(d) relates to connections R and L in parallel. For each of the combinations presented in Fig. 5(a)–5(d), the optimal values of parameters are correspondingly:

$$\text{Fig. 5(a)} - n_{\text{opt}} = 0.85, \quad Q_{\text{opt}} = 10, \quad Q_{1\text{opt}} = 0.55;$$

$$\text{Fig. 5(b)} - n_{\text{opt}} = 0.84, \quad Q_{\text{opt}} = 0.15, \quad Q_{1\text{opt}} = 0.5;$$

$$\text{Fig. 5(c)} - n_{\text{opt}} = 1.1, \quad Q_{\text{opt}} = 2.0, \quad Q_{1\text{opt}} = 6;$$

$$\text{Fig. 5(d)} - n_{\text{opt}} = 0.85, \quad Q_{\text{opt}} = 0.75, \quad Q_{1\text{opt}} = 9.$$

In Figs. 5(a)–5(d) along the axes of ordinates, the normalized values $(v/v_{\max})$ are laid, where $v_{\max}$ is the maximal absolute value of the vibrational velocity during the pulse duration.

It can be seen that for all cases considered, the acoustic pulses have almost the same form, and do not exceed 5.5–6 one-half periods. Hence, there exist approximately equal opportunities for obtaining pulses of short durations, when using any of considered above variants of connection R-L circuits to the piezoelectric plate.

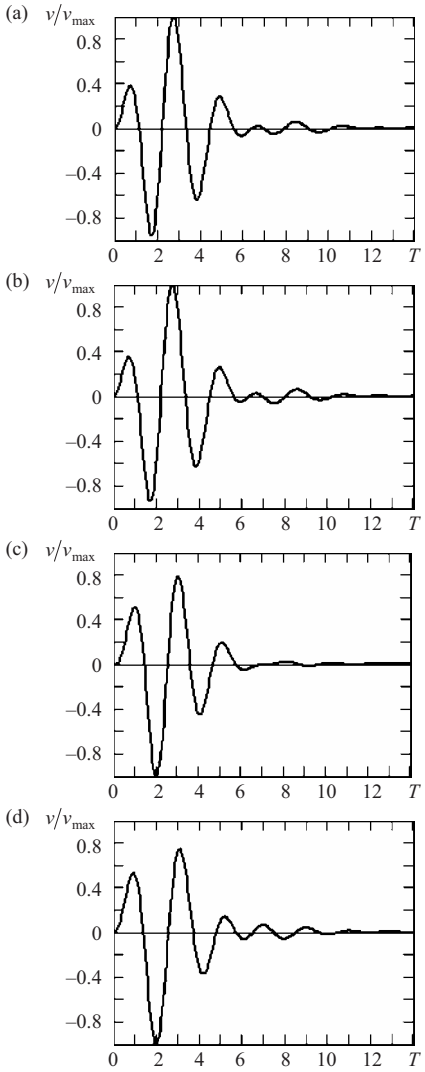


FIG. 5. The radiated acoustic pulses of vibrational velocity for optimal values of parameters Q, Q_1, and n.

The results of numerical investigation allow stating that under optimal values of parameters, the amplitudes of radiated pulses occur to have very close values, independent of series or parallel type of R-L circuit itself. This is true for both parallel and series connection the R-L circuit to the piezoelectric plate. However, connection of the electric load in parallel to the piezoelectric plate provides the amplitude of the radiated pulse about two times less than in the case of connection to the plate in series. This may be explained by the fact, that in the case of parallel connection the electric load to the piezoelectric plate, the optimal value R_i is significantly greater than in the case of connection in series. This results in decrease in the voltage transfer coefficient from the generator to the piezoelectric plate.

IV. CONCLUSION

In this paper, the possibility of radiating a short pulse by connecting R-L circuits to the piezoelectric plate in various combinations was investigated. On the base of numerical computation, the specific estimates for durations and ampli-

tudes of radiated signals were obtained, and the optimal values of parameters, characterizing the system, were determined.

ACKNOWLEDGMENT

Authors are thankful to Professor Abbakumov K. E. for his support and advice.

[1]S. I. Konovalov and A. G. Kuz'menko, "Effect of the way of connecting the electric load elements on the pulse duration generated by a piezoelectric transducer," Acoust. Phys. **53**, 245–247 (2007).
[2]S. I. Konovalov and A. G. Kuz'menko, "Effect of an electric load on the duration of the acoustic pulse radiated by a piezoelectric plate," Acoust. Phys. **50**, 112–113 (2004).
[3]S. I. Konovalov and A. G. Kuz'menko, "Damping of a piezoelectric plate and the use of an electric circuit to obtain a short acoustic pulse," Acoust. Phys. **51**, 719–721 (2005).
[4]S. I. Konovalov and A. G. Kuz'menko, "The effect of series and parallel connection of a correcting circuits elements on the duration of a piezoelectric probe's acoustic signal," Russ. J. Nondestruct. Test., **42**, 737–743 (2006).
[5]S. I. Konovalov and A. G. Kuz'menko, "Study of the possibility of producing a short acoustic pulse using an inductive-resistive circuit connected to the input of a radiator loaded on a solid through a contact layer," Russ. J. Nondestruct. Test. **40**, 729–731 (2004).
[6]D. A. Berlincourt, D. R. Curran, and H. Jaffe, in *Physical Acoustics*, edited by W. P. Mason (Academic, New York, 1964), Vol. **1A**.

Conformal Fourier wavenumber decompositions on continuous differentiable surfaces

Anthony J. Romano,[a] P. B. Abraham, N. P. Valdivia, and B. H. Houston
Naval Research Laboratory, Physics Acoustics Branch, Washington, DC 20375-0001

J. A. Bucaro
SET Inc., Springfield, Virginia 22151

S. Dey
SFA Inc., Crofton, Maryland 21114

(Received 13 May 2008; revised 4 December 2008; accepted 12 December 2008)

In this manuscript, a method is introduced for the evaluation of Fourier wavenumber decompositions on C^1 vibrating surfaces for spatial-spectral analysis. Whereas typical Fourier analysis is restricted to geometries that are separable for meaningful interpretations of the corresponding wave motion, this approach allows for conformal spectral analysis along curved surfaces. This is accomplished by restricting the wavevectors to lie within the local tangent to the surface and to be spatially dependent. The theoretical development is presented and it is demonstrated that commonly utilized kernels appropriate for some simple geometries can be recovered. Additionally, this approach is applied in the analysis of the vibration and radiation of a point driven, fluid loaded cone, where the displacements and pressures have been obtained using the finite element method. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3068450]

PACS number(s): 43.40.At [EGW] Pages: 1461–1472

I. INTRODUCTION

Most of the tools (both standard as well as specialized) for understanding radiation, scattering, and the relations between sound, elastic vibration, and structural detail exploit knowledge about wavenumber space (spatial Fourier analysis) and the extrapolation of wavefields from one surface to the other.[1]

Up to now, the application of Fourier decomposition and analysis and the powerful techniques based on them have been restricted to separable geometries (planes, spheres, cylinders, etc.). This is primarily because the Fourier kernels (and their corresponding basis functions) for the spatial transforms of these simple geometries provide a direct and physical interpretation of the vibrational behavior. For nonseparable geometries, however, such a direct interpretation is not yet available.

In the past, efforts to adapt simple Fourier decompositions to accommodate more complex structures have involved rather *ad hoc* remedies. These include, for example, the approximation of complex, curved structures with locally planar, piecewise continuous transform "elements" (or plates), and the axial "weighting" of the radius within a cylindrical transform as an estimation for the transforms along conical and ellipsoidal structures. Although these representations seemed logical and provided particular feature spaces, their physical interpretations were left to speculation.

Concerning the extrapolation of wavefields from one surface to another (as in nearfield acoustical holography), several advanced methods such as boundary element methods (BEMs) and equivalent source methods[2,3] have been developed which can accommodate these projections; however, they do not provide a direct analysis and interpretation of the vibrational fields in a conformal wavenumber spectrum.

In this paper, we introduce a new method which enables two-dimensional spatial Fourier decompositions on C^1 surfaces. This is an extension of our previous work in waveguide constrained magnetic resonance elastography,[4] wherein we performed dispersion analysis along particular fiber pathways within a tapered waveguide. That study was able to extract spatially dependent velocities of propagation even though the material comprising the structure did not vary significantly throughout. Here, we wish to extend that development to surfaces.

In Sec. II, we present the theoretical development of this approach from fundamental principles using differential geometry, and in Sec. III, we demonstrate how the Fourier transform representations of some simple geometries can be recovered. In Sec. IV, we develop the particular form of this generalized transform which is applicable to a conical surface, and in Sec. V, we analyze the vibration of a point driven, fluid loaded conical structure whose displacements and pressures were calculated using the finite element method. Finally, in Sec. VI, we analyze the radiation behavior from the conical structure and utilize our transform method to evaluate the farfield radiating components.

II. THEORETICAL DEVELOPMENT

In this section, we briefly describe the background for this method and provide the theoretical development for the

[a]Electronic mail: anthony.romano@nrl.navy.mil

extension to surface transforms. We begin by outlining a previously published method we developed for dispersion analysis along arbitrarily oriented fibers and waveguides, which is provided below in Sec. II A. In Secs. II B and II C, we subsequently demonstrate how this approach can be extended to arbitrary surfaces.

A. Background

As presented in Ref. 4, we were posed the problem of analyzing wave propagation along a fibrous, tapered waveguide. We were provided with a knowledge of the three-dimensional elastic displacements throughout a volume containing the structure, as well as the orientations and locations of the individual fibers comprising the structure. To evaluate the propagation behavior, we developed a spatial-spectral filter which provided only those wave components which were traveling parallel along the individual fibers within the structure. With this information, a sliding window spatial Fourier transform was then performed along the individual fiber pathways utilizing arclength, τ, as a parameter, for localized dispersion analysis, and the transform appeared as

$$U_i(k_\tau) = \int_{\epsilon_L} u_i(\boldsymbol{r}(\tau)) e^{-\imath k_\tau \tau} d\tau, \tag{1}$$

where $u_i(\boldsymbol{r}(\tau))$, $i=1,2,3$, are the local displacement components, $U_i(k_\tau)$ their corresponding transformed quantities, k_τ is the wavenumber component corresponding to the arclength variable τ, and ϵ_L is the sliding aperture window length. This specific representation provided an analysis of the zero-order waveguide components[5] along individual pathways within the structure. The method yielded results indicating that the wave velocities were spatially varying along the length of the structure even though the material parameters comprising the structure did not vary significantly.

B. Extension to surfaces

In this publication, we wish to extend this formulation to C^1 surfaces for two-dimensional wavenumber decomposition. Whereas the previous representation required that the wavevectors depend on the unit tangent vector along a fiber pathway (i.e., a line), here, we require that our wavevectors lie within a local surface tangent plane.

We begin by requiring that our surface be characterizable as C^1. Specifically, this implies that (1) the surface can be parametrized providing the position, and (2) that this parametric representation allows for first order spatial derivatives providing the corresponding surface tangent vectors.

Next, we define two surface tangent vectors $\boldsymbol{T}_1(\tau_1,\tau_2)$ and $\boldsymbol{T}_2(\tau_1,\tau_2)$, where τ_1 and τ_2 are the parametric variables defining the surface. We can represent a surface parametrically by means of a function $\mathbf{r}:\mathcal{T}\rightarrow R^3$, where $\mathcal{T}\subset R^2$. So, for each $(\tau_1,\tau_2)\in\mathcal{T}$, we have

$$\mathbf{r}(\tau_1,\tau_2) = \langle x_1(\tau_1,\tau_2), x_2(\tau_1,\tau_2), x_3(\tau_1,\tau_2)\rangle, \tag{2}$$

where x_1, x_2, x_3 are the rectangular coordinate functions of the representation, τ_1, τ_2 are the parameters, and the brackets $\langle\ \rangle$ imply that they are vector components.[6] As (τ_1,τ_2) ranges over $\mathcal{T}$, $\boldsymbol{r}(\tau_1,\tau_2)$ gives the position vector of points on the surface. The surface tangent vectors may now be defined as

$$\boldsymbol{T}_1(\tau_1,\tau_2) = \frac{\partial \boldsymbol{r}(\tau_1,\tau_2)}{\partial \tau_1} = \frac{\partial x_1(\tau_1,\tau_2)}{\partial \tau_1}\hat{e}_1 + \frac{\partial x_2(\tau_1,\tau_2)}{\partial \tau_1}\hat{e}_2 + \frac{\partial x_3(\tau_1,\tau_2)}{\partial \tau_1}\hat{e}_3, \tag{3}$$

and

$$\boldsymbol{T}_2(\tau_1,\tau_2) = \frac{\partial \boldsymbol{r}(\tau_1,\tau_2)}{\partial \tau_2} = \frac{\partial x_1(\tau_1,\tau_2)}{\partial \tau_2}\hat{e}_1 + \frac{\partial x_2(\tau_1,\tau_2)}{\partial \tau_2}\hat{e}_2 + \frac{\partial x_3(\tau_1,\tau_2)}{\partial \tau_2}e_3, \tag{4}$$

where $\hat{e}_1$, $\hat{e}_2$, and $\hat{e}_3$ are the orthogonal unit vector components, while the normalized representations of the surface tangent vectors will be labeled as $\hat{T}_1(\tau_1,\tau_2)$ and $\hat{T}_2(\tau_1,\tau_2)$, respectively. These quantities allow us to define a unit normal vector, $\hat{N}(\tau_1,\tau_2)$, to the surface as

$$\hat{N}(\tau_1,\tau_2) = \hat{T}_1(\tau_1,\tau_2) \times \hat{T}_2(\tau_1,\tau_2). \tag{5}$$

At each point $\mathbf{r}(\tau_1,\tau_2)$ on the surface, the vectors, $\hat{T}_1(\tau_1,\tau_2)$, $\hat{T}_2(\tau_1,\tau_2)$, and $\hat{N}(\tau_1,\tau_2)$ define the "Gauss trihedron"[6] which acts as a basis for the space of vectors at $\boldsymbol{r}(\tau_1,\tau_2)$. On an arbitrary surface, the vectors $\boldsymbol{T}_1(\tau_1,\tau_2)$ and $\boldsymbol{T}_2(\tau_1,\tau_2)$ [and correspondingly $\hat{T}_1(\tau_1,\tau_2)$ and $\hat{T}_2(\tau_1,\tau_2)$] may not necessarily be orthogonal vectors. However, any two vectors may be orthogonalized either by specific parametrization or by utilizing such methods as, for example, Gram–Schmidt.[7]

C. The surface transform

Since the waves which we wish to transform have propagating components along a surface, we need only to concern ourselves with transforms along particular arclengths, as demonstrated above in Eq. (1), and account for their corresponding surface areas. Continuing with our development, we see that the arclengths labeled as $s_1(\tau_1,\tau_2)$ and $s_2(\tau_1,\tau_2)$, which are traced along the surface of the structure, are related to their respective surface tangent vectors, Eqs. (3) and (4), as

$$s_1(\tau_1,\tau_2) = \int |\boldsymbol{T}_1(\tau_1,\tau_2)| d\tau_1, \tag{6}$$

and

$$s_2(\tau_1,\tau_2) = \int |\boldsymbol{T}_2(\tau_1,\tau_2)| d\tau_2, \tag{7}$$

respectively, where we have expressed these quantities in terms of their indefinite integrals. As a consequence of these definitions, the associated wavenumbers may assume two specific and individual forms, k_1 and k_2, where k_1 multiplies the arclength $s_1(\tau_1,\tau_2)$ in one transform kernel, while k_2 multiplies the arclength $s_2(\tau_1,\tau_2)$ in a second kernel. It should be mentioned that the wavenumbers, k_1 and k_2, may be spatially

dependent as well along the surfaces as will be demonstrated below. Additionally, the corresponding surface area is related to the magnitude of the cross product of the surface tangent vectors as $|\boldsymbol{T}_1(\tau_1,\tau_2)\times\boldsymbol{T}_2(\tau_1,\tau_2)|d\tau_1 d\tau_2$. Using these definitions,[8,9] we see that the forward two-dimensional spatial Fourier transform of a surface quantity, $f(\mathbf{r}(\tau_1,\tau_2))$, lying on the surface assumes the following form:

$$F(k_1,k_2)=\int\int f(\mathbf{r}(\tau_1,\tau_2))e^{-\imath k_1 s_1(\tau_1,\tau_2)}e^{-\imath k_2 s_2(\tau_1,\tau_2)} \times|\boldsymbol{T}_1(\tau_1,\tau_2)\times\boldsymbol{T}_2(\tau_1,\tau_2)|d\tau_1 d\tau_2. \quad (8)$$

Here, the function $f(\mathbf{r}(\tau_1,\tau_2))$ can represent surface pressure, surface displacements or velocities, or the parametric representation of a surface itself. Additionally, the vector displacements or velocities may be dotted with a locally rotating reference frame as demonstrated in Ref. 4.

The generalized inverse transform of Eq. (8) may be represented as

$$f(\mathbf{r}(\tau_1,\tau_2))=\frac{1}{(2\pi)^2}\int\int F(k_1,k_2) \times e^{\imath k_1 s_1(\tau_1,\tau_2)}e^{\imath k_2 s_2(\tau_1,\tau_2)}dk_1 dk_2. \quad (9)$$

A generalized analytical proof that Eqs. (8) and (9) form a Fourier transform pair is not a simple undertaking, particularly in light of the fact that the wavenumbers, k_1 and k_2 can, themselves, be spatially dependent. Additionally, there may be issues of orthogonality requirements as mentioned above. In what follows, however, we will demonstrate how these representations can identically recover the wavevector representations of several well-known simple geometries, and then extend this approach to a conical surface for vibration and radiation analysis.

III. APPLICATION TO SIMPLE GEOMETRIES

In this section, we demonstrate that our new development reduces to the well- known formalisms on structures which are defined by separable geometries such as a plate, a cylinder, and a sphere. Here, we define the particular parametrizations, tangent vectors, surface areas, and wavenumbers appropriate for these simple cases, and demonstrate how their corresponding forms of the spatial Fourier transform can be identically recovered.[1]

A. Cartesian geometry

Consider first a two-dimensional surface lying in the $\boldsymbol{r}=\langle x,y,0\rangle$ plane. In this case, the parametric variables may be defined as $\tau_1=x$ and $\tau_2=y$. The surface tangent vectors [Eqs. (3) and (4)] become $\boldsymbol{T}_1(x,y)=\langle 1,0,0\rangle$ and $\boldsymbol{T}_2(x,y)=\langle 0,1,0\rangle$, respectively, while their normalized representations appear as $\hat{T}_1(x,y)=\langle 1,0,0\rangle$ and $\hat{T}_2(x,y)=\langle 0,1,0\rangle$. As can be verified by inspection, the two tangent vectors are orthogonal (since $\boldsymbol{T}_1(x,y)\cdot\boldsymbol{T}_2(x,y)=\hat{T}_1(x,y)\cdot\hat{T}_2(x,y)=0$), while the curl of their normalized representations defines the surface normal, $\hat{N}(x,y)=\hat{T}_1(x,y)\times\hat{T}_2(x,y)=\langle 0,0,1\rangle$. In this case, the three vectors $\hat{T}_1(x,y)$, $\hat{T}_2(x,y)$, and $\hat{N}(x,y)$ define an orthogonal reference frame. Letting $k_1=k_x$, $k_2=k_y$, $d\tau_1=dx$, $d\tau_2=dy$, $s_1(x,y)=x$, $s_2(x,y)=y$, and $|\boldsymbol{T}_1(x,y)\times\boldsymbol{T}_2(x,y)|=1$, Eq. (8) becomes

$$F(k_x,k_y)=\int\int f(\mathbf{r}(x,y))e^{-\imath k_x x}e^{-\imath k_y y}dxdy, \quad (10)$$

which is the two-dimensional spatial Fourier transform representation of the function $f(\mathbf{r}(x,y))$.

The generalized inverse transform of Eq. (10) may be obtained from Eq. (9) and appears as

$$f(\mathbf{r}(x,y))=\frac{1}{(2\pi)^2}\int\int F(k_x,k_y)e^{\imath k_x x}e^{\imath k_y y}dk_x dk_y, \quad (11)$$

which is the two-dimensional inverse Fourier transform representation of the function $F(k_x,k_y)$.

B. Cylindrical geometry

Consider next a cylindrical surface lying in the $\mathbf{r}=\langle\rho,\theta,z\rangle$ coordinate system with $x=\rho\cos\theta$, $y=\rho\sin\theta$, and $z=z$. In this case, the parametric variables of the surface at a fixed radius, $\rho=a$, may be defined as $\tau_1=\theta$ and $\tau_2=z$. The surface tangent vectors [Eqs. (3) and (4)] become $\boldsymbol{T}_1(\theta,z)=\langle -a\sin\theta,a\cos\theta,0\rangle$ and $\boldsymbol{T}_2(\theta,z)=\langle 0,0,1\rangle$, respectively, while their normalized representations appear as $\hat{T}_1(\theta,z)=\langle -\sin\theta,\cos\theta,0\rangle$ and $\hat{T}_2(\theta,z)=\langle 0,0,1\rangle$. As can been seen by inspection, the two tangent vectors are orthogonal [since $\boldsymbol{T}_1(\theta,z)\cdot\boldsymbol{T}_2(\theta,z)=\hat{T}_1(\theta,z)\cdot\hat{T}_2(\theta,z)=0$], while the curl of their normalized representations defines the surface normal, $\hat{N}(\theta,z)=\hat{T}_1(\theta,z)\times\hat{T}_2(\theta,z)=\langle\cos\theta,\sin\theta,0\rangle$, and the three vectors $\hat{T}_1(\theta,z)$, $\hat{T}_2(\theta,z)$, and $\hat{N}(\theta,z)$ define an orthogonal reference frame. Letting $k_1=k_\theta$, $k_2=k_z$, $d\tau_1=d\theta$, $d\tau_2=dz$, $s_1(\theta,z)=a\theta$, $s_2(\theta,z)=z$, and $|\boldsymbol{T}_1(\theta,z)\times\boldsymbol{T}_2(\theta,z)|=a$, Eq. (8) becomes

$$F(k_\theta,k_z)=a\int_0^{2\pi}d\theta\int f(\mathbf{r}(a,\theta,z))e^{-\imath k_\theta a\theta}e^{-\imath k_z z}dz, \quad (12)$$

which is the two-dimensional spatial Fourier transform representation of the function $f(\mathbf{r}(a,\theta,z))$.

The generalized inverse transform of Eq. (12) may be obtained from Eq. (9) and appears as

$$f(\mathbf{r}(a,\theta,z))=\frac{1}{(2\pi)^2}\int\int F(k_\theta,k_z)e^{\imath k_\theta a\theta}e^{\imath k_z z}dk_\theta dk_z, \quad (13)$$

which is the two-dimensional inverse Fourier transform representation of the function $F(k_\theta,k_z)$.

C. Spherical geometry

Consider next a spherical surface lying in the $\mathbf{r}=\langle\rho,\theta,\phi\rangle$ coordinate system with $x=\rho\sin\theta\cos\phi$, $y=\rho\sin\theta\sin\phi$, and $z=\rho\cos\theta$. In this case, the parametric variables of the surface at a fixed radius, $\rho=a$, may be defined as $\tau_1=\theta$ and $\tau_2=\phi$. The surface tangent vectors [Eqs. (3) and (4)] become $\boldsymbol{T}_1(\theta,\phi)=\langle a\cos\theta\cos\phi, a\cos\theta\sin\phi,-a\sin\theta\rangle$ and $\boldsymbol{T}_2(\theta,\phi)=\langle -a\sin\theta\sin\phi,a\sin\theta\cos\phi,0\rangle$, respectively, while their

normalized representations appear as $\hat{T}_1(\theta,\phi)=\langle\cos\theta\cos\phi,\cos\theta\sin\phi,-\sin\theta\rangle$ and $\hat{T}_2(\theta,\phi)=\langle-\sin\phi,\cos\phi,0\rangle$. As can be verified by inspection, the two tangent vectors are orthogonal [since $\boldsymbol{T}_1(\theta,\phi)\cdot\boldsymbol{T}_2(\theta,\phi)=\hat{T}_1(\theta,\phi)\cdot\hat{T}_2(\theta,\phi)=0$], while the curl of their normalized representations defines the surface normal, $\hat{N}(\theta,\phi)=\hat{T}_1(\theta,\phi)\times\hat{T}_2(\theta,\phi)=\langle\sin\theta\cos\phi,-\sin\theta\sin\phi,\cos\theta\rangle$. In this case, the three vectors $\hat{T}_1(\theta,\phi)$, $\hat{T}_2(\theta,\phi)$, and $\hat{N}(\theta,\phi)$ define an orthogonal reference frame. Letting $k_1=k_\theta$, $k_2=k_\phi$, $d\tau_1=d\theta$, $d\tau_2=d\phi$, $s_1(\theta,\phi)=a\theta$, $s_2(\theta,\phi)=a\sin\theta\phi$, and $|\boldsymbol{T}_1(\theta,\phi)\times\boldsymbol{T}_2(\theta,\phi)|=a^2\sin\theta$, Eq. (8) becomes

$$F(k_\theta,k_\phi)=a^2\int_0^{2\pi}d\phi\int_0^{\pi}f(\boldsymbol{r}(a,\theta,\phi))e^{-\imath k_\theta a\theta}e^{-\imath k_\phi a\sin\theta\phi}\times\sin\theta d\theta, \tag{14}$$

which is the two-dimensional spatial Fourier transform representation of the function $f(\boldsymbol{r}(a,\theta,\phi))$. Generally, a surface spherical Fourier transform is accomplished using spherical harmonic basis functions;[1] however, it can be shown that such a representation is equivalent to the representation shown in Eq.(14).

The generalized inverse transform of Eq. (14) may be obtained using Eq. (9) and appears as

$$f(\boldsymbol{r}(a,\theta,\phi))=\frac{1}{(2\pi)^2}\int\int F(k_\theta,k_\phi)e^{\imath k_\theta a\theta}e^{\imath k_\phi a\sin\theta\phi}dk_\theta dk_\phi. \tag{15}$$

which is the two-dimensional inverse Fourier transform representation of the function $F(k_\theta,k_\phi)$.

IV. APPLICATION TO A CONICAL GEOMETRY

In this section, we apply the generalized expression, as shown in Eq. (8), to a conical structure for vibration and radiation analysis. Generally, this analysis can be performed in either a global or a localized fashion. In the global analysis, the integral representation can be performed over the entire surface area of a continuous structure being interrogated, while in the localized analysis, the transform, as shown in Eq. (8), can be applied over individual segments of a structure, with each segment being continuous within itself. Additionally, a local window may be moved along the surface in a specific manner so as to investigate the spatial dependence of the dispersion spectra as was demonstrated in Ref. 4, which utilized a sliding window along a particular pathway. Here, we demonstrate the transform in a global fashion on the conical section of an endcapped, fluid loaded, point driven structure.

Consider a conical surface lying in the $\boldsymbol{r}=\langle\rho(z),\theta,z\rangle$ coordinate system with $x=\rho(z)\cos\theta$, $y=\rho(z)\sin\theta$, and $z=z$. Let us define the function $\rho(z)$ as $\rho(z)=a_0[1-z/l_z]$, where a_0 is the radius of the base of the cone and l_z is the length of the cone. In this case, the parametric variables of the surface may be defined as $\tau_1=\theta$ and $\tau_2=z$. The surface tangent vectors [Eqs. (3) and (4)] become

$$\boldsymbol{T}_1(\theta,z)=\left\langle-a_0\left[1-\frac{z}{l_z}\right]\sin\theta,a_0\left[1-\frac{z}{l_z}\right]\cos\theta,0\right\rangle \tag{16}$$

and

$$\boldsymbol{T}_2(\theta,z)=\left\langle-\frac{a_0}{l_z}\cos\theta,-\frac{a_0}{l_z}\sin\theta,1\right\rangle, \tag{17}$$

respectively, while their normalized representations appear as

$$\hat{T}_1(\theta,z)=\langle-\sin\theta,\cos\theta,0\rangle \tag{18}$$

and

$$\hat{T}_2(\theta,z)=\frac{\left\langle-\frac{a_0}{l_z}\cos\theta,-\frac{a_0}{l_z}\sin\theta,1\right\rangle}{\sqrt{1+\left(\frac{a_0}{l_z}\right)^2}}. \tag{19}$$

As can be seen by inspection, the two tangent vectors are orthogonal [since $\boldsymbol{T}_1(\theta,z)\cdot\boldsymbol{T}_2(\theta,z)=\hat{T}_1(\theta,z)\cdot\hat{T}_2(\theta,z)=0$], while the curl of their normalized representations defines the surface normal,

$$\hat{N}(\theta,z)=\hat{T}_1(\theta,z)\times\hat{T}_2(\theta,z)=\frac{\left\langle\cos\theta,\sin\theta,\frac{a_0}{l_z}\right\rangle}{\sqrt{1+\left(\frac{a_0}{l_z}\right)^2}}, \tag{20}$$

and the three vectors $\hat{T}_1(\theta,z)$, $\hat{T}_2(\theta,z)$, and $\hat{N}(\theta,z)$ define an orthogonal reference frame. Letting $k_1=k_\theta$ (representing the wavenumber around the circumferential direction of the cone), $k_2=k_z$ (representing the wavenumber along the axial surface length of the cone), $d\tau_1=d\theta$, $d\tau_2=dz$, $s_1(\theta,z)=a_0[1-z/l_z]\theta$, $s_2(\theta,z)=[\sqrt{1+(a_0/l_z)^2}]z$, and $|\boldsymbol{T}_1(\theta,z)\times\boldsymbol{T}_2(\theta,z)|=a_0[1-z/l_z]\sqrt{1+(a_0/l_z)^2}$, Eq. (8) becomes

$$F(k_\theta,k_z)=a_0\sqrt{1+\left(\frac{a_0}{l_z}\right)^2}\int_0^{2\pi}d\theta\int f\left(\boldsymbol{r}\left(a_0\left[1-\frac{z}{l_z}\right],\theta,z\right)\right)\times e^{-\imath k_\theta a_0[1-z/l_z]\theta}e^{-\imath k_z[\sqrt{1+(a_0/l_z)^2}]z}\left[1-\frac{z}{l_z}\right]dz. \tag{21}$$

In this representation, due to circular periodicity requirements, it can be seen that the circumferential wavenumber must assume the form $k_\theta=2n\pi/(2\pi a_0[1-z/l_z])$, n is an integer, while the axial wavenumber is represented as $k_z=2m\pi/L_{ax}$, where m is an integer and L_{ax} is the length of interrogation on the conical surface along the z-axis (i.e., $L_{ax}=\sqrt{l_z^2+a_0^2}$). It should be mentioned that these wavenumbers resemble those provided by Godzevich[10] in his Galerkin formulation of the Donnel–Mushtari shell equations for a conical structure; however, his wavenumbers are constant with respect to the structure, while those provided here are spatially dependent, taking into account the variation of radius with respect to axial location.

The generalized inverse to Eq. (21) may be obtained using Eq. (9) and appears as

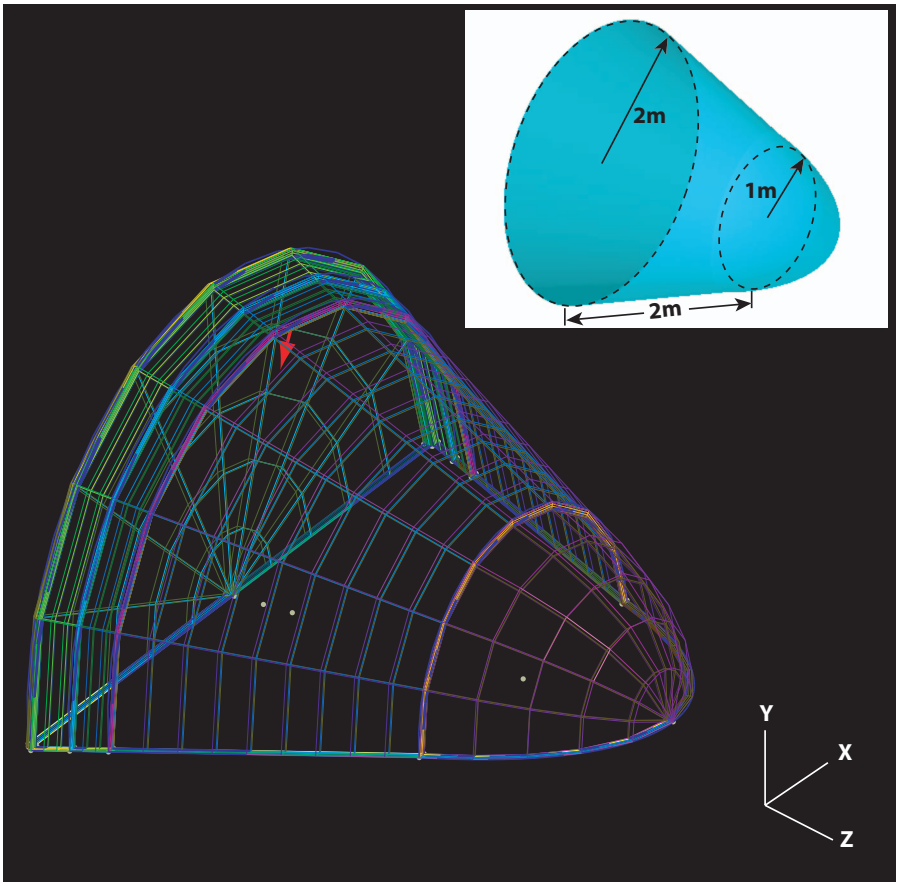


FIG. 1. Images of the conical structure used in the finite element study showing both the dimensions as well as the finite element mesh. Red arrow indicates location and direction of the applied point force.

$$f\left(\boldsymbol{r}\left(a_0\left[1-\frac{z}{l_z}\right],\theta,z\right)\right)=\frac{1}{(2\pi)^2}\int\int F(k_\theta,k_z)e^{\imath k_\theta a_0[1-z/l_z]\theta}\times e^{\imath k_z[\sqrt{1+(a_0/l_z)^2}]z}dk_\theta dk_z, \quad (22)$$

where, for a finite structure, $dk_\theta \sim dk_\theta/dn=2\pi/2\pi a_0[1-z/l_z]$, and $dk_z \sim dk_z/dm=2\pi/L_{\mathrm{ax}}$, respectively. In what follows, we will demonstrate the application of this transform pair.

V. DEMONSTRATION USING NUMERICAL DATA

In this section, we demonstrate the application of the spatial wavenumber decomposition on a finite element simulation of a point driven, fluid-loaded, conical structure. With reference to Fig. 1, we see the image and dimensions of the physical structure, as well as an image of the finite element mesh utilized in this study. As can be seen in this figure, the conical length is 2 m, the larger conical end has a radius of 2 m, and the smaller conical radius is 1 m. The left end is terminated with a flat, circular plate, while the right end is terminated with a hemispherical shell. The spatial parameter $l_z=4$ m, and $L_{\mathrm{ax}}=\sqrt{5}$ m. The thickness of all three parts of the structure is 1 cm, and they are all composed of nickel with Young's modulus of 2.2×10^{11} Pa, a density of 7800 kg/m^3, Poisson's ratio of 0.3, and a damping ratio of 0.01. The conical structure was immersed in water with a density of 1000 kg/m^3 and a velocity of 1500 m/s.

Concerning the numerical study, a parallel hp-version of a finite element method[11,12] was utilized to determine the structural-acoustic response due to a point force applied to the interior surface of a conical shell. After a hp-convergence analysis, an approximation using fourth degree approximation for both the pressure and displacement fields was utilized. The spatial mesh, shown in Fig. 1, has 6720 elastic and 13 440 fluid elements. 3360 unconjugated infinite elements[13] of radial degree 2 were utilized to truncate the infinite exterior fluid domain. The resulting discrete model consists of 659 772 complex-valued unknown coefficients. The data were delivered to the transform over a grid of 201 locations axially and 720 locations circumferentially. As a demonstration of our transform method, the surface displacements and pressures were determined over the conical structure for two frequencies: 50 and 500 Hz. With regard to Figs. 2 and 3, we see the real part of the normal surface displacements and pressures, respectively, for 50 and 500 Hz. As can be seen, two images per dataset are provided. While one image displays the image as a cone, to better portray the modal structures, the conical sections were unwrapped onto a corresponding two-dimensional plane.

There are numerous, interesting observations provided by this new transform method. With regard to Eq. (21), we see by the definition of the parametrization and the wavenumbers that the conical structure can be seen to support helical wave propagation. Unlike helical waves on a uniform cylindrical surface,[1] however, these modes are spatially dependent and their resulting velocities vary as a function of position. To demonstrate this, let us define a helical wave velocity, c_h, as

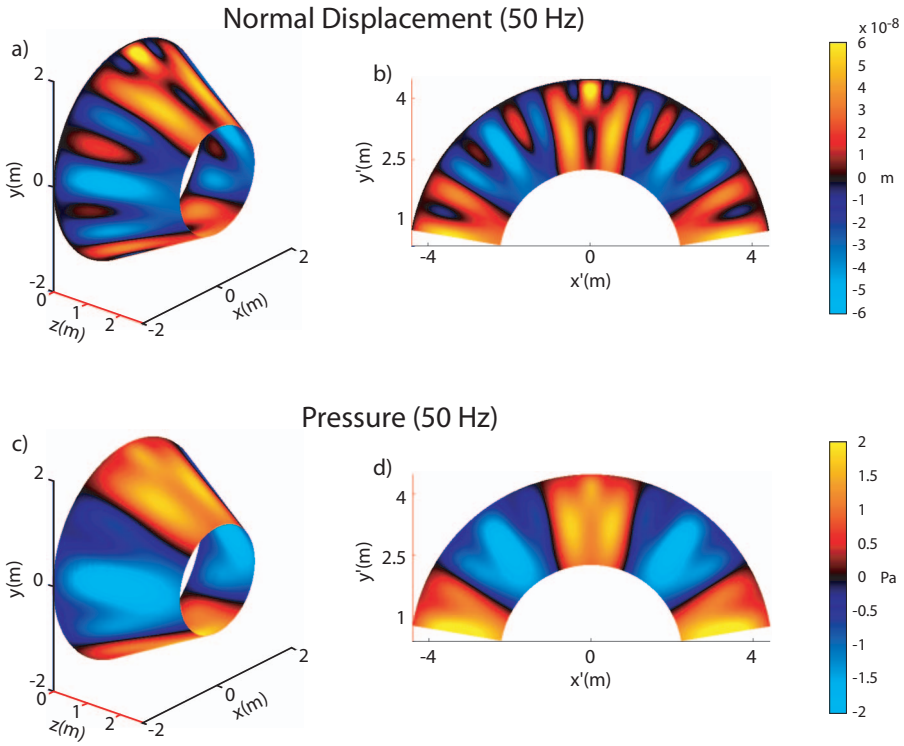


FIG. 2. Normal surface displacement [(a) and (b)] and surface pressure [(c) and (d)] over the conical surface of the structure for 50 Hz excitation. Images (a) and (c) represent a perspective image of the cone while images (b) and (d) represent the unwrapping of the conical surface onto a plane with relative axes x' and y'.

$$c_h(\omega,m,n,z) = \frac{\omega}{\sqrt{k_z^2 + k_\theta^2}} = \frac{\omega}{\sqrt{\left(\frac{2m\pi}{L_{\text{ax}}}\right)^2 + \left(\frac{n}{a_0\left[1 - \frac{z}{l_z}\right]}\right)^2}}. \quad (23)$$

Since k_θ is dependent on the z-coordinate, so are the helical wave velocities. In Fig. 4, we show the velocity profiles of these helical modes as functions of some low order circumferential and axial mode numbers, and axial location using Eq. (23). As can be observed, for the $n=0$ breathing mode, there is no spatial variation of the corresponding helical wave velocity [with regard to Figs. 4(b) and 4(d)]. However, for the higher order circumferential modes, we see that the helical wave velocities become slower as the waves propagate toward the smaller end of the cone. Additionally, as can

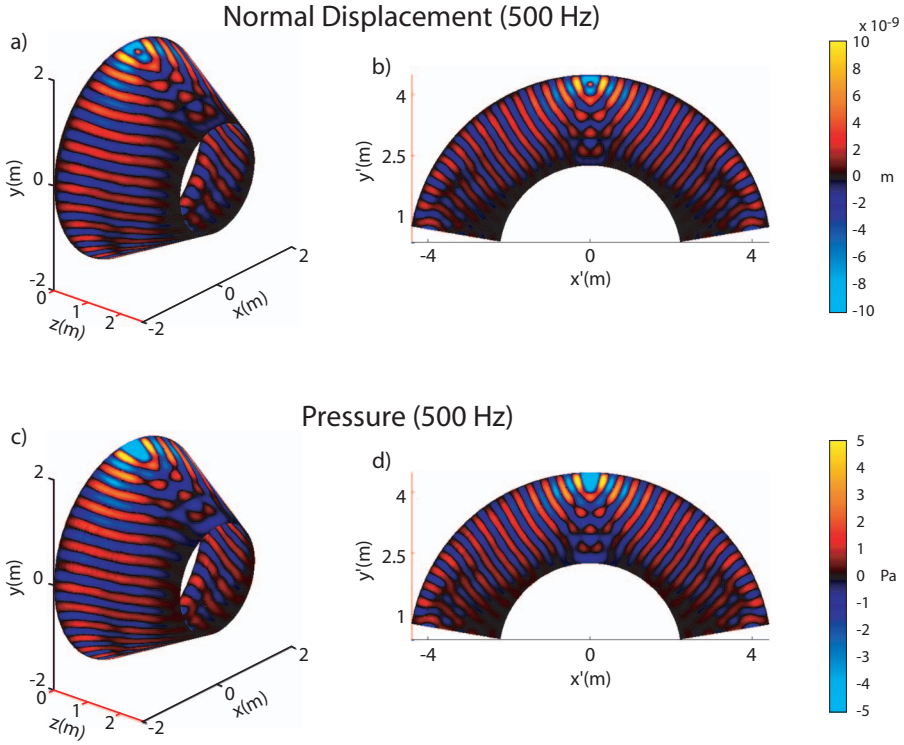


FIG. 3. Normal surface displacement [(a) and (b)] and surface pressure [(c) and (d)] over the conical surface of the structure for 500 Hz excitation. Images (a) and (c) represent a perspective image of the cone while images (b) and (d) represent the projection of the conical surface onto a plane with relative axes x' and y'.

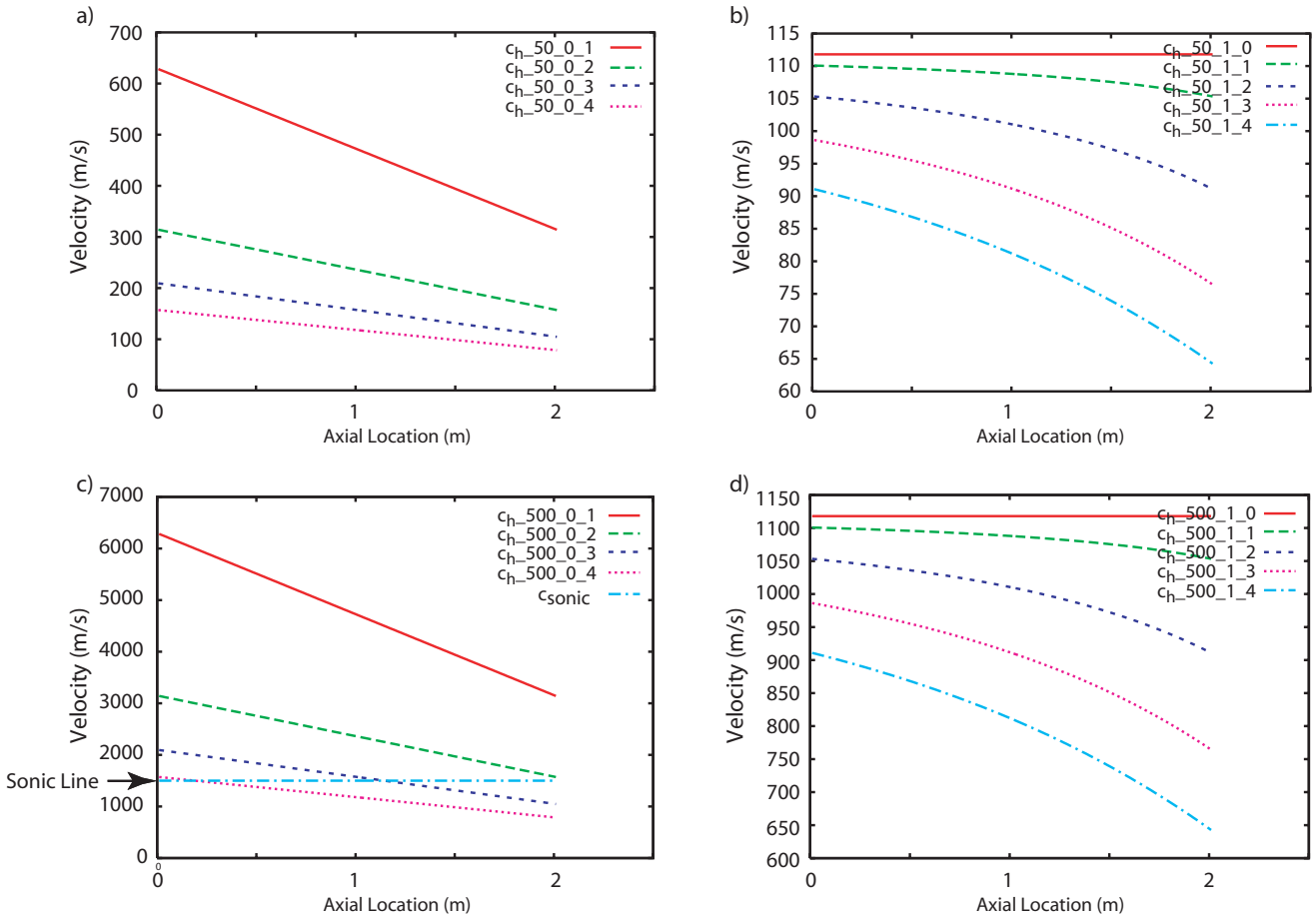


FIG. 4. Images displaying helical wave velocity, c_h [Eq. (23)], for the two frequencies of excitation and various modal combinations as functions of axial location on the cone. (a) c_h at 50 Hz for modes $m=0$, $n=1,2,3,4$, respectively. (b) c_h at 50 Hz for modes $m=1$, $n=0,1,2,3,4$, respectively. (c) c_h at 500 Hz for modes $m=0$, $n=1,2,3,4$, respectively, as well as the velocity of sound in the fluid. (d) c_h at 500 Hz for modes $m=1$, $n=0,1,2,3,4$, respectively.

be observed from Eq. (23), if a conical structure actually continued to a point, then the helical wave velocities become identically zero at the focus. This may seem counterintuitive, however because the radius decreases as the cone tapers, k_θ increases, thereby representing a slowing of the helical wave velocity. Another interesting result is that the sets of modes have supersonic "cutoff" locations, as shown in Fig. 4(c), wherein the $m=0$, $n=4$ mode and the $m=0$, $n=3$ mode drop below the velocity of wave propagation in the fluid. This will be discussed later in Sec. VI in reference to farfield radiation contributions. We continue with our analysis by performing the integration of Eq. (21) over the circumferential variable first, resulting in the function $F(n,z)$. As shown in Figs. 5 and 6, we can see the axial dependence of the various circumferential modes. For the 50 Hz case, there appear to be only two dominant circumferential modes appearing for $n=\pm 2$ and $n=\pm 9$, and are essentially constant along the length of the cone. For the 500 Hz case, however, there is a broad presence of circumferential harmonics near the driver axial location, with the $n=\pm 31$ component dominating. Moving down the cone away from the base, the modes shift toward lower n values, a phenomena observed by Lindholm and Hu[14] which they attributed to the higher stiffness and shorter distance between consecutive nodal meridians in this region which tends to suppress high n transverse motion. In any event, this decrease in mode number results in a constant

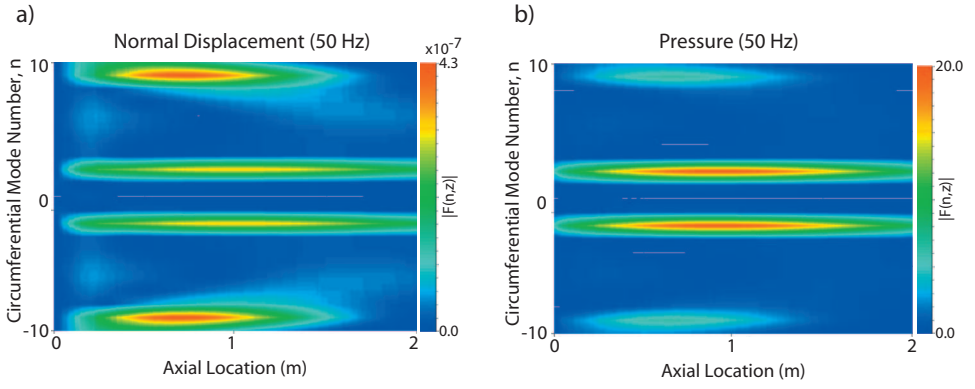


FIG. 5. Integration of the conformal transform for a cone [Eq. (21)] over the circumferential variable for (a) 50 Hz normal displacement and (b) 50 Hz pressure. The image portrays the function $|F(n,z)|$, which represents circumferential mode number as a function of axial location.

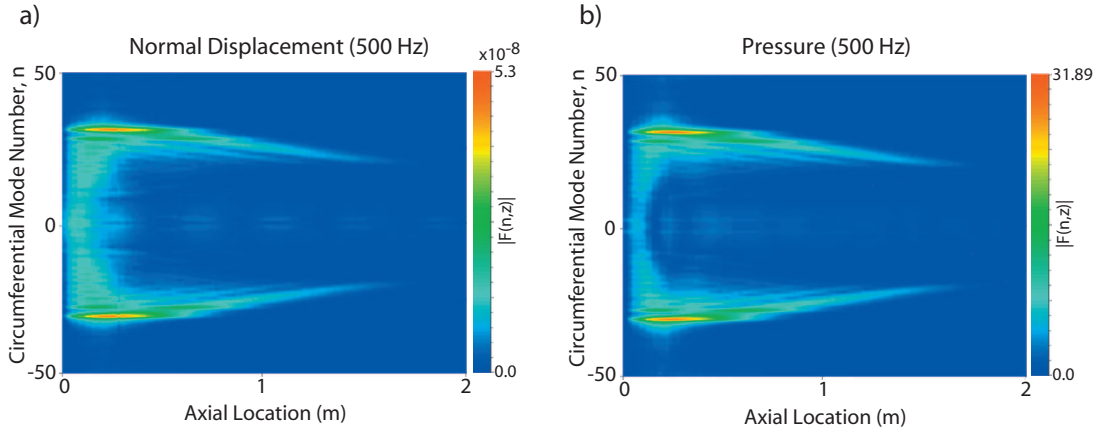


FIG. 6. Integration of the conformal transform for a cone [Eq. (21)] over the circumferential variable for (a) 500 Hz normal displacement and (b) 500 Hz pressure. The image portrays the function $|F(n,z)|$, which represents circumferential mode number as a function of axial location.

velocity of the wave as it propagates along the conical structure. That is, as it maintains a constant velocity down the cone, it assumes lower order circumferential modes as its propagation mechanism. This result was determined using Eq. (23) which indicated that for the $m=0$, $n=31$ mode, the velocity was 200 m/s near the driver. As the wave propagated down the cone, to maintain this same velocity, it converted, in a linear fashion, to lower order circumferential modes, becoming an $m=0$, $n=16$ mode at the smaller end of the cone.

Next, application of the axial integration of the surface transform [as shown in Eq. (21)] over the conical section results in the function $F(n,m)$, and is shown in Figs. 7 and 8. As can be observed, for the vibration at 50 Hz, the modes are concentrated near $k_z=0$, and have two distinct circumferential components for $n=\pm 2$ and $n=\pm 9$, which is not surprising from the result shown in Fig. 5. For the 500 Hz excitation, the character of the wavenumber spectrum is quite different from that observed at the lower frequency. First, there are noticeable components at low n ($n=\pm 2$) for $m\approx -5$. These are clearly backward-propagating waves, presumably related to reflection of these much shorter wavelength ($\lambda \approx l_z/5$) structure-borne waves from the smaller end of the cone. Second, although here too there exists a dominant n mode ($n=\pm 31$), in this case, there is significant additional modal structure whose striations suggest modal coupling effects.

Finally, we demonstrate the convergence of the inverse transform numerically. In Fig. 9, we show a slice in the circumferential direction near the axial drive location for the displacement at 500 Hz. The red line is the total displacement, while the subsequent lines provide the inversion of the modes including up to $m=\pm 10$, $n=\pm 10$; $m=\pm 20$, $n=\pm 20$; and $m=\pm 50$, $n=\pm 50$, respectively. As can be seen, the inverse is very well converged (to within a 1 percent total variation) for the final set of modes.

While it should be mentioned that much work has been performed on the vibration of conical structures in the past (notably the work of Godzevich,[10] Goldberg and Bogdanoff,[15] a review chapter by Leissa,[16] and more recently Guo[17,18]), our results are somewhat difficult to compare to theirs. Although their studies mention modal cutoffs as being spatially dependent as do we, the transform kernels we have defined in this study do not directly compare to dispersion profiles or specific mode shapes they provide, particularly since we have only investigated two individual frequencies. If we examined a broad spectrum of frequencies, we should be able to portray the families of wavetypes as a result of

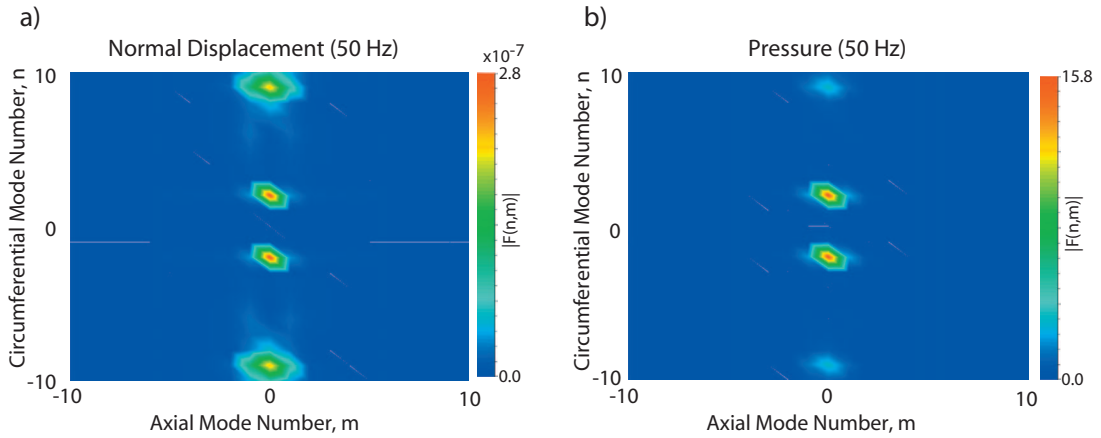


FIG. 7. Integration of the conformal transform for a cone [Eq. (21)] over both spatial variables for (a) 50 Hz normal displacement and (b) 50 Hz pressure. The image portrays the function $|F(n,m)|$, as functions of axial and circumferential mode numbers, m and n, respectively.

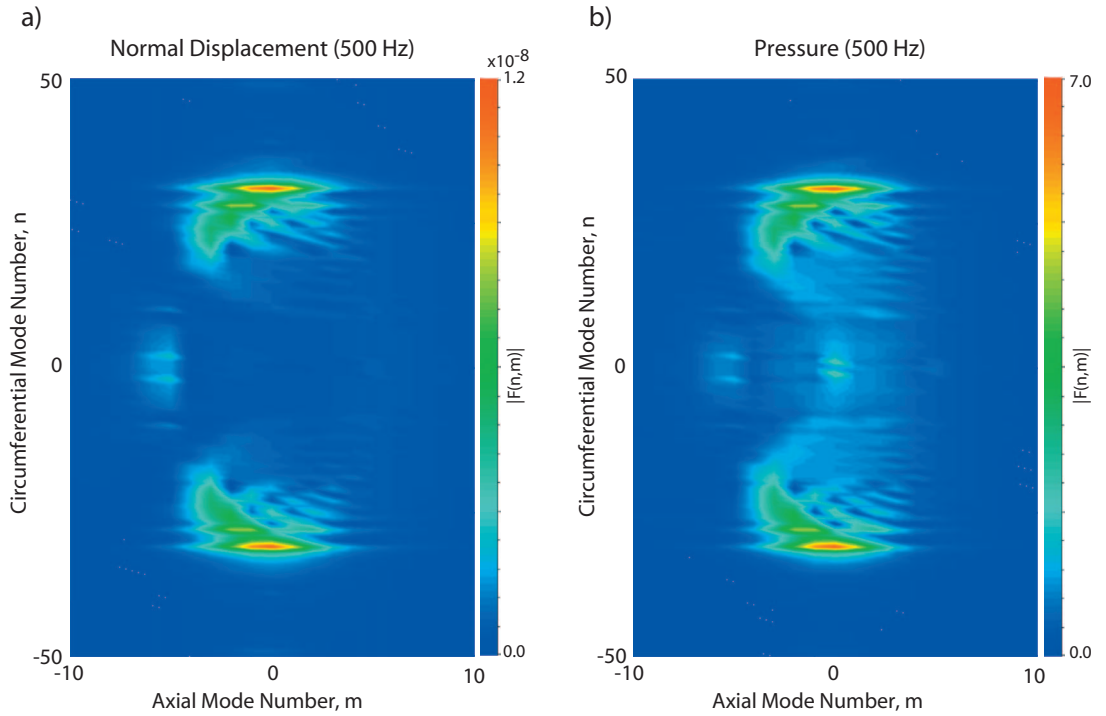


FIG. 8. Integration of the conformal transform for a cone [Eq. (21)] over both spatial variables for (a) 500 Hz normal displacement and (b) 500 Hz pressure. The image portrays the function $|F(n,m)|$, as functions of axial and circumferential mode numbers, m and n, respectively.

dispersion images. Guo,[18] however, demonstrated the axial dependence of wavenumbers (compressional, shear, and flexural) and showed that on a fluid-loaded cone, the compressional wavenumbers slow as one approaches the apex. Additionally, he demonstrated that for the $n=0$ circumferential harmonic, the compressional wavenumber displays no axial dependence on the fluid-loaded cone, a characteristic behavior we portray in Fig. 4 for the $m=1$, $n=0$ helical wave mode. Suffice it to say that one should be aware of these previous studies when placing the current work in context.

VI. FIELD SEPARATION USING THE HELMHOLTZ INTEGRAL

As a final component, we demonstrate the utility of the previously described transform method in the analysis of the farfield radiation behavior of the conical surface. Consider the farfield expression for the Helmholtz integral as shown below:

$$p_\infty(\hat{r}) = \frac{1}{4\pi}\int_\Gamma p(\boldsymbol{r}_0)\boldsymbol{n}\cdot\nabla_0 G(\hat{r},\boldsymbol{r}_0) - i\omega\rho G(\hat{r},\boldsymbol{r}_0)v_n(\boldsymbol{r}_0)d\Gamma, \tag{24}$$

where $\boldsymbol{r}_0$ is the surface location, $p(\mathbf{r}_0)$ is the surface pressure, $v_n(\mathbf{r}_0)$ is the normal surface velocity, $\mathbf{n}$ the outwardly pointing normal vector, and $G(\hat{r},\mathbf{r}_0)$ is the farfield Green's function $e^{-\imath k\hat{r}\cdot\boldsymbol{r}_0}e^{ik|\boldsymbol{r}|}/|\boldsymbol{r}|$ of the surrounding medium, $\hat{r}$ is the normalized field point, and Γ the surface of integration. This particular form can be shown to be the farfield expansion of

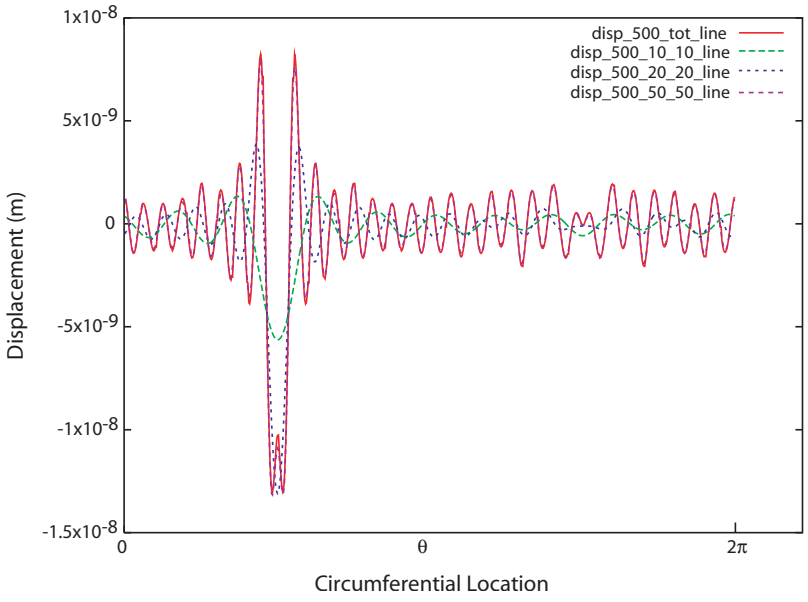


FIG. 9. Convergence demonstration of the inverse transform [Eq. (22)] of the decomposition of the normal displacement at 500 Hz [as shown in Figure 8(a)] for various modes compared to the total field [as shown in Figs. 3(a) and 3(b)]. Red line represents the original displacement near the axial location of the driver and for each circumferential position around the cone. Green dashed line represents the inverse transform including modes within $m=\pm 10$ and $n=\pm 10$. Blue dashed line represents the inverse transform including modes within $m=\pm 20$ and $n=\pm 20$. Red dashed line represents the inverse transform including modes within $m=\pm 50$ and $n=\pm 50$. Field is very well converged for the final set of modes.

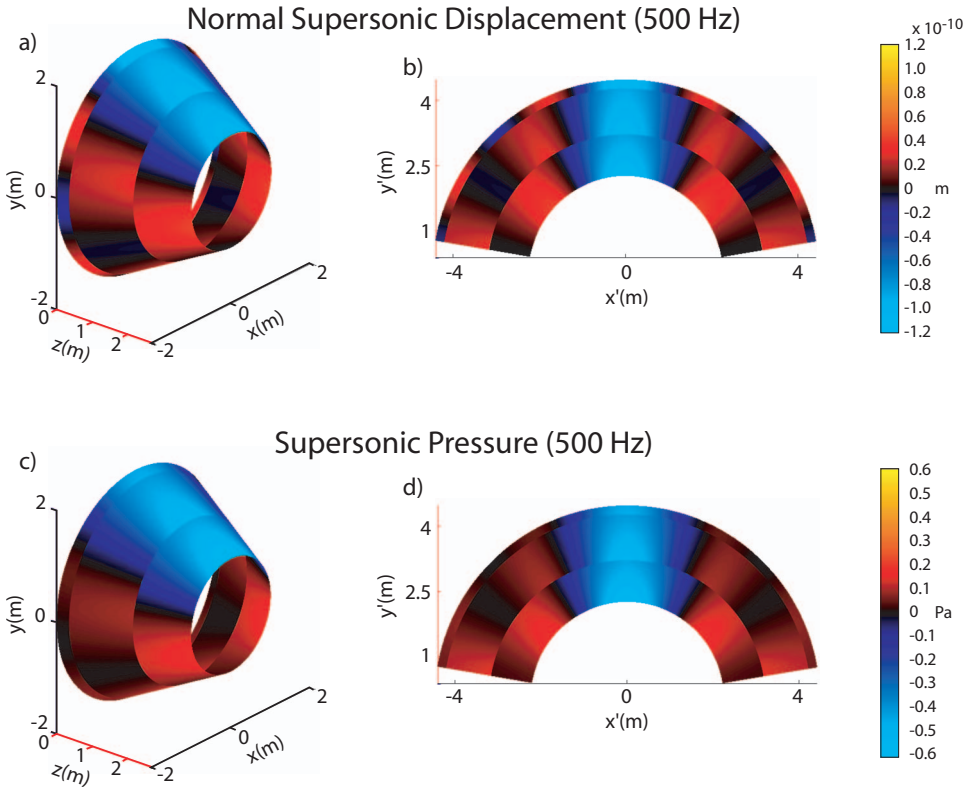


FIG. 10. Normal supersonic displacement [(a) and (b)] and supersonic pressure [(c) and (d)] for 500 Hz resulting from modal cutoff criteria, as shown in Fig. 4(c).

the free-space Green's function, $G(\boldsymbol{r},\boldsymbol{r}_0)=e^{ik|\boldsymbol{r}-\boldsymbol{r}_0|}/|\boldsymbol{r}-\boldsymbol{r}_0|$, for $|\boldsymbol{r}|\gg|\boldsymbol{r}_0|$.[19] We wish to use this expression to quantify those components out of the entire wavenumber spectrum which are responsible for farfield radiation.

It is well known[1] that for an infinite surface or structure, the wavenumber components which will more efficiently contribute to the farfield are those contained within the radiation circle, i.e., whose radius in wavenumber space is determined by the wavenumber in the surrounding medium. These components are often called "supersonic," because there structural wavenumbers give rise to velocities on the structure which are greater than the velocity of propagation in the fluid. All other wavenumber components represent velocities on the structure which are slower than the velocity of propagation in the fluid. These latter waves tend to decay exponentially away from the vibrating surface, and are called "evanescent." For a circularly symmetric structure, however, *slightly* subsonic circumferential waves on the structure can still radiate efficiently to the farfield because as they begin to radiate radially they can become trace matched acoustically a short distance away from the circular surface. A general "rule of thumb"[20] for empirically fitting such a circumferential wave has been given as

$$n \approx ka + \epsilon, \tag{25}$$

where n is the circumferential harmonic such that $n\lambda=2\pi a$, a is the radius of the structure, k is the wavenumber in the fluid, and ϵ is a fitting parameter.

In Sec. V, we discussed the dependence of wave velocity on the radius of the cone, which is subsequently dependent on the axial location. Therefore, in this case, Eq. (25) assumes the particular form

$$n(z) \approx ka_0\left(1-\frac{z}{l_z}\right)+\epsilon. \tag{26}$$

In Fig. 4(c) [with regard to Eq. (23)], at 500 Hz, one can actually see at which point the wave velocities drop below the sonic line as a function of position. If we use this cutoff criteria in the inverse transform of the displacement and pressure for the 500 Hz case, the resulting surface fields are shown in Fig. 10. These are comprised of the $k_z=0$ wavenumber, coupled with various lower order circumferential harmonics such as $n=0,1,2,3,4$, whose presence is spatially dependent. Here, we see that the supersonic modes are actually "quantized" as a function of axial location. A comparison of the farfield radiation [Eq. (24)] due to the total fields (computed from the fields as shown in Fig. 3) and the radiation due to these supersonic components is shown in Figs. 11(a) and 11(b) for a projection in the x-y and the y-z planes, respectively. They obviously do not agree very well. If we implement Eq. (26) and perform a few iterations, we arrive at the subsequent farfield patterns shown in this figure. For a determined farfield convergence of the axial mode contributions contained within $m=\pm 20$, we vary the parameter ϵ for the values of 0, 2, and 4, respectively. At this latter value, the farfield patterns have converged to the total field to within a 5% total error. From the values of k_{ax} and $a(z)$ for our cone, this implies that there are within $n=\pm 8$ circumferential harmonics which contribute at the larger end of the cone, and within $n=\pm 6$ circumferential harmonics which contribute at the smaller end. Using these limits, the inverse transform for the radiating components are shown in Fig. 12. Therefore, using this transform method in conjunction with the Helmholtz integral provides a method for determining

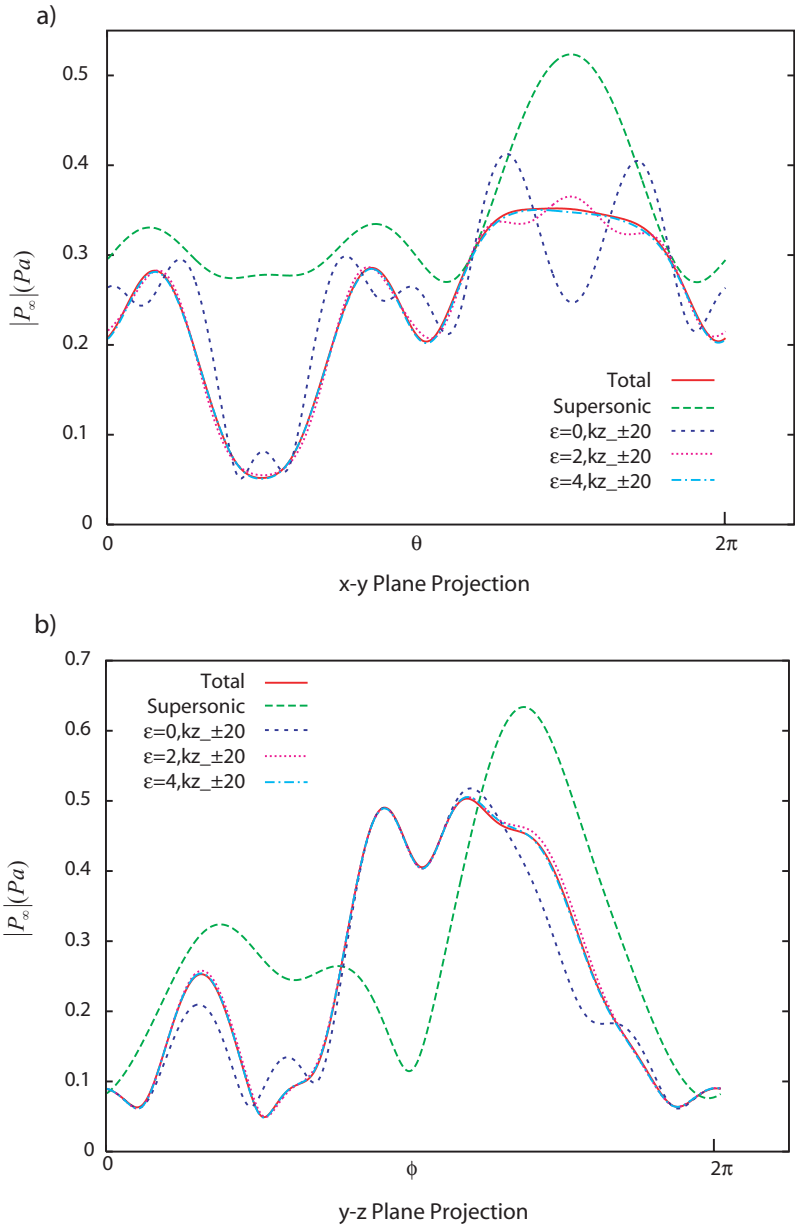


FIG. 11. Farfield radiation pattern [Eq. (24)] for projection in (a) the x-y plane and (b) in the y-z plane for the 500 Hz case. Red line indicates radiation due to total field [Figs. 3(a) and 3(c)]. Green dashed line indicates radiation due to supersonic components [Figs. 10(a) and 10(c)]. Dark blue dashed line indicates radiation due to modes included using Eq. (26) for $\epsilon=0$ and axial mode numbers within $m=\pm 20$. Red dashed line indicates radiation due to modes included using Eq. (26) for $\epsilon=2$ and axial mode numbers within $m=\pm 20$. Green dashed line indicates radiation due to modes included using Eq. (26) for $\epsilon=4$ and axial mode numbers within $m=\pm 20$.

the radiating components on C^1 surfaces for radiation analysis and/or feature space determination.

VII. CONCLUSIONS

In this paper, we developed and implemented a formulation for conformal Fourier wavenumber decompositions on C^1 surfaces. The generalized forward and inverse representations were derived from fundamental principles using differential geometry, and it was shown that the representations of the spatial Fourier transforms for simple structures such as a plate, a cylinder, and a sphere can be identically recovered with the appropriate parametrization.

The transform was then applied to a point driven, fluid loaded, conical structure, whose displacements and pressures were obtained using the finite element method. From the analytical representation for the kernels, the cone could be seen to support spatially dependent helical wave propagation, and that individual modes tended to slow as waves propagated toward the tapered end. Additionally, the inverse transform was utilized, in conjunction with the farfield Helmholtz radiation integral, to determine the radiating components from the total surface fields on the cone. As a result of this latter study, it was seen that many additional modes were required for farfield convergence than are represented by the so-called supersonic radiation condition. For more complex vibrating surfaces, it is anticipated that many more modes will be required for this farfield convergence and will provide an interesting metric for radiation analysis.

Future research will attempt to extend this method to more complex vibrating structures. Additionally, although we have provided a numerical demonstration that our forward and inverse transform representations form a transform pair in the case of a conical structure, many questions remain. In particular, although we have presented orthogonal parametrizations for simple surfaces, it is not yet known under what conditions a forward and inverse transform pair can exist. Therefore, a generalized analytical proof of the existence of this transform pair and the criteria which must be met for

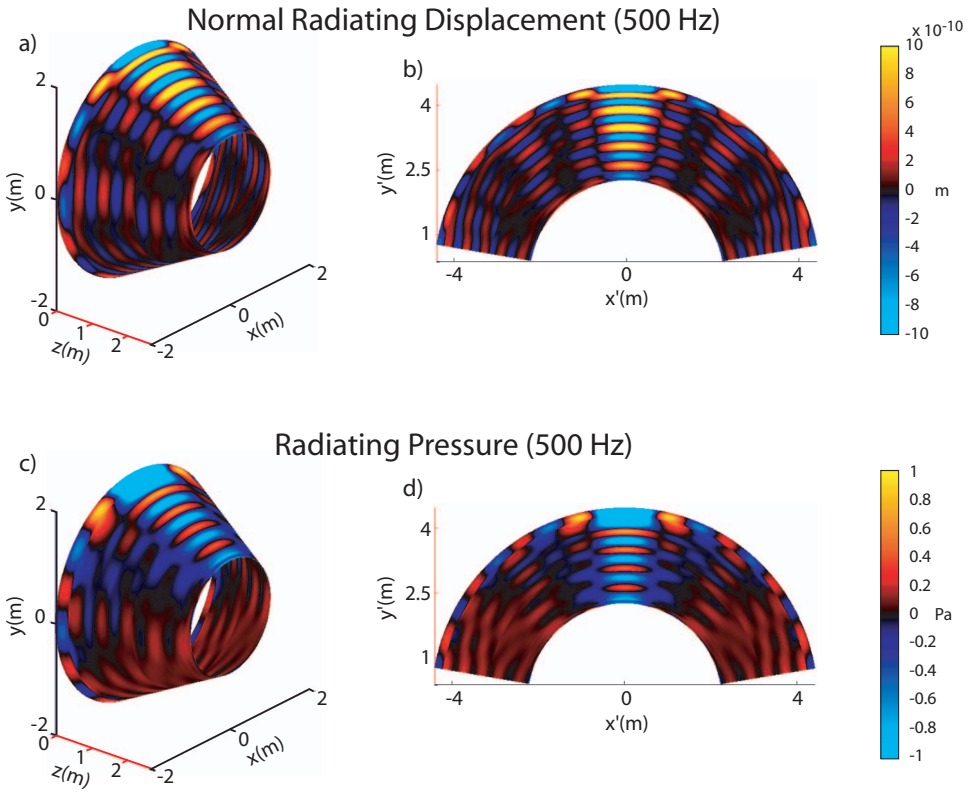


FIG. 12. Normal radiating displacement [(a) and (b)] and radiating pressure [(c) and (d)] for 500 Hz using the modal superposition for the converged case in Fig. 11, where $\epsilon=4$ and axial mode numbers within $m=\pm 20$.

them to exist is currently being investigated. While many challenges remain, this approach shows great promise in the study of the vibration of, and radiation from, C^1 surfaces, for which current methods of modal analysis are either insufficient or nonexistent.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Earl G. Williams and Dr. Louis R. Dragonette for their very useful discussions and insight provided throughout this research. This work was supported by the Office of Naval Research and the SERDP Program.

[1]E. G. Williams, *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography* (Academic, London, 1999).
[2]N. P. Valdivia and E. G. Williams, "Study of the comparison of the methods of equivalent sources and boundary element methods for near-field acoustic holography," J. Acoust. Soc. Am. **120**, 3694–3705 (2006).
[3]N. P. Valdivia, E. G. Williams, and P. C. Herdic, "Approximations of inverse boundary element methods with partial measurements of pressure field," J. Acoust. Soc. Am. **123**, 109–120 (2008).
[4]A. J. Romano, P. B. Abraham, P. J. Rossman, J. A. Bucaro, and R. L. Ehman, "Determination and analysis of guided wave propagation using magnetic resonance elastography," Magn. Reson. Med. **54**, 893–900 (2005).
[5]P. M. Morse and U. K. Ingard, *Theoretical Acoustics* (McGraw-Hill, New York, 1968), pp. 492–522.
[6]W. Ledermann and S. Vajda, *Combinatorics and Geometry* (Wiley, New York, 1985), Vol. **V**, Pt. B, Chap. 12, pp. 423–467.
[7]L. N. Trefethen and D. Bau III, *Numerical Linear Algebra* (SIAM, Philadelphia, 1997), pp. 48–51.
[8]W. Kaplan, *Advanced Calculus*, 2nd ed. (Addison-Wesley, Reading, MA, 1973), pp. 276–281.
[9]E. Kreyszig, *Differential Geometry* (Dover, New York, 1991), pp. 15–71.
[10]V. G. Godzevich, "Free oscillations of circular conical shells," *Theory of Shells and Plates*, Proceedings of the Fourth All-Union Conference, Yerevan, Armenian SSR, 24–31 Oct., 1962 [Yerevan, Izdatel'stvo Akademii Nauk Armianskoi SSR , pp. 378–382 (1964)].
[11]S. Dey and D. K. Datta, "A parallel HP-FEM infrastructure for three-dimensional structural acoustics," Int. J. Numer. Methods Eng. **68**, 583–603 (2006).
[12]S. Dey, J. J. Shirron, and L. S. Couchma, "Mid-frequency structural acoustic and vibration analysis in arbitrary, curved three-dimensional domains," Comput. Theor. Polym. Sci. **79**, 617–629 (2001).
[13]J. J. Shirron and S. Dey, "Acoustic infinite elements for non-separable geometries," Comput. Methods Appl. Mech. Eng. **191**, 4123–4139 (2002).
[14]U. S. Lindholm and W. C. L. Hu, "Nonsymmetric transverse vibrations of truncated conical shells," Technical Report No. 3, Southwest Research Institute, San Antonio, TX, March 1965.
[15]J. E. Goldberg and J. L. Bogdanoff, "On the calculation of the axisymmetric modes and frequencies of conical shells," J. Acoust. Soc. Am. **32**, 738–742 (1960).
[16]A. Leissa, *Vibration of Shells* (Acoustical Society of America, Melville, NY, 1993), pp. 331–401.
[17]Y. P. Guo, "Normal mode propagation on conical shells," J. Acoust. Soc. Am. **96**, 256–264 (1994).
[18]Y. P. Guo, "Fluid-loading effects on waves on conical shells," J. Acoust. Soc. Am. **97**, 1061–1066 (1995).
[19]D. Colton and R. Kress, *Inverse Acoustic and Electromagnetic Scattering Theory* (Springer-Verlag, Berlin, 1992), pp. 15–20.
[20]J. E. Hansen, *Spherical Near-Field Antenna Measurements* (Peter Peregrinus Ltd., London, 1988), pp. 16–18.

Entropy in statistical energy analysis

Alain Le Bot[a)]
Laboratoire de Tribologie et Dynamique des Systèmes, CNRS, École Centrale de Lyon, 36 Avenue Guy de Collongue, 69134 Ecully, France

(Received 21 May 2008; revised 2 October 2008; accepted 30 December 2008)

In this paper, the second principle of thermodynamics is discussed in the framework of statistical energy analysis (SEA). It is shown that the "vibrational entropy" and the "vibrational temperature" of sub-systems only depend on the vibrational energy and the number of resonant modes. A SEA system can be described as a thermodynamic system slightly out of equilibrium. In steady-state condition, the entropy exchanged with exterior by sources and dissipation exactly balances the production of entropy by irreversible processes at interface between SEA sub-systems.
 [DOI: 10.1121/1.3075613]

PACS number(s): 43.40.At, 43.55.Br, 43.20.Ks [RLW] Pages: 1473–1478

I. INTRODUCTION

Statistical energy analysis[1] (SEA) is a statistical theory of vibroacoustics. When modes are too numerous and when waves are uncorrelated, the vibrational energy tends to be uniformly distributed inside the system. This is the state of diffuse field[2] or equipartition of modal energy.[3] Furthermore, when two sub-systems with diffuse field are connected together, the power flow between them is proportional to the difference of their modal energies.[4]

It has been recognized for a long time[5] that this proportionality relationship for power flow is the foundation of an analogy with thermodynamics provided that the "vibrational temperature" is defined as the modal energy. The "heat" or vibrational energy then flows from hot sub-system to cool sub-system.

SEA is entirely based on the application of the energy balance that is the first principle of thermodynamics. And the natural question that arises is the following: What about the second principle of thermodynamics? This question was raised by Carcaterra in Ref. 6. In Ref. 7, it has been shown that the mechanical statistics approach leads to a strict definition of the vibrational temperature and gives an argument in favor of the proportionality relationship of power flow with the difference of modal energies.

This aim of this study is to explicit the notion of entropy in SEA for systems at equilibrium but also for systems in non-equilibrium, which is the only case of interest in SEA.

II. ENTROPY OF A SINGLE SEA SUB-SYSTEM

The first question to solve is the following: How can one define the entropy of a single SEA sub-system?

A. Thermodynamic entropy

Before giving a rigorous answer to this question, let us develop some simple arguments based on classical thermodynamics. Clausius' principle states that thermal energy always flows from high temperature to low temperature. In SEA, the net power exchanged by two sub-systems i and j is given by

$$P_{ij} = \omega \eta_{ij} n_i \left(\frac{E_i}{n_i} - \frac{E_j}{n_j} \right), \tag{1}$$

where E_i is the total vibrational energy contained in sub-system i, n_i is the modal density in mode per rad/s, ω is the central frequency, and η_{ij} is the so-called coupling loss factor in which all the physics of the problem at hand is included. It is clear that the relevant quantity, which governs the flow of vibrational energy, is the modal energy E_i/N_i where $N_i = n_i \Delta\omega$ is the number of modes contained in the frequency band of width $\Delta\omega$. The modal energy in SEA is therefore analogous to the temperature in thermodynamics and for this reason, it can be called vibrational temperature,

$$T \propto \frac{E}{N}. \tag{2}$$

In thermodynamics, at least two kinds of energy can be exchanged with a system: heat and work. Work is the energy carried by external forces. This is the case when a pressure p is applied to the boundary and leads to a change in volume dV, the work is $\delta W = -pdV$ or when the system of mass m is raised to height dh and therefore increases its potential energy by $\delta W = mgdh$. In all cases, it is tacitly assumed that the force is static or at least slowly varying. This means that the period of the variation is very large compared with the period of thermal vibrations. In vibroacoustics, it is irrelevant to consider the deformation of the sub-system or to take into account the gravitational potential energy. The only way to provide energy to sub-systems is by driving forces for structures and noise sources in rooms. The forces are then time-varying and the energy they provide must be considered as the equivalent of heat. The "vibrational heat" provided by a source of power P^i during dt is then

$$\delta Q = P^i dt. \tag{3}$$

The "vibrational entropy" can now be introduced in the same way Clausius did a century and half ago in thermodynamics. Entropy is a state function that depends on the ex-

a)Electronic mail: alain.le-bot@ec-lyon.fr

tensive variables of the sub-system. These are the total vibrational energy E and the number of modes N. For a reversible process, Clausius definition of entropy is

$$dS = \frac{\delta Q}{T}. \tag{4}$$

Since the vibrational sources do not provide work, $\delta W = 0$ and the first principle of thermodynamics reads

$$dE = \delta Q. \tag{5}$$

By substituting Eqs. (2) and (5) into Eq. (4), it yields

$$dS \propto N\frac{dE}{E}. \tag{6}$$

Finally, the variation of entropy between two states of energy E_2 and E_1 is therefore

$$\Delta S \propto N \log\left(\frac{E_2}{E_1}\right). \tag{7}$$

This relationship gives the vibrational entropy of SEA subsystems. It has been established for reversible processes. But, the entropy is a function of state and, therefore, the same relationship must also hold for irreversible processes.

It could be inferred from Eq. (7) that the absolute entropy $S(E,N)$ is $N \log E$. This expression is compatible with the variation of entropy derived here, but, it is inconsistent with the requirement that the entropy is an extensive quantity in classical thermodynamics. An extensive quantity must verify the equality $S(2E,2N)=2S(E,N)$, which is obviously not the case with $N \log E$.

Equation (7) has been derived by considering a small variation dE of internal energy. The number of modes N has been tacitly assumed to be constant during this transformation. Thus, if one fixes a reference energy E_0, the absolute entropy $S(E,N)=\Delta S(E,E_0)+S(E_0,N)$ is known from Eq. (7) apart from a function $S(E_0,N)$, which now only depends on the number of modes N. But, it is not possible to go beyond and, in particular, to fully derive the function $S(E,N)$ with Clausius' approach. This is what is done in Sec. II B by introducing the statistical entropy.

B. Statistical entropy

In statistical mechanics,[8] the microscopic properties of a system are known as soon as the Hamiltonian of the system is established. A SEA sub-system is a set of N modes of frequencies ω_i $i=1,\ldots,N$ in the band $[\omega-\Delta\omega/2,\omega+\Delta\omega/2]$. Since the modes are orthogonal, the vibrational energy of the sub-system can be decomposed on the modes. If q_i is the modal factor and $p_i=m_i\dot{q}_i$ is its conjugate momentum, the Hamiltonian of the sub-system is

$$H(q_1,\ldots,q_N,p_1,\ldots,p_N) = \sum_{i=1}^{N} \frac{m_i\omega_i^2}{2}q_i^2 + \frac{1}{2m_i}p_i^2, \tag{8}$$

where m_i is the modal mass of the mode i. For normalized modes, $m_i=1$ or m the mass of the sub-system depending on the choice of the norm.

In the phase space Γ defined by all positions and momenta $q_1,\ldots,q_N,p_1,\ldots,p_N$ of the sub-system, the hypersurface of constant energy $H(q_1,\ldots,q_N,p_1,\ldots,p_N)=E$ is an ellipsoid of dimension $2N-1$ with semi-axes $(2m_iE)^{1/2}$ and $(2E/m_i\omega_i^2)^{1/2}$. Its volume $V(E)$ is

$$V(E) = \int_{H\leq E} d\Gamma = \frac{(2\pi)^N}{N!\prod_{i=1}^{N}\omega_i}E^N. \tag{9}$$

In SEA, the set of modes is considered to be a random population of modes. The frequencies ω_i have a uniform probability density over the band $[\omega-\Delta\omega/2,\omega+\Delta\omega/2]$. The product $\Pi_i\omega_i$ can therefore be substituted by its expected value ω^N. The volume becomes

$$V(E) = \left(\frac{2\pi}{\omega}\right)^N\frac{E^N}{N!} \tag{10}$$

and the structure function $\Omega(E)$ is

$$\Omega(E) = \frac{dV}{dE} = \left(\frac{2\pi}{\omega}\right)^N\frac{E^{N-1}}{N-1!}. \tag{11}$$

All the information useful for a macroscopic description of the sub-system is contained in the structure function Ω.

The general method of statistical mechanics is to consider that the position of the system in Γ-space can never be known exactly. The only available information on the system is the probability density function of presence at any point in Γ-space. At this stage, it is common in statistical mechanics to distinguish two statistical ensembles. The microcanonical ensemble is the statistical distribution of positions in the Γ-space of an isolated system for which the energy is constant. And the canonical ensemble is the statistical distribution of a system in equilibrium with a thermal bath. Whatever the distribution of probability ρ_i of microstates (positions in Γ-space), the statistical entropy is always defined as

$$S = \sum_i -k\rho_i \log \rho_i, \tag{12}$$

where the sum runs over all microstates. This is the expectation of information $I_i=-k \log \rho_i$ associated with an event of probability ρ_i.

Indeed, the two statistical ensembles do not lead to the same probability distribution ρ_i. And the resulting entropy that is the mean loss of information of system whose the only macroscopic state is known is *a priori* not the same for both ensembles. But, this is a well-known result in statistical mechanics that the microcanonical entropy and the canonical entropy only differ with a negligible term.[8] For this reason, in the following, the SEA sub-system is assumed to be isolated and the microcanonical approach is adopted.

In the microcanonical ensemble, all the positions in the Γ-space around the surface $H(q_1,\ldots,q_N,p_1,\ldots,p_N)=E$ are equally probable. This is the so-called fundamental assumption of statistical mechanics. If the position of modes q_i and p_i is known with the uncertainties δq_i and δp_i such that $\delta q_i\delta p_i=h$ and the energy of the system is known with the

uncertainty δE, the number of microstates W is the volume of the shell of thickness δE and surface $\Omega(E)$ divided the volume of a microstate h^N,

$$W = \frac{\Omega(E)}{h^N}\delta E. \tag{13}$$

The entropy is then defined by substituting the uniform probability $\rho_i = 1/W$ in Eq. (12), which gives the Boltzmann's equation $S = k \log W$,

$$S = kN \log\frac{2\pi E}{\omega h} - k \log(N-1)\,! + k \log\frac{\delta E}{E}. \tag{14}$$

For a large population $N \simeq N-1$ and Stirling's approximation gives $\log N! \simeq N \log N - N$. Furthermore, the term $\log \delta E/E$ is usually discarded since it is related to the choice of the relative uncertainty on the energy, which is not a physical process. The final result is

$$S(E,N) = kN\left[1 + \log\left(\frac{2\pi}{h\omega}\frac{E}{N}\right)\right]. \tag{15}$$

This is the microcanonical entropy of a SEA sub-system. It can be remarked that this relationship well agrees with Eq. (7), which stems from the difference of $S(E_2,N) - S(E_1,N)$. Equation (15) gives the complete expression of the vibrational entropy of an isolated SEA sub-system.

The temperature is obtained as for any thermodynamic system with

$$\frac{1}{T} = \left(\frac{\partial S}{\partial E}\right)_N, \tag{16}$$

with the result

$$T = \frac{E}{kN}. \tag{17}$$

As it was expected from Eq. (1) giving the power flow between two SEA sub-systems, the vibrational temperature is well defined by the modal energy. This result was obtained from an analogy between SEA and thermodynamics. This is now a logical consequence of the expression of the entropy obtained from Boltzmann's definition. The same result has been derived in Ref. 7 with Khinchin's definition[9] of entropy.

III. SECOND PRINCIPLE OF THERMODYNAMICS IN SEA

The second principle of thermodynamics states that an isolated system (with no exchange of energy) always evolves in such a way that its entropy increases. To illustrate what is the second principle of thermodynamics in the particular framework of SEA, let us consider two isolated acoustical cavities. Isolated means that the cavities do not exchange energy and, in particular, that there is no noise source inside and that they do not dissipate the sound. As usual, it is also assumed that the thermal equilibrium is realized within the cavities, that is, the sound is diffuse. Therefore, these cavities are perfect reverberation chambers with infinite reverberation times; the sound pressure level is constant, but indeed, may have different values in the two cavities. The entropy of the acoustical cavities $i=1,2$ are $S(E_i,N_i)$ where S is given in Eq. (15). For an acoustical cavity, the modal density is

$$n_i = \frac{V_i\omega^2}{2\pi^2 c_0^3}, \tag{18}$$

where c_0 is the sound speed and V_i is the volume of the cavity.

Now, let us consider that the two rooms are adjacent and that they are separated by a small closed window. The problem is to observe the evolution of the coupled rooms after the window has been opened. By opening the aperture between the two cavities, the energy is mixed and, since there is no cause of dissipation, the global system has energy $E = E_1 + E_2$. The total volume is also $V = V_1 + V_2$ so that the modal density of the total system is

$$n = \frac{(V_1+V_2)\omega^2}{2\pi^2 c_0^3} = n_1 + n_2. \tag{19}$$

Thus, the number of modes of the global system $N = n\Delta\omega = (n_1+n_2)\Delta\omega$ is the sum of the numbers of modes of the two cavities $N = N_1 + N_2$.

When the equilibrium is reached, the entropy of the entire system becomes

$$S(E,N) = S(E_1+E_2, N_1+N_2). \tag{20}$$

The difference between the final entropy and the sum of initial entropy is therefore the entropy created during the mixing process.

$$\Delta S = S(E,N) - [S(E_1,N_1) + S(E_2,N_2)]. \tag{21}$$

By substituting Eq. (15), it yields

$$\Delta S = k(N_1+N_2)\log\frac{E_1+E_2}{N_1+N_2} - kN_1 \log\frac{E_1}{N_2} - kN_2 \log\frac{E_2}{N_2}. \tag{22}$$

This is the entropy created by mixing the energy of two acoustical cavities. It is expected to be non-negative.

To prove this statement, it is necessary to introduce the function $f(x,y) = y \log(x/y)$ and to rewrite the entropy of mixing ΔS,

$$\Delta S = kf\left(\frac{E_1+E_2}{2}, \frac{N_1+N_2}{2}\right) - \frac{k}{2}[f(E_1,N_1) + f(E_2,N_2)]. \tag{23}$$

The function f, defined for $x>0$ and $y>0$, has the following derivatives:

$$\frac{\partial^2 f}{\partial x^2} = -\frac{y}{x^2}, \quad \frac{\partial^2 f}{\partial x\,\partial y} = \frac{1}{x}, \quad \frac{\partial^2 f}{\partial y^2} = -\frac{1}{y}, \tag{24}$$

so that the quadratic form tangent to f is

$$q(X,Y) = -\frac{y}{x^2}X^2 + \frac{1}{x}XY - \frac{1}{y}Y^2 = -\left(\frac{\sqrt{y}}{x}X - \frac{1}{\sqrt{y}}Y\right)^2. \tag{25}$$

Since $q(X,Y) < 0$ for all X and Y, $-f$ is convex. The convexity of a function means that its value at the middle point of a

segment is lower than the mean of the values at the extremities. This reads

$$f\left(\frac{x_1+x_2}{2},\frac{y_1+y_2}{2}\right)-\frac{1}{2}f(x_1,y_1)-\frac{1}{2}f(x_2,y_2)>0. \quad (26)$$

This proves the inequality $\Delta S>0$. The entropy of mixing is therefore always non-negative. The evolution of an isolated system (which does not receive neither dissipates energy) always leads to an increasing of the entropy. This is the second principle of thermodynamics in the particular framework of SEA.

IV. ENTROPY PRODUCTION AT INTERFACE

Up to now, we have considered that sub-systems are isolated and that they have reached their equilibrium state, or, in other words, that the equipartition of energy among modes is verified. Indeed, this situation is an ideal case, which has been introduced only to obtain the explicit expression of the entropy of a single sub-system.

But in vibroacoustics, a sub-system can never been considered as isolated. For instance, in Sabine's theory of reverberation chambers, the room is first filled up with noise and, after the source has been switched off, the energy starts to decrease. The rate of dissipation gives the reverberation time of the room. This is therefore not an isolated room. In steady-state SEA, the sub-systems can receive vibrational energy and in the same time, they can dissipate the energy as well as they can exchange energy with neighbor sub-systems. SEA sub-systems are therefore not isolated systems.

A. Single sub-system

Let us first consider a single sub-system submitted to input power and dissipated power. This sub-system is assumed to be in steady-state condition, that is, all macroscopic quantities such as input power, dissipated power, vibrational energy, vibrational temperature, and entropy are constant in time. The sub-system is not in thermodynamic equilibrium in the meaning accepted in the first two sections since a flux of energy can flow through it. This is the simplest problem in non-equilibrium thermodynamics.

If the flux is not too important, the sub-system is in local equilibrium. This means that the relaxation processes inside the sub-system such as the mixing of rays are more rapid than the breakdown of equilibrium imposed by sources and dissipation. This assumption of local equilibrium is important since it allows to define a unique vibrational temperature for the sub-system.

The transfer of vibrational energy by sources and dissipation is therefore a reversible process. At any time,

$$dS=\frac{\delta Q^i+\delta Q^d}{T}, \quad (27)$$

where δQ^i is the vibrational heat supplied by sources to the sub-system and δQ^d is the vibrational heat extracted from the sub-system by dissipation. This net vibrational heat is the sum of the gain δQ^i and the loss δQ^d. Each term is the product of a power by the infinitesimal time dt,

$$\delta Q^i=P^i dt \quad (28)$$

for the gain term and

$$\delta Q^d=-P^d dt \quad (29)$$

for the loss term.

The vibrational entropy may therefore be separated into two terms: the entropy dS^i provided by sources and the entropy dS^d consumed by dissipation.

$$dS=dS^i+dS^d=\frac{P^i-P^d}{T}dt. \quad (30)$$

The first term dS^i is positive, and the sources provide heat to the sub-system and therefore warm it and increase its entropy. The second term dS^d is negative, and the dissipation extracts heat from the sub-system. This cooling process makes the entropy to decrease.

In steady-state condition, the power balance reduces to

$$P^i=P^d, \quad (31)$$

where $P^d=\eta\omega E^d$ is the power being dissipated. This simple equation is the SEA equation in the trivial case of a single sub-system. Substitution into Eq. (30) gives

$$\frac{dS}{dt}=0. \quad (32)$$

Thus, any SEA system composed of a single sub-system and submitted to sources and dissipation has a null production entropy.

B. Coupled sub-systems

Let us now consider a couple of two adjacent sub-systems. Once again, the sub-systems are assumed to be in local equilibrium allowing to define a vibrational temperature for all sub-systems. But, these vibrational temperatures T_1 and T_2 may be different. The conditions under which a local equilibrium is reached while a global equilibrium is not is discussed in Ref. 10.

By applying Eq. (4) to sub-system 1, it yields

$$dS_1=\frac{P_1^i-P_1^d-P_{12}}{T_1}dt \quad (33)$$

and for sub-system 2

$$dS_2=\frac{P_2^i-P_2^d+P_{12}}{T_2}dt. \quad (34)$$

Entropy is an extensive quantity. Therefore, for the entire system, the entropy dS is at any time the sum of the entropy of its parts dS_1+dS_2.

$$dS=\frac{P_1^i-P_1^d-P_{12}}{T_1}dt+\frac{P_2^i-P_2^d+P_{12}}{T_2}dt. \quad (35)$$

This sum may be re-combined into two terms, $dS=dS^{\mathrm{rev}}+dS^{\mathrm{irr}}$. The first term dS^{rev} is the sum of the net entropy provided by the exchange of vibrational heat with exterior (sources and dissipation).

$$dS^{\text{rev}} = \frac{P_1^i - P_1^d}{T_1}dt + \frac{P_2^i - P_2^d}{T_2}dt. \tag{36}$$

This entropy is supplied to the system by a reversible process since equality (4) applies. The second term,

$$dS^{\text{irr}} = P_{12}\left(\frac{1}{T_2} - \frac{1}{T_1}\right)dt, \tag{37}$$

is the entropy created during the exchange of vibrational heat between the two sub-systems. This term is not related to an exchange of heat with exterior. This is the entropy created inside the system by the mixing of vibrational energy. This entropy is therefore the entropy created by irreversible processes.

By substituting Eqs. (1) and (17)

$$\frac{dS^{\text{irr}}}{dt} = k\eta_{12}\omega N_1\left(\frac{E_1}{N_1} - \frac{E_2}{N_2}\right)\left(\frac{N_2}{E_2} - \frac{N_1}{E_1}\right). \tag{38}$$

This relationship gives the production rate of entropy at the interface between two adjacent SEA sub-systems. The right-hand side can be re-arranged,

$$\frac{dS^{\text{irr}}}{dt} = k\eta_{12}\omega N_1\frac{N_1N_2}{E_1E_2}\left(\frac{E_1}{N_1} - \frac{E_2}{N_2}\right)^2. \tag{39}$$

It is now clear that the right-hand side is always positive. Therefore,

$$\frac{dS^{\text{irr}}}{dt} \geq 0. \tag{40}$$

This result generalizes the second principle of thermodynamics to the case of non-equilibrium SEA sub-systems.

The procedure followed to obtain Eq. (38) is similar to the one adopted in thermodynamics of linear irreversible processes[11] except that the power flow has been substituted by the correct SEA power flow given in Eq. (1).

C. Entropy balance

The last question to examine is the following: How to write an entropy balance for the entire SEA system? Since the steady-state condition is assumed to apply, the entropy is constant in time and therefore the production of entropy dS/dt for the entire system must vanish.

Any sub-system i receives an infinitesimal entropy supplied by sources,

$$dS_i^i = \frac{P_i^i}{T_i}dt, \tag{41}$$

and also loses entropy by dissipation,

$$dS_i^d = -\frac{P_i^d}{T_i}dt. \tag{42}$$

At any interface, the entropy created by the irreversible process is

$$dS_{ij}^{\text{irr}} = P_{ij}\left(\frac{1}{T_j} - \frac{1}{T_i}\right)dt. \tag{43}$$

Now, consider the entire system. Since entropy is an extensive variable, the entropy of the whole system is the sum of the entropy exchanged with exterior by sub-systems and the entropy created at interfaces by irreversible processes.

$$dS = \sum_{i=1}^{N} dS_i^i + dS_i^d + \sum_{i>j} dS_{ij}^{\text{irr}}. \tag{44}$$

The last sum runs for $i>j$ since each interface must be counted only one time. By substituting Eqs. (41)–(43) into Eq. (44), it yields

$$\frac{dS}{dt} = \sum_{i=1}^{N} \frac{P_i^i - P_i^d}{T_i} + \sum_{i>j} P_{ij}\left(\frac{1}{T_j} - \frac{1}{T_i}\right). \tag{45}$$

By splitting the last sum and remarking that $P_{ij}=-P_{ji}$,

$$\frac{dS}{dt} = \sum_{i=1}^{N} \frac{1}{T_i}\left(P_i^i - P_i^d + \sum_{j\neq i} P_{ij}\right). \tag{46}$$

But the energy balance of any sub-system reads

$$P_i^i - P_i^d + \sum_{j\neq i} P_{ij} = 0. \tag{47}$$

Finally,

$$\frac{dS}{dt} = 0. \tag{48}$$

This is the expected result. It means that there is no production of entropy for the entire SEA system. But there is an exchange of entropy with exterior, which exactly balances the production of entropy by irreversible processes.

V. CONCLUSION

In this paper, it has been shown that the second principle of thermodynamics can be stated in the particular context of SEA. Explicit relationships have been derived for the vibrational entropy of sub-systems and the production of vibrational entropy by irreversible processes at interfaces of sub-systems. The mixing of vibrational energy of two sub-systems as well as the exchange of power flow between two sub-systems always lead to an increase in the entropy.

Beyond the work of Carcaterra[6] who demonstrated that the proportionality of power flow with the difference of modal energies can be derived from the statistical mechanics, this study highlights the importance of entropy in SEA. The thermodynamical analogy on which SEA is based could not be completed without the introduction of the second principle.

In SEA, the study of perfectly isolated sub-systems in the meaning of classical thermodynamics is of no interest. Systems of interest are always crossed by a flux of energy supplied by sources and lost in dissipation process. These thermodynamic systems are maintained in non-equilibrium condition by external constraints. But, the diffuse field assumption in SEA requires that all sub-systems remain in

equilibrium. From this point of view, SEA is therefore the thermodynamics of linear irreversible processes in vibrational systems.

ACKNOWLEDGMENTS

The author wishes to acknowledge the National Research Agency (ANR, France) to support the project DyVIn.

[1]R. Lyon and R. DeJong, *Theory and Application of Statistical Energy Analysis* (Butterworth-Heinemann, Boston, MA, 1995).
[2]J. Polack, "Modifying the chambers to play billiards, or the foundations of reverberation theory," Acustica **76**, 257–270 (1992).
[3]L. Maxit and J. Guyader, "Extension of sea model to subsystems with non-uniform modal energy distribution," J. Sound Vib. **265**, 337–358 (2003).
[4]R. Lyon and G. Maidanik, "Power flow between linearly coupled oscillators," J. Acoust. Soc. Am. **34**, 623–639 (1962).
[5]R. Lyon, "Fluctuation theory and (very) early statistical energy analysis," J. Acoust. Soc. Am. **113**, 2401–2403 (2003).
[6]A. Carcaterra, "An entropy approach to statistical energy analysis," in Inter-Noise 98, Christchurch, New-Zealand (1998).
[7]A. Carcaterra, "An entropy formulation for the analysis of energy flow between mechanical resonators," Mech. Syst. Signal Process. **16**, 905–920 (2002).
[8]W. Pauli, *Statistical Mechanics* (Dover, New York, 1973).
[9]A. Khinchin, *Mathematical Foundations of Statistical Mechanics* (Dover, New York, 1949).
[10]A. Le Bot, "Derivation of statistical energy analysis from radiative exchanges," J. Sound Vib. **300**, 763–779 (2007).
[11]I. Prigogine, *Introduction to Thermodynamics of Irreversible Processes* (Wiley, New York, 1962).

Verifying the attenuation of earplugs *in situ*: Method validation on human subjects including individualized numerical simulations

Annelies Bockstael[a)]
Department of Oto-Rhino-Laryngology and Department of Information Technology, Ghent University, De Pintelaan 185, 9000 Ghent, Belgium

Timothy Van Renterghem and Dick Botteldooren
Department of Information Technology, Ghent University, Sint-Pietersnieuwstraat 41, 9000 Ghent, Belgium

Wendy D'Haenens, Hannah Keppler, Leen Maes, Birgit Philips, Freya Swinnen, and Bart Vinck
Department of Oto-Rhino-Laryngology, Ghent University, De Pintelaan 185, 9000 Ghent, Belgium

(Received 25 August 2008; revised 4 November 2008; accepted 30 December 2008)

The microphone in real ear (MIRE) protocol allows the assessment of hearing protector's (HPD) attenuation *in situ* by measuring the difference between the sound pressure outside and inside the ear canal behind the HPD. Custom-made earplugs have been designed with an inner bore to insert the MIRE probe containing two microphones, the reference microphone measuring the sound pressure outside and the measurement microphone registering the sound pressure behind the HPD. Previous research on a head and torso simulator reveals a distinct difference, henceforth called transfer function, between the sound pressure at the MIRE measurement microphone and the sound pressure of interest at the eardrum. In the current study, similar measurements are carried out on humans with an extra microphone to measure the sound pressure at the eardrum. The resulting transfer functions confirm the global frequency dependency found earlier, but also show substantial variability between the ears with respect to the exact frequency and amplitude of the transfer functions' extrema. In addition, finite-difference time-domain numerical models of an ear canal with earplug are developed for each individual ear by including its specific geometrical parameters. This approach leads to a good resemblance between the simulations and their corresponding measurements. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3075603]

PACS number(s): 43.50.Hg, 43.66.Vt [BSF] Pages: 1479–1489

I. INTRODUCTION

Several studies clearly indicate that exposure to excessive noise is an important cause of hearing loss (Śliwinska-Kowalska and Kotylo, 2007). Furthermore, Tak and Calvert (2008) estimated that this hearing loss is not seldom induced by noise exposure during employment. Additionally, noise damages hearing in an irreversible way, to date irreparable by pharmacological intervention (Śliwinska-Kowalska and Kotylo, 2007).

Therefore, dealing with noise-induced hearing loss comes down to prevention by lowering the sound level at the worker's eardrum. In this regard, individual hearing protection is often the most feasible solution due to economical and practical constraints (Arezes and Miguel, 2002). The European Noise Directive (2003-10-EC) on exposure limit values also acknowledges the use of hearing protectors at the work floor by stipulating that the worker's effective exposure must take account of the attenuation provided by his hearing protectors.

However, it seems less advisable to calculate this exposure from the hearing protector's attenuation reported by the manufacturer, since the discrepancy between the attenuation measured in laboratory conditions and the actual protection offered to the individual user has been well-established (Franks, 2003; Witt, 2007).

This issue can be handled by measuring the attenuation of hearing protectors *in situ* (Witt, 2007; Casali *et al.*, 1995). The usability of any *in situ* measurement is not only determined by its accuracy but also by its rapidity and the required acoustical conditions (Casali *et al.*, 1995). Among the different alternatives, the microphone in real ear (MIRE) approach offers a quick and objective way to evaluate the attenuation (Berger, 2005).

MIRE testing may be carried out with one or two microphones. In the single microphone technique, the receiver is placed in the ear canal during separate, consecutive measurements with and without a hearing protector. Using two microphones, one is placed inside the ear canal underneath the hearing protector, and the other measures simultaneously the sound level outside the ear. Both methods have proved to be successful with earmuffs, but the application with earplugs often requires extra adaptations (Pääkkönen *et al.*,

[a)]Electronic mail: annelies.bockstael@ugent.be

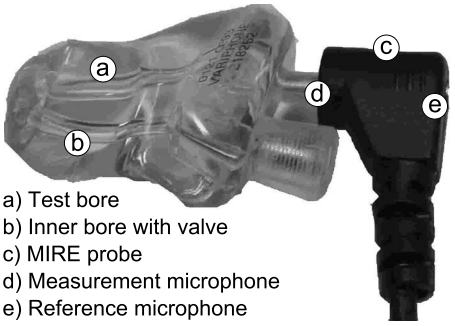


FIG. 1. Earplug with two inner bores; one to adjust the attenuation with an adjustable valve (b) and the other test bore (a) for insertion of the MIRE probe (c) with measurement (d) and reference (e) microphone. The measurement microphone measures the sound level in the ear canal behind the hearing protector whereas the reference microphone registers the incoming sound level.

2000). By contrast, Voix (2006) and others (Berger *et al.*, 2007) described a custom-made earplug with an inner bore that allows the insertion of a miniature microphone registering sound pressure levels inside the ear canal behind the hearing protector. In practice, this microphone is mounted in a probe that also contains a reference microphone measuring the sound pressure outside the ear canal (see Fig. 1).

Since this test design becomes more widespread (Berger *et al.*, 2007), a more thorough investigation of the underlying acoustical mechanisms is needed. A critical point in the accuracy of this measurement technique is the relationship, henceforth called transfer function, between the sound pressure registered by the MIRE measurement microphone and the sound pressure of interest at the eardrum. As literature already suggested (Hammershøi and Møller, 1996; Hellstrom and Axelsson, 1993), the sound pressure at the MIRE measurement microphone differs indeed manifestly from the level registered by a head and torso simulator's (HATS) microphone mimicking the human eardrum (Bockstael *et al.*, 2008).

Although the thus measured transfer functions appear stable and reproducible, it is unwise to generalize the results obtained on a HATS to human subjects without additional research. Despite the technological progress in the design of HATS's (Buck and Dancer, 2007), these devices are never an exact replica of the human body's acoustical features (Berger, 1986, 2005). Moreover, the intersubject variation might also be significant (Hammershøi and Møller, 1996).

In this paper, the transfer function previously obtained for the HATS is now also registered for humans by simultaneously measuring the sound pressure at the eardrum and at the MIRE measurement microphone for a sample of volunteers.

Furthermore, the influence of the morphology of an individual's hearing protector and ear canal on the transfer function will be assessed using finite-difference time-domain (FDTD) simulations of the outer ear canal occluded by an earplug with inner bore, analogous to the earlier developed numerical simulations for the HATS (Bockstael *et al.*, 2008). It will be verified whether numerical FDTD simulations representing a specific individual's occluded ear are able to provide satisfying correspondence with measured transfer functions. Furthermore, the possible gain in likeness will be assessed when using personalized numerical models versus the numerical model of the HATS. Finally, the influence of different dimensional parameters on the accuracy of the simulations will be addressed.

II. GENERAL METHODOLOGY

This study has been approved by the Ethical Committee of the University Hospital of Ghent (Belgium).

A. Subjects

Nineteen subjects, eleven female and eight male, between 18 and 48 years old are randomly selected from students and employees at Ghent University. Fifteen of them were inexperienced with respect to hearing protectors; none of them has a history of otological problems. All participated voluntarily and signed an informed consent.

Before the actual measurements take place, otoscopy is carried out and the hearing of the volunteers is tested with pure-tone audiometry performed by a qualified audiologist in accordance with the modified Hughson-Westlake technique (ISO 8253-1, 1989). This takes place in a sound-proof audiometric cabin at the Audiology Center of the Department of Oto-rhino-laryngology of Ghent University Hospital, using a regularly calibrated Orbiter 922 audiometer. All subjects have normal pure-tone hearing thresholds better than 20 dB HL for octave frequencies between 125 and 8000 Hz (Harrel, 2002).

Furthermore, the status of the middle ear is verified by carrying out tympanometry at 226 Hz with a ZODIAC 901 tympanometer of Madsen Electronics. All ears show normal patterns, suggesting that the reflection of sound at the eardrum is not perceptibly influenced by abnormalities of the eardrum or middle ear (Fowler and Shanks, 2002).

B. Hearing protector devices

The custom-made earplugs used in this study are made of hypo-allergic acrylic and have two inner bores (see Fig. 1), the test bore and a second bore with an adjustable valve determining the attenuation. Measurements by the Germane Institute for Occupational Safety and Health (BGIA) following the ISO standard (ISO 4869-2, 1994) reveal that an almost open valve can offer 25 dB attenuation [single number rating (SNR)] whereas a completely closed valve reaches a SNR of 31 dB. The test bore is closed in normal wearing conditions, but it allows the insertion of a MIRE probe to measure the hearing protector's attenuation.

Custom-made earplugs are manufactured for each participant on the basis of an impression of the ear canal taken by a well-trained audiologist. The fitting of each protector is tested using an attenuation control unit (ACU). Via the test bore, this device builds up a pressure of 10 mbars in the residual part of the ear canal behind the hearing protector. If this pressure holds stable for 2 s, the fitting of the hearing protector is considered satisfactory. If not, a new impression of the ear canal has to be made for a new hearing protector. This procedure is repeated until each participant has a pair of perfectly fitting hearing protectors.

III. MEASUREMENT METHODOLOGY

A. Material and measurement setup

All measurements are carried out in an anechoic room to prevent disturbances from background noise and reflected sound. They are performed with a laptop PC connected to a four input channel data acquisition front-end of Brüel & Kjær (type 3560-C) linking all sound equipment. The signals from the microphones are registered by the Brüel & Kjær's PULSE LABSHOP software version 7.0. Linear averaging is carried out over 3000 samples and overloads are rejected. In the frequency range between 0 Hz and 10 kHz, the responses are spectrally analyzed using fast fourrier transform (FFT) (6400 points).

The test stimulus is low pass filtered pink noise with a cut-off frequency of 12.8 kHz generated on the PC using PULSE LABSHOP software. The signal is then transmitted via the front-end and a Pioneer A-607 R direct energy metal-oxide-semiconductor (MOS) amplifier through a Renkus-Heinz (model CM 81) loudspeaker. The quality of the sound generation system is not critical since the sound signal will be calibrated out in all measurements. The signal is set sufficiently loud to ensure that the lower working sound limit of each microphone is exceeded. On the other hand, it is also verified that the test signals are not harmful to the participants' hearing.

The MIRE measurements are performed with a probe containing two Knowles low noise FG-3652 microphones. The so-called measurement microphone registers via the earplug's test bore the sound pressure level in the residual part of the ear canal behind the hearing protector whereas the reference microphone captures the incoming sound level. The difference between both levels indicates the noise reduction offered by the earplug. Previous measurements have shown that the attenuation of the hearing protector is not altered by the insertion of the probe (Bockstael *et al.*, 2008).

Further, a GN ReSound Aurical microphone is used to measure the sound pressure at the eardrum. This device consists of a flexible silicone tube (outer diameter of 0.85 mm) to be inserted in the ear canal, connected to an ear piece with microphone. According to the Aurical's manual, the tube is inserted 31 mm in the ear canal for male subjects and 28 mm for female participants.

Finally, two prepolarized free-field 1/2 in. microphones type 4189 (Brüel & Kjær) with preamplifier (type 2669C, Brüel & Kjær) are used to calibrate unwanted influences out of the measured transfer functions (see Sec. III B).

To perform the measurements, the test subject with his test ear toward the loudspeaker is seated diagonally in front of the loudspeaker at a distance of 1.61 m and one free-field microphone is placed symmetrically at the other side. The height of this free-field microphone is set equal to the height of the test ear. For this test, the sound incident direction is less important because Hammershøi and Møller (1996) showed that the transmission of sound to the eardrum from any point between the eardrum and a point 6 mm outside the ear canal can be considered directional independent.

B. Calibration and signal processing

The transfer functions calculated from these measurements should be absolutely independent of the test signal, the test space, and the microphones' characteristics. To fulfill these conditions, several calibration steps are carried out.

The first step takes place while the measurements are performed. The test setup described previously allows to calculate the frequency response function between, on the one hand, the MIRE measurement microphone, the MIRE reference microphone, and the Aurical, and on the other hand, the free-field microphone, respectively, $H_{mf}^{(1)}$, $H_{rf}^{(1)}$, and $H_{af}^{(1)}$. The equation for the frequency response function can in general be written as

$$H_{xy} = \sqrt{\frac{G_{xy}(k)}{G_{xx}(k)}\frac{G_{yy}(k)}{G_{xy}^{*}(k)}}, \tag{1}$$

where H_{xy} is the frequency response, $G_{xx}(k)$ and $G_{yy}(k)$ are the autospectra, $G_{xy}(k)$ is the cross-spectrum, and $G_{xy}^{*}(k)$ is its complex conjugate.

Ideally, the reference microphone would have been placed very close to the test subject to ensure that all the microphones receive the same signal. However, the signal of the free-field microphone was heavily distorted by sound reflection from the test subject's body and therefore it had to be placed at the opposite direction.

To compensate for a possible inhomogeneous sound distribution across the anechoic room, a second calibration step is carried out. The test subject is replaced by the second free-field microphone at the same height as the subject's test ear. The frequency response function [Eq. (1)] is then calculated between the second free-field microphone and the first one, the latter placed at its original position. With this transfer function H_{Ff}^{c}, the measured transfer functions $H_{xf}^{(1)}$ can be calibrated using the following expression:

$$H_{xF}^{(2)} = \frac{H_{xf}^{(1)}}{H_{Ff}^{c}}, \tag{2}$$

with x being, respectively, m, r, or a. F stands for the second reference microphone and f for the first one.

Further, the influence of the different microphone characteristics has to be accounted for. For the MIRE measurement microphone, the MIRE reference microphone and the Aurical, the frequency response function is measured by subsequently closely mounting these microphone with the second free-field microphone placed straight in front of the loudspeaker. This procedure yields to a calibration function over the frequency range of interest for each microphone, namely H_{xF}^{c}, with x being, respectively, m, r, or a. With this function the measurements can be corrected yielding to

$$H_{xF}^{(3)} = \frac{H_{xF}^{(2)}}{H_{xF}^{c}}. \tag{3}$$

Measurements of the H_{xF}^{c} frequency response are repeated regularly during the entire test period. The nearly identical H_{xF}^{c} responses measured on different test days indi-

cate that the characteristics of the different microphones are very stable across time and different environmental conditions.

Finally, the transfer function between the MIRE measurement microphone and the Aurical (H_{ma}) on the one hand and between the MIRE measurement microphone and the MIRE reference microphone (H_{mr}) on the other can be calculated, namely,

$$H_{ma} = \frac{H_{mF}^{(3)}}{H_{aF}^{(3)}} \tag{4}$$

and

$$H_{mr} = \frac{H_{mF}^{(3)}}{H_{rF}^{(3)}}. \tag{5}$$

C. Measurement procedure

At the beginning of each test day, the microphones are calibrated using a 4228 pistonphone from Brül & Kjær. Otoscopy and 226 Hz tympanometry are carried out for each test subject to ensure that no indication for outer or middle ear abnormalities is present. Next, the ability of the test subjects to fit his hearing protector correctly is checked with the ACU. Finally, the hearing protector is removed from the ear and the subject is seated with his test ear toward the loudspeaker as described previously.

The Aurical is inserted at the appropriate depth, 28 mm for female and 31 mm for male participants, to measure the sound pressure at the eardrum for the unoccluded ear canal. Then, the earplug, once with open, once with closed valve, is placed by the test subject. The MIRE probe is slided at a fixed depth into the test bore by the investigator. The position of the probe is visually inspected and contact with the pinna is avoided as much as possible. In addition, the responses of both MIRE microphones are also registered with the Aurical removed from the ear canal. The order of the different measurements is randomized across subjects and the earplugs are fitted carefully for each test. The investigator always leaves the anechoic room between the different measurement steps. After all tests are completed for one ear, the Aurical's tube is disinfected.

It is worthwhile realizing that the flexible tube of the Aurical breaks the seal between hearing protector and ear canal and hence might lower the earplug's attenuation. This is not problematic since the focus of this study lies on the transfer function between sound pressure at the MIRE measurement microphone and the sound pressure at the eardrum, which appears to be independent of the earplug's attenuation (Bockstael *et al.*, 2008).

IV. NUMERICAL SIMULATION METHODOLOGY

The sound pressure distribution in an ear canal occluded by an earplug with two inner bores is numerically simulated using the FDTD technique. The model used for the individualized simulations is based on the model designed previously for the HATS. The general features of the simulations are summarized in Sec. IV A, a more detail description can be found in Bockstael *et al.* (2008).

A. General characteristics

A key factor of the numerical FDTD simulation is that both pressure p and particle velocity $\mathbf{u}$ are discretized in Cartesian grids. These grids are staggered by shifting the grid for discretizing $\mathbf{u}_\alpha$ over half of a grid step, $d\alpha/2$, in direction α with respect to the grid chosen for discretizing p. In time, staggering is obtained by calculating p at $t=ldt$ and $\mathbf{u}$ at $t=\left(l+\frac{1}{2}\right)dt$. The resulting equations

$$u_\alpha^{l+1/2}\left(\alpha+\frac{1}{2}\right) = u_\alpha^{l-1/2}\left(\alpha+\frac{1}{2}\right) - \frac{dt}{\rho_0 d\alpha}\{p^l(\alpha+1) - p^l\}, \tag{6}$$

$$p^{l+1} = p^l - \sum_{\beta=x,y,z} \frac{\rho_0 c^2 dt}{d\beta} \cdot \left\{u_\beta^{l+1/2}\left(\beta+\frac{1}{2}\right) - u_\beta^{l+1/2}\left(\beta-\frac{1}{2}\right)\right\}, \tag{7}$$

with c the speed of sound and ρ_0 the density of air, allow to step in time replacing old values by newly calculated ones without much memory overhead. The brief notation $(\alpha+q)$ is used to indicate that the value is taken at a point shifted by q spatial steps $d\alpha$ in the α-direction with respect to the reference location referred to by indices (i,j,k). The first equation is repeated for $\alpha=x$, y, and z, where x and y are defined along the cross-section of the ear canal and z represents the longitudinal axis.

In the numerical model of the occluded ear, two points of interest are defined where the sound pressure is registered, namely, in front of the eardrum and at the end of the test bore where in reality the MIRE measurement microphone is placed.

The acoustically important features of the different materials in the model can be introduced by boundary impedance of the form

$$Z = j\omega Z_1 + Z_0 + \frac{Z_{-1}}{j\omega}, \tag{8}$$

which can be easily implemented in FDTD (Botteldooren, 1995). The impedance of the hearing protector's material is based on data from Hillström *et al.* (2000, 2003) and the impedance of the MIRE measurement microphone terminating the test bore is derived from Brüel & Kjær (1996). The impedance of the ear canal is based on the impedance of bone (Wit *et al.*, 1987) because the ear canal's cartilaginous part is filled by the hearing protector and hence relevant sound propagation occurs mainly in the ossicular part. For the influence of the middle and inner ear, the impedance at the eardrum is calculated from the one-dimensional circuit model by Kringlebotn (1988) because of the close resemblance with experimental results (Farmer-Fedor and Rabbitt, 2002; Voss and Allen, 1994) and in accordance to Hiselius (2004). Finally, the surface impedance of the earplug terminating the residual part of ear canal between hearing protec-

tor and eardrum is not simply set equal to material impedance (Hiselius, 2005), but measured with an impedance tube using the two microphone transfer function according to the ISO standard (ISO 10534-2, 1998).

In addition, extra attention has been devoted to the sound field in the earplug's two inner bores. Since these bores are very narrow, the effects of viscosity and heat conductivity become potentially important and are therefore included in the model.

Eardrum impedance, earplug impedance, and viscothermal boundary layers are implemented in FDTD using digital filters (Bockstael *et al.*, 2008).

B. Individual characteristics

Some characteristics of the hearing protector and the outer ear clearly depend on the individual under study, in contrast with the features described in Sec. IV A. It will be verified whether this mostly geometrical variation can explain and predict the interindividual variation of the transfer functions. Therefore, the most striking features of the hearing protector that are thought to influence sound propagation are accurately measured for each earplug and included in the simulations with a 0.35 mm gridcell size.

First, the length of the test bore and second bore determining the attenuation is measured with a caliper accurate up to 0.01 mm.

Second, the end of the earplug toward the eardrum has a particular shape for each individual. This tip of the hearing protector is not flat, but forms a very small pit in which the two inner bores end. The cross-section of this pit is in general more or less elliptic, but the length of the major and minor axis varies, as does the pit's depth. Moreover, the distance between the two bores' terminus differs between subjects, just like their distance to the pit's edge. Finally, the width of the acrylic rim around the pit appears to be typical for each hearing protector. To determine all relevant dimensions of the earplug's tip, measurements are carried out with the coordinate-measurement machine VM-250 Nexiv, manufactured by Nikon and accurate up to 0.1 μm.

The residual part of the ear canal between hearing protector and eardrum is for all participants modeled as a straight tube with uniform cross-section. The length of this tube is the difference between the length of the hearing protector and the length of the unoccluded ear canal. The former is measured with the caliper, and the latter is estimated from the first maximum of the frequency response of the unoccluded ear (Pickles, 1988). Finally, the diameter of the ear canal is based on the earplug's diameter.

V. STATISTICAL ANALYSIS

The results from the measurements and the simulations are explored with different statistical techniques which are discussed in detail in Sec. VI.

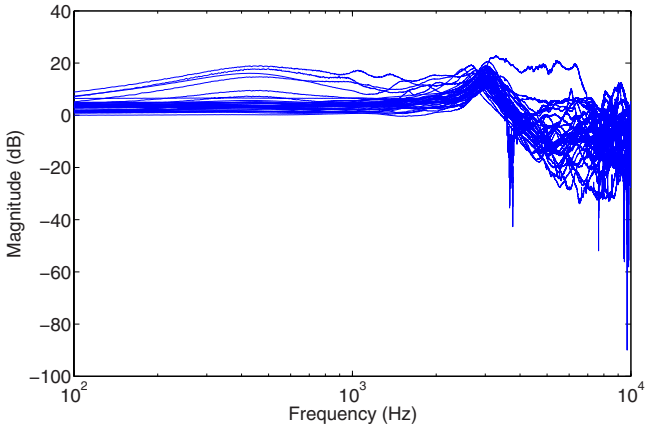


FIG. 2. (Color online) Magnitude of all measured transfer functions between the MIRE measurement microphone and the Aurical microphone for earplugs with open valve.

VI. RESULTS

A. Measured transfer functions

Figure 2 shows for all ears the transfer function with open valve between the MIRE measurement microphone and the Aurical (H_{ma}) which is very similar to the measurements with closed valve. A positive function value implies that the MIRE measurement microphone registers a higher sound pressure level than the Aurical, obviously the opposite is true for a negative value.

It becomes clear that the majority of all transfer functions has the same global form with a distinct maximum between 2500 and 3500 Hz and multiple minima above 4500 Hz, with often the most distinct minimum between 4500 and 6500 Hz. This morphology is in accordance with the transfer function measured on the HATS (Bockstael *et al.*, 2008).

Combining the results from both studies leads to the conclusion that the first maximum in Fig. 2 is most probably caused by resonance in the test bore, picked up by the MIRE measurement microphone, but not by the Aurical. Additionally, the most distinct minimum is most likely due to resonances in the residual part of the ear canal behind the hearing protector, registered by the Aurical but not by the MIRE measurement microphone.

Despite this common global shape of most transfer functions, the intersubject variability appears to be substantial with respect to the exact frequency and amplitude of the extrema.

Within one ear, the measurements with open and closed valves resemble each other closely for most subjects and only minor shifts are seen with respect to the frequencies at which the extrema occur, see, for example, Fig. 3. Minimal deviations in the first maximum's frequency might be due to the position of the MIRE probe in the test bore. Although the probe is as accurately as possible inserted at a fixed depth, minor differences in the exact position cannot be excluded. In a similar way, the position of the earplug in the ear canal might also slightly differ between measurements, causing differences in the exact frequency of the minima above 4500 Hz.

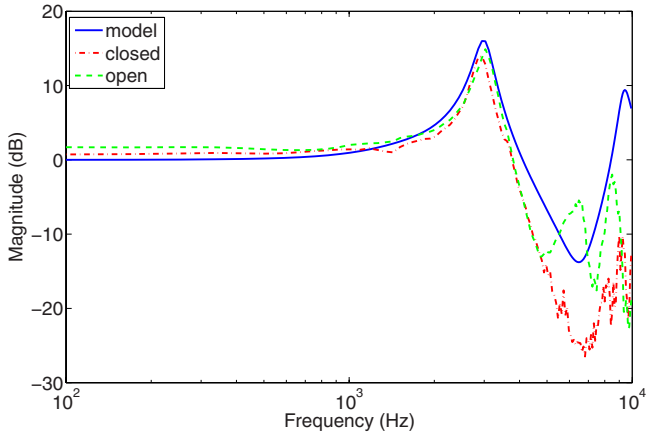


FIG. 3. (Color online) Example of similarity between simulation and measurements for one particular ear: magnitude of the measured transfer functions with open ("open") and closed ("closed") valve and magnitude of the simulated transfer function ('model').

However, for some participants, the amplitude of the transfer functions with closed valve tends to be more negative than with open valve in the higher frequency region (Fig. 3). Previous results suggest that this difference is not a true feature of the transfer functions for different positions of the valve since they appear to be independent of the applied attenuation in HATS measurements (Bockstael *et al.*, 2008). These results also reveal that the MIRE measurement microphone cannot be held responsible for the observed deviation since similar aberrations have never systematically occurred when measuring with the HATS.

Hence, the Aurical microphone seems to register in proportion a higher sound pressure level in the condition with closed valve to the measurements with open valve. This might be at least partially caused by sound pressure reaching the Aurical microphone unwantedly through the walls of the flexible tube. This sound level becomes potentially important in the conditions where the sound pressure behind the hearing protector is very low, i.e., for earplugs with closed valve. This hypothesis is supported by the fact that the effect is merely seen in the higher frequency range where the attenuation of a passive hearing protector reaches a maximal value. Therefore, the transfer functions measured with open valve might be more reliable in the higher frequencies.

To sum up, the close resemblance between the measurements with open and closed valves suggests that the transfer functions for a particular ear can be registered in consistent and hence reliable way. Nevertheless, the results also suggest that errors up to 10 dB can be expected when predicting an individual's transfer function from the mean transfer function measured on a sample of test subjects. More detailed prediction models, for instance, in the form numerical FDTD simulations, are clearly needed.

B. Numerical FDTD simulations

For the ears which measured transfer functions follow the global trend, the simulated transfer functions are very similar to the measured ones, an example is given in Fig. 3. The frequency dependence at the lower frequencies is very well predicted. As for the amplitude, all the numerical simulations have a constant amplitude of 0 dB whereas most measurements reach constant values between 0 and 5 dB. It is experimentally verified that bending the silicone tube can indeed lower the response of the Aurical microphone up to 5 dB in the lower frequency range. Flexures of the tube could not be avoided due to the relatively difficult positioning of the tube in an ear canal occluded by an earplug, but the influence of the bends can be clearly identified and hence the difference between the simulated and measured transfer function below 1500 Hz is not considered critical.

Further, the frequency and amplitude of the first maximum are very well approached by the model, as is the frequency of the most distinct minimum between 4500 and 6500 Hz. However, model and measurements tend to differ in the exact amplitude of this minimum. For frequencies above 6500 Hz, the numerical model and the measurements still show resembling frequency dependence but the resemblance is decreased compared to the frequency region below 6500 Hz. These findings clearly confirm the results obtained with the HATS (Bockstael *et al.*, 2008).

By contrast, for the ears for which the measured transfer functions clearly differ from the average in shape, the numerical simulations are unable to predict the measured frequency dependence. Hence, the question arises whether the obtained transfer functions are reliable and the numerical simulations produce a wrong forecast, or, by contrast, whether the measured transfer functions are incorrect and therefore cannot be approached by the numerical model.

Figure 2 shows that all anomalous transfer functions have a more positive difference between the MIRE measurement microphone and the Aurical, meaning either that the MIRE measurement microphone registers a higher sound pressure level or that the Aurical receives a lower level.

Following the first hypothesis, the attenuation derived from the transfer function between the two MIRE microphones (H_{mr}) is expected to be lower for the ears with abnormal results. However, this is not the case, the ears with normal transfer functions appear to have a systematically lower attenuation, especially for the frequencies below 1500 Hz (see Fig. 4).

To verify whether the observed difference in attenuation is only due to random error in this sample or whether it can be generalized, a two-way analysis of variance ($\alpha=0.05$) is carried out with ear (left or right) and difference from the average transfer function in shape (low or high) as fixed factors and the perceived attenuation below 1500 Hz as dependent variable. Before the results are interpreted, the aptness of the model is checked by looking at the standardized and studentized residuals (Kutner *et al.*, 2004). No interaction effect between the factors "ear" and "difference" ($P=0.14$) is found; hence the main effects may be interpreted. No significant difference is found for the attenuation of left and right ears ($P=0.41$). However, the mean attenuation is clearly higher in the ears with a distinctly different transfer function ($P<0.001$).

From this analysis, it becomes clear that the irregular transfer functions are not caused by an excessive registration of the sound pressure by the MIRE measurement microphone but by an underestimation of the sound level by the

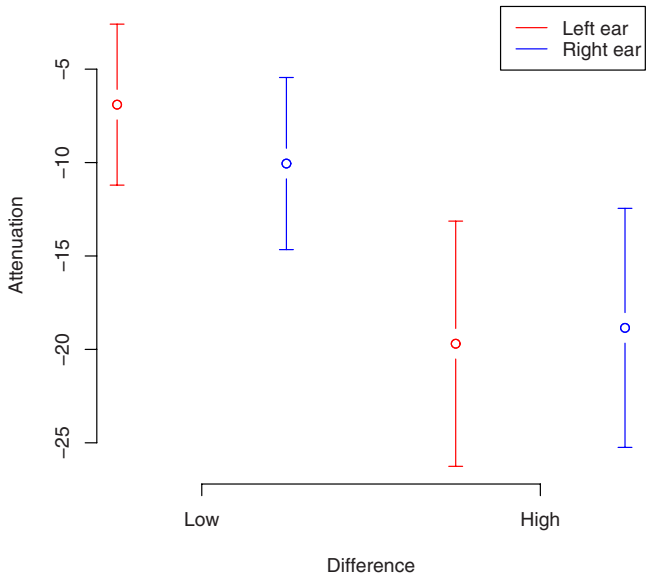


FIG. 4. (Color online) Attenuation below 1500 Hz (mean value and standard deviation) for ears with a high ("high") and low difference in shape ("low") from the average transfer function.

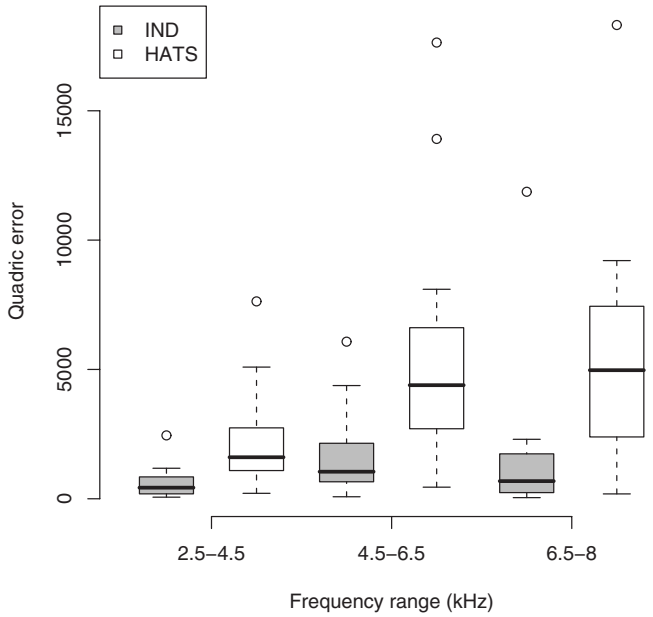


FIG. 5. Quadric error for the right ears between the measured transfer function with open valve and, respectively, the individualized model ("IND") and the model based on the HATS ("HATS"), calculated for the frequency region from 2.5 to 4.5 kHz, from 4.5 to 6.5 kHz, and 6.5 to 8 kHz. The boxes span the middle half of the ordered observations and the thick black lines inside represent the median. The whiskers extend to the most extreme data point which is no more than 1.5 times the interquartile range from the boxes. The circles represent data points that fall outside these limits.

Aurical. In the frequency range below 1500 Hz, the attenuation is the most sensitive to leakage because lower frequencies are more easily transmitted by the air column between hearing protector and ear canal (Apfel, 1997). Hence, the difference in attenuation suggests that the flexible tube of the Aurical causes less leakage for ears with deviant transfer function than for ears with regular ones. This leads to the assumption that the Aurical's tube must be more or less squeezed in the former ears, hindering the transfer of sound to the Aurical's microphone, decreasing the finally registered sound pressure level.

The above reasoning yields to the conclusion that the aberrant transfer functions do not reflect the true differences in sound pressure between the MIRE measurement microphone and the eardrum, but are caused by measurement error. Hence, these results will not be taken into account for further analysis.

C. Aptness of an individualized model

As described in Sec. IV, the numerical simulations are on the one hand based on general characteristics of an ear canal occluded by an earplug (Sec. IV A) and on the other on a set of parameters unique for each ear (Sec. IV B). In the current section, the aptness of an individualized model will be studied.

For this analysis, the quadratic error is calculated by squaring the difference between the model and the corresponding measurements. Further, this difference is summed for the four frequency regions that can be clearly distinguished; the lower frequency region between 100 and 1500 Hz where the function has a merely constant value, the frequency region between 2500 and 4500 Hz which encloses the first maximum, the region between 4500 and 6500 Hz where the most distinct minimum is seen, and finally the frequencies between 6500 and 8000 Hz.

Since the acoustical features of the HATS mimic the characteristics of an average human body, it might be possible to approach the transfer function of an individual with numerical simulations based on the acoustical and geometrical characteristics of a HATS. To find out whether the individualization of these geometrical features does enhance the resemblance between measured and predicted transfer functions, the quadric error between the measurements and, respectively, the individualized model and the model of the HATS is compared. This is done by separate pairwise tests for the different ears and the different positions of the valves. Because of the relatively small sample sizes (less than 20), normality tests might be less reliable and hence it is not always possible to state with sufficient accuracy that the conditions for the paired student t-test are fulfilled. Therefore, the non-parametric Wilcoxon signed-rank test is also performed (Moore and McCabe, 1999). For both statistical test, α is set at 0.05.

These analyses are not carried out for the lowest frequency range because the individualized models and the model of the HATS have exactly the same constant amplitude around 0 dB.

For the four combinations of right and left test ear with open and closed valves, the quadric error based on HATS model seems to exceed the quadric error calculated from the individualized model. This is illustrated by Fig. 5, depicting the quadric error for the right ear with open valve.

From Table I it can be seen that both the parametric and non-parametric analyses consequently yield to a statistically significant higher quadric error for the majority of the test situations when the model of the HATS is used ($P<0.05$).

TABLE I. For the different test situations [left ear with closed valve (L.C.), left ear with open valve (L.O.), right ear with closed valve (R.C.), and right ear with open valve (R.O.)], the quadric error between the measurements and the corresponding individualized model is statistically compared to the quadric error between measurements and the HATS model. The *P*-values from the paired student t-test (upper number) and from the Wilcoxon signed-rank test (lower number) are tabulated, as is the number of data points (between brackets).

	2.5–4.5 kHz	4.5–6.5 kHz	6.5–8 kHz
L.C. (13)	0.0003	0.09	0.7
	0.0004	0.05	0.3
L.C. (15)	0.002	0.0003	0.04
	6×10^{-05}	0.0001	0.04
R.C. (15)	0.0001	0.01	0.01
	0.0002	0.008	0.02
R.O. (16)	0.001	0.002	0.0004
	3×10^{-05}	9×10^{-05}	0.001

Only the left ears with closed valve fail to give statistically significant results in the two higher frequency regions ($P>0.05$), possibly due to the smaller amount of data in this group (Moore and McCabe, 1999). This analysis yields to the conclusion that individualized numerical simulations will in general lead to a more accurate prediction of the transfer function between the MIRE measurement microphone and the sound pressure of interest at the eardrum.

Apart from this conclusion, the question also arises whether the general characteristics of the model are suitable for both men and women. In particular, different studies state different conclusions upon the gender dependency of energy reflectance at the eardrum. Some declare that no statistically significant differences are found (Johansson and Arlinger, 2003), while others state the opposite (Feeney and Sanford, 2004). Possible influence of gender in this study is assessed by testing the difference in quadric error for men and women for different ears and different positions of the valve. The comparison is made for the frequency region between 100 Hz and 1.5 kHz, between 2.5 and 4.5 kHz, between 4.5 and 6.5 kHz, and finally between 6.5 and 8 kHz. For each separate category, a non-parametric Mann–Whitney test ($\alpha=0.05$) is carried out because of the limited dataset (Moore and McCabe, 1999). None of the analysis yields to a statistical significant difference between the quadric error for men and women ($P>0.05$); hence from this dataset it seems unnecessary to provide different male and female general characteristics for the numerical simulations.

D. Influence of geometrical characteristics

A final issue is the possible relationship between the quadric error and the geometrical parameters. The sound propagation in ears with more extreme values for certain geometrical parameters might be more difficult to model correctly; hence the quadric error between the simulated and the measured transfer functions might increase.

This question is addressed by drawing scatter plots between each geometrical characteristic and the quadric error for the different frequency regions described earlier. No distinction is made between left and right ears because each ear has its individual set of parameters. Further, the quadric error is calculated for both the measurements with open and with closed valve to include as much information as possible. On these scatter plots, a polygon that best fits the distribution of the points is drawn automatically to aid the eye in seeing important patterns (Wickham, 2008).

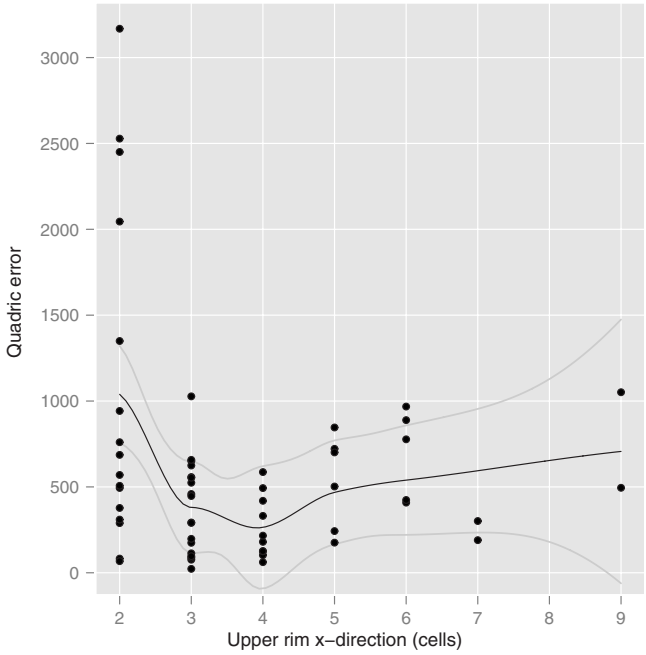


FIG. 6. Relationship between the quadric error for the frequency region from 2.5 to 4.5 kHz and the number of cells in the *x*-direction between the ear canal and the starting point of air in the pit at the end of the hearing protector.

The vast majority of the scatter plots shows randomly distributed points and hence no relationship between these particular geometrical variables and the quadric error can be deduced from this dataset, making further statistical analysis pointless. However, some plots do reveal a certain correlation between the variable under study and the quadric error for one or more frequency regions; this relationship is assessed in more detail.

Figure 6 reveals that the quadric error from 2.5 to 4.5 kHz tends to be higher for hearing protectors with a limited number of 0.35 mm gridcells in the *x*-direction between the ear canal's wall and the starting point of air in the pit at the end of the hearing protector. This quadric error also seems to increase slightly with increasing cell numbers. Comparing the difference in quadric error for ears with a two-cell rim on the one hand and ears with three and four cells on the other yields to statistically significant differences for both the two sample t-test ($df=16.115$; $P=0.012<0.05$) and the Mann–Whitney test ($P=0.0071<0.05$). A similar comparison between the ears with five or more cells and the ears with three and four cells also leads to statistical significant results for the t-test ($df=25.016$; $P=0.012<0.05$) and the Mann–Whitney test ($P=0.0085<0.05$).

Although these data have to be interpreted with some caution because of the possible correlation between the different points, they do seem to confirm the tendencies visible

in Fig. 6. This might be explained by the fact that a very small rim or very large rim implies a more abrupt transition between the earplug's inner bore and the ear canal. The boundary layer theory used in this approach is less suitable to model the viscothermal effects at more curved edges (Bockstael *et al.*, 2008), possibly increasing the difference between simulations and measurements.

The quadric error for the frequency region between 6.5 and 8 kHz seems less randomly spread for 11 geometrical parameters, all related to dimensions in the *x*- or *y*-direction. However, the observed trends are in essence introduced by the left and right ear of one participant apparently having small cross-sectional characteristics and an enlarged quadric error in this frequency region. The fact that the aberrant data belong to one participant somewhat undermines the hypothesis of a causal relationship between the ear canal's dimensions and the associated quadric error. This idea is further refuted by the finding that for most parameters more extreme values are also found for ears with only a limited quadric error.

VII. DISCUSSION

Hearing protectors are a useful tool in the prevention of noise-induced hearing loss (Arezes and Miguel, 2002), provided that they are worn correctly (Rabinowitz *et al.*, 2007). Unfortunately, the actual use of protectors at the workplace is seldom optimal (Hong *et al.*, 2008). Feelings of discomfort are an important reason for this (Hong *et al.*, 2008), in addition Abel *et al.* (1988) showed that the size of the protectors often does not adequately meet the needs of the consumers.

Custom-made earplugs are by definition adapted to the individual's ear canal geometry and are rated higher with regard to comfort (Hsu *et al.*, 2004); hence they deserve extra attention in the prevention of hearing loss due to noise.

An additional advantage is the possibility of drilling a test bore over the total length of the earplug which can be used to measure the protector's attenuation *in situ* following the MIRE approach. Field testing might not only bridge the gap between the laboratory estimate of the attenuation and the actual protection achieved by the workers; an individualized approach might also improve the results of hearing conservation programs (Witt, 2007).

One important issue in this approach is the difference between the sound pressure measured by the MIRE measurement microphone and the sound pressure of interest at the eardrum. This project aims to predict the transfer function of an individual based on general characteristics of an ear canal occluded by an earplug and on individual geometry. The accuracy of the FDTD model used for this purpose is verified by comparing the numerical prediction with measured transfer functions.

The measured transfer functions vary substantial between themselves, but most of them follow the same global frequency dependence. This is in accordance with results from Hammershøi and Møller (1996) who have measured the sound transmission to the eardrum from the entrance of the ear canal for different subjects.

As for the measurement equipment, previous research has shown that the MIRE microphones provide stable and reliable measurement results (Bockstael *et al.*, 2008). In addition, different studies support the measurement procedure with the Aurical microphone. Hellstrom and Axelsson (1993) demonstrated that the miniature microphone with a probe tube attached is useful for sound measurements in the outer ear canal. Their results indicate that a probe tube with a larger outer diameter than the tube in this study, respectively, 1.5 and 0.85 mm, has negligible influence on the sound pressure in front of the eardrum. Moreover, this study's registration point close to the tympanic membrane induces least errors in the case of small movements of the probe tube (Hellstrom and Axelsson, 1993). Finally, the generalization of sound pressure in front of the entire eardrum from the sound transmitted by the probe tube seems justified. Koike *et al.* (2002) demonstrated that the distribution of pressure in front of the eardrum is almost uniform at 0.1 kHz and at 7 kHz differences are within a 2-dB-range.

In general, the numerical approach seems to predict the measured frequency dependence quite accurately, all the more because the transfer functions that clearly deviate both from their model and the global trend are most likely due to experimental error and do not depict a true transfer function between MIRE measurement microphone and eardrum (see Sec. VI A). In the frequency region below 5500 Hz, measured and simulated transfer functions resemble each other closely. Above this frequency, the numerical simulations still follow the trend of the measurements, but differences in amplitude increase.

It is indeed possible that the numerical model is less accurate for higher frequencies. In the current approach, the outer ear canal is modeled as a straight tube with uniform cross-section, whereas in reality the ear canal is bended with variable diameter (Koike *et al.*, 2002). The bended character is less important in this project because the cartilaginous part is merely filled with the hearing protector and the second bend marks the start of the transition between this part and the ossicular canal (Dillon, 2001). Still, variation in cross-section might influence the transfer function in the higher frequencies (Koike *et al.*, 2002). In addition, the vibration of the tympanic membrane becomes more complex with increasing frequency and clearly differs from the relatively simple low-frequency pattern for frequencies higher than 5000 Hz (Tonndorf and Khanna, 1972).

Besides these considerations, the question arises whether it is actually possible to approach the true transfer function correctly in this higher frequency region. Leaving aside the small intrasubject variability (Hellstrom and Axelsson, 1993) and measurement difficulties for high attenuation values (Sec. IV A), even very small variations in the position of the earplug may distinctly alter the frequency response of the ear canal's residual part, changing the appearance of the transfer function for frequencies above 4500 Hz. Berger *et al.* (2007) also described the issue of increasing variability with increasing frequency when measuring hearing protector's attenuation.

Some characteristics of the numerical simulations are common for all subjects and based on the model for a HATS

(Bockstael *et al.*, 2008), for instance, the impedance defining the ossicular part of the ear canal. Since none of the participants has a history of otological problems and none of the audiological tests reveals any abnormalities, it seems correct to apply the impedance of bone in all cases.

Further, the same model for the impedance at the eardrum has been used for the different simulations because all subjects have normal tympanometric results. As discussed in Sec. 4 C, some studies have found small gender differences in middle ear impedance. However, in this study, the quadric error between measurements and simulations does not differ significantly between men and women, suggesting that for this application the model of the eardrum impedance is adequate for both genders. A similar discussion exists upon the influence of age on the characteristics of the middle ear (Feeney and Sanford, 2004). Since the possible aging effect is mostly suspected for elderly subjects, it does not seem critical for this model designed for the working population.

The individualized features of the model are based on specific geometrical dimensions of each ear and hearing protector. The comparisons between the quadric error for the measurements and both the HATS model and the individualized model reveal that the individualized approach matches the measurements better and therefore appears to be advisable. Moreover, the models seem suitable for a considerable range of possible values for the vast majority of geometrical characteristics.

The satisfying concordance between the numerical approach and the registered transfer functions clearly shows that the transfer function between MIRE microphone and eardrum might be accurately predicted using this FDTD model. Further research will be undertaken to implement this approach in practice.

VIII. CONCLUSION

Measurements on human subject reveal an apparent difference between the sound pressure measured by the MIRE measurement microphone and the sound pressure of interest at the eardrum. For an individual ear, these transfer functions appear to be stable and reproducible. Moreover, the majority of the transfer functions shows a frequency dependency similar to previous measurements on a HATS. This yields to the conclusion that characteristics common to most transfer functions can be explained by similar acoustical phenomena. Distinct aberrations from the average in shape are traced down to experimental error.

Despite the similar frequency dependency, substantial intersubject variation is seen with respect to the the exact frequency and amplitude of the extrema. However, the measured transfer function of an individual ear can be accurately predicted by FDTD simulations. The numerical models are on the one hand based on general acoustical properties of an ear canal occluded by an earplug with two inner bores and on the other on individual geometrical parameters. The general features of the model appear to be suitable for both men and women. The introduction of individualized characteristics clearly enhances the similarity between measurements and simulations. Furthermore, the numerical model can handle a wide range of values for the vast majority of the included geometrical parameters.

Further research should be dedicated to the practical implementation of the numerical model.

ACKNOWLEDGMENTS

The authors would like to thank the firm Variphone for the fabrication of the custom-made earplugs.

2003-10-EC (**2003**). "Directive 2003-10-EC on the minimum health and safety requirements regarding the exposure of workers to the risks arising from physical agents (noise)," (European Parliament and Council, Brussels).

Abel, S., Alberti, P., and Rokas, D. (**1988**). "Gender differences in real-world hearing protector attenuation," J. Otolaryngol. **17**, 86–92.

Apfel, R. E. (**1997**). *Encyclopedia of Acoustics* (Wiley, New York), pp. 161–164.

Arezes, P. M., and Miguel, A. S. (**2002**). "Hearing protectors acceptability in noisy environments," Ann. Occup. Hyg. **6**, 531–536.

Berger, E. H. (**1986**). "Methods of measuring the attenuation of hearing protection devices," J. Acoust. Soc. Am. **79**, 1655–1687.

Berger, E. H. (**2005**). "Preferred methods for measuring hearing protector attenuation," *Environmental Noise Control* (Inter Noise, Brazil).

Berger, E. H., Voix, J., and Kieper, R. W. (**2007**). "Methods of developing and validating a field-mire approach for measuring hearing protector attenuation," *Spectrum*, 32nd Annual Conference of the National Hearing Conservation Association, Savannah, GA, Vol. **24**, Suppl. 1.

Bockstael, A., de Greve, B., Van Renterghem, T., Botteldooren, D., D'haenens, W., Keppler, H., Maes, L., Philips, B., and Vinck, B. (**2008**). "Verifying the attenuation of earplugs in situ: method validation using artificial head and numerical simulations," J. Acoust. Soc. Am. **124**, 973–981.

Botteldooren, D. (**1995**). "Finite-difference time-domain simulation of low-frequency room acoustic problems," J. Acoust. Soc. Am. **98**, 3302–3308.

Brüel, & Kjær (**1996**). "Microphone handbook," Technical Report, Denmark, http://www.bksv.com.

Buck, K., and Dancer, A. (**2007**). "Performance of different types of hearing protectors undergoing high-level impulse noise," in Noise at Work 2007—Proceedings, First European Forum on Effective Solutions for Managing Occupational Noise, Lille, pp. 1397–1406.

Casali, J. G., Mauney, D. W., and Alton, B. J. (**1995**). "Physical vs. psychophysical measurements of hearing protection attenuation—a.k.a. mire vs. reat," Sound Vib. July, 20–27.

Dillon, H. (**2001**). *Hearing Aids* (Thieme Medical, New York).

Farmer-Fedor, B., and Rabbitt, R. (**2002**). "Acoustic intensity, impedance and reflection coefficient in the human ear canal," J. Acoust. Soc. Am. **112**, 600–620.

Feeney, M., and Sanford, C. (**2004**). "Age effects in the human middle ear: Wideband acoustical measures," J. Acoust. Soc. Am. **116**, 3546–3558.

Fowler, C. G., and Shanks, J. E. (**2002**). *Handbook of Clinical Audiology* (Lippincott Williams & Wilkins, Baltimore, MD). pp. 175–204.

Franks, J. (**2003**). "Comparison of the regulatory noise reduction rating (NRR) and the required ANSI S3.19 test method with real world outcomes and results from testing with the new ANSI S12.6B method," *Workshop on Hearing Protector Devices* (United States Environmental Protection Agency, Washington, DC).

Hammershøi, D., and Møller, H. (**1996**). "Sound transmission to and within the human ear canal," J. Acoust. Soc. Am. **100**, 408–427.

Harrel, R. W. (**2002**). *Handbook of Clinical Audiology*, (Lippincott Williams & Wilkins, Baltimore, MD), pp. 71–87.

Hellstrom, P.-A., and Axelsson, A. (**1993**). "Miniature microphone probe tube measurements in the external auditory canal," J. Acoust. Soc. Am. **93**, 907–919.

Hillström, L., Mossberg, M., and Lundberg, B. (**2000**). "Identification of complex modulus from measured strains on an axially impacted bar using least squares," J. Sound Vib. **230**, 689–707.

Hillström, L., Valdek, U., and Lundberg, B. (**2003**). "Estimation of the state vector and identification of the complex modulus of a beam," J. Sound Vib. **261**, 653–673.

Hiselius, P. (**2004**). "Method to assess acoustical two-port properties of ear-

plugs," Acta. Acust. Acust. **90**, 137–151.
Hiselius, P. (**2005**). "Attenuation of earplugs—Objective predictions compared to subjective reat measurements," Acta. Acust. Acust. **91**, 764–770.
Hong, O., Samo, D., Hulea, R., and Eakin, B. (**2008**). "Perception and attitudes of firefighters on noise exposure and hearing loss," J. Occup. Environ. Hyg. **5**, 210–215.
Hsu, Y.-L., Huang, C.-C., Yo, C.-Y., Chen, C.-J., and Lien, C.-M. (**2004**). "Comfort evaluation of hearing protection," Int. J. Ind. Ergonom. **33**, 543–551.
ISO 10534-2 (**1998**). "Acoustics—Determination of sound absorption coefficient and impedance in impedance tubes—Part 2: Transfer-function method," (International Standard Organisation, Switzerland).
ISO 4869-2 (**1994**). "Acoustics—Hearing protectors—Part 2: Estimation of effective A-weighted sound pressure level when hearing protectors are worn" (International Standard Organisation, Switzerland).
ISO 8253-1 (**1989**). "Acoustics—audiometric test methods—Part 1: Basic pure tone air and bone conduction threshold audiometry," (International Standard Organisation ISO 8253-1:1989, Switzerland).
Johansson, M., and Arlinger, S. (**2003**). "Otoacoustic emissions and tympanometry in a general adult population in Sweden," Int. J. Audiol. **42**, 448–64.
Koike, T., Wada, H., and Kobayashi, T. (**2002**). "Modeling of the human middle ear using the finite-element method," J. Acoust. Soc. Am. **111**, 1306–1317.
Kringlebotn, M. (**1988**). "Network model of the human middle ear," Scand. Audiol. **17**, 75–85.
Kutner, M. H., Nachtscheim, C. J., Neter, J., and Li, W. (**2004**). *Applied Linear Statistical Models*, 5th ed. (McGraw-Hill, New York).
Moore, D. S., and McCabe, G. P. (**1999**). *Introduction to the Practice of Statistics*, 3rd ed. (W.H. Freeman and Company, New York).
Pääkkönen, R., Savolainen, S., Myllyniemi, J., and Lehtomäki, K. (**2000**). "Ear plug fit and attenuation—An experimental study," Acust. Acta Acust. **86**, 481–484.
Pickles, J. O. (**1988**). *An Introduction to the Physiology of Hearing* (Academic, London).
Rabinowitz, P., Galusha, D., Dixon-Ernst, C., Slade, M., and Cullen, M. (**2007**). "Do ambient noise exposure levels predict hearing loss in a modern industrial cohort ?," Occup. Environ. Med. **64**, 53–59.
Śliwinska-Kowalska, M., and Kotylo, P. (**2007**). "Evaluation of individuals with known or suspected noise damage to hearing," Audiological Medicine **5**, 54–65.
Tak, S., and Calvert, G. (**2008**). "Hearing difficulty attributable to employment by industry and occupation: An analysis of the national health interview survey-United States, 1997 to 2003," J. Occup. Environ. Med. **50**, 46–56.
Tonndorf, J., and Khanna, S. M. (**1972**). "Tympanic-membrane vibrations in human cadaver ears studied by time-averaged holography," J. Acoust. Soc. Am. **52**, 1221–1233.
Voix, J. (**2006**). "Mise au point d'un bouchon d'oreille 'intelligent' (Development of a 'smart' earplug)," Ph.D. thesis, Ecole de Technologie Supérieure Université du Québec.
Voss, S., and Allen, J. (**1994**). "Measurement of acoustic impedance and reflectance in the human ear canal," J. Acoust. Soc. Am. **95**, 372–384.
Wickham, H. (**2008**). "The ggplot package," Technical Report 0.4.2, The R Project for Statistical Computing, www.r-project.org (Last viewed August **2008**).
Wit, H. P., Damme, K. J., and Van Spoor, C. W. (**1987**). *Fysica voor de fysiotherapeut (Physics for the physiotherapist)* (Bunge, Utrecht).
Witt, B. (**2007**). "Putting the personal back into PPE: Hearing protector effectiveness," Occup. Health Saf. **76**, 90–92.

Normal incidence sound transmission loss evaluation by upstream surface impedance measurements

Raymond Panneton[a)]
GAUS, Department of Mechanical Engineering, Université de Sherbrooke, Sherbrooke, Quebec J1K 2R1, Canada

(Received 22 August 2008; revised 20 November 2008; accepted 9 January 2009)

A method is developed to obtain the normal incidence sound transmission loss of noise control elements used in piping systems from upstream surface impedance measurements only. The noise control element may be a small material specimen in an impedance tube, a sealing part in an automotive hollow body network, an expansion chamber, a resonator, or a muffler. The developments are based on a transfer matrix (four-pole) representation of the noise control element and on the assumption that only plane waves propagate upstream and downstream the element. No assumptions are made on its boundary conditions, dimensions, shape, and material properties (i.e., the element may be symmetrical or not along its thickness, homogeneous or not, isotropic or not). One-load and two-load procedures are also proposed to identify the transfer matrix coefficients needed to obtain the true transmission loss of the tested element. The method can be used with a classical two-microphone impedance tube setup (i.e., no additional downstream tube and downstream acoustical measurements). The method is tested on three different noise control elements: two impedance tube multilayered specimens and one expansion chamber. The results found using the developed method are validated using numerical simulations.
 [DOI: 10.1121/1.3076929]

PACS number(s): 43.50.Yw, 43.55.Rg, 43.55.Ev, 43.58.Bh [KVH] Pages: 1490–1497

I. INTRODUCTION

One important global acoustic indicator to evaluate the noise insulation properties of noise control elements used in piping systems (e.g., expanding sealing parts in automotive hollow body networks, small material specimens in an impedance tube, expansion chambers, resonators, and mufflers) is the normal sound transmission loss (nSTL). It is generally measured by placing the element to test between an upstream tube and a downstream tube, where only plane acoustic waves propagate—see Fig. 1. Based on this general setup, with proper acoustic measurements in both tubes and different termination conditions, methods were developed to deduce the nSTL. Two classes of methods have been developed to measure this nSTL: the wave field decomposition methods[1–6] and the four-pole (transfer matrix) methods.[6–10] The common denominator to these methods is that they all require acoustic measurements in an upstream tube and a downstream tube. When considering a perfectly anechoic termination, only three microphone positions are required (two in the upstream tube and one in the downstream tube). Since such an end condition is difficult to obtain experimentally, these methods generally use four-microphone positions (two in both tubes). Consequently, compared to the standard impedance tube setup, where only an upstream tube with two microphones is required, a tube extension (i.e., the downstream tube) with one or two microphones needs to be added to apply the aforementioned methods.

With a view to use directly the standard impedance tube setup and to limit the number of measuring points to 2, a method is developed. This method, referred as the upstream surface impedance (USI) method, theoretically yields the exact nSTL. It uses the classical impedance tube shown in Fig. 1(a) and the standard two-microphone technique.[11,12] Unlike existing methods, the USI method does not require the additional downstream tube and microphones shown in Fig. 1(b). Compared to other existing methods, the proposed method only uses two upstream microphones with a total of 6 microphone signals. This is 2 signals less than the exact four-microphone two-load[4,10] and two-source[7] methods for general asymmetric elements (other four-microphone methods requiring less than 8 microphone signals are not, strictly speaking, exact for asymmetric elements, though they may yield good evaluation of nSTL).

II. THEORY

A. Surface impedance of a multilayer

To introduce the method, the multilayer depicted in Fig. 2 is used. In this figure, four acoustic waves propagating outside the multilayer are defined: two x-positive propagating plane waves, A and C, and two x-negative propagating plane waves, B and D. Assuming an anechoic termination on the downstream side, wave D vanishes, and wave C becomes equal to the sound transmission coefficient τ. For all other types of termination, wave C is the sum of the transmission coefficient and an infinite number of back reflected waves between the termination and the multilayer. For the

a)Author to whom correspondence should be addressed. Electronic mail: raymond.panneton@usherbrooke.ca

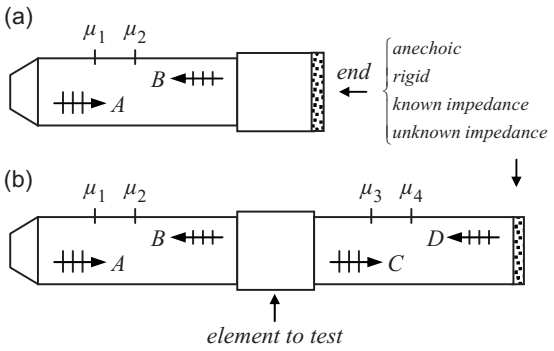


FIG. 1. (a) Standard two-microphone impedance tube (e.g., the BK 4206 impedance tube). (b) Typical transmission tube with 3 or 4 microphones (e.g., the BK 4206 T transmission loss tube).

multilayer in Fig. 2, Dun and Davern[13] showed that the acoustic surface impedance in front of the ith layer is given by

$$Z_i = W_i \frac{Z_{i-1} \coth \gamma_i L_i + W_i}{Z_{i-1} + W_i \coth \gamma_i L_i}, \tag{1}$$

where W_i, γ_i, and L_i are the characteristic impedance, propagation constant, and thickness of the ith medium. From this equation, it can be shown that the surface impedance in front of the multilayer can always be expressed in the form

$$Z_n = \frac{\zeta_1 Z_{\text{end}} + \zeta_2}{Z_{\text{end}} + \zeta_3}, \tag{2}$$

where Z_{end} is the surface impedance of the termination, and the zeta coefficients are impedance parameters. They will be discussed later on. If the termination is a plenum of air backed by a rigid wall

$$Z_{\text{end}} = \rho_0 c_0 \coth jk_0 L_0, \tag{3}$$

where ρ_0, c_0, and k_0 are the density, sound speed, and wave number in the air at rest. If the termination is a perfectly reflecting hard wall, $Z_{\text{end}}=\infty$ and Eq. (2) reduces to $Z_n=\zeta_1$.

B. Transfer matrix representation of a noise control element

1. Arbitrary end termination surface impedance

Assuming the previous multilayer or another noise control element is modeled using a global transfer matrix **T** and the acoustic excitation comes from the left-hand side (i.e., wave A), the acoustic pressures (p_a and p_b) and velocities (u_a and u_b) at points a and b are related by the transfer matrix (or four-pole) relation

$$\begin{Bmatrix} p_a \\ u_a \end{Bmatrix} = \begin{bmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{bmatrix} \begin{Bmatrix} p_b \\ u_b \end{Bmatrix}, \tag{4}$$

where T_{ij} are the coefficients of the global transfer matrix **T**. Defining the zeta coefficients

$$\zeta_1 = \frac{T_{11}}{T_{21}}, \quad \zeta_2 = \frac{T_{12}}{T_{21}}, \quad \zeta_3 = \frac{T_{22}}{T_{21}}, \tag{5}$$

Eq. (4) is rewritten as

$$\begin{Bmatrix} p_a \\ u_a \end{Bmatrix} = T_{21} \begin{bmatrix} \zeta_1 & \zeta_2 \\ 1 & \zeta_3 \end{bmatrix} \begin{Bmatrix} p_b \\ u_b \end{Bmatrix}. \tag{6}$$

Dividing the first equation in Eq. (6) by its second one yields the surface impedance at point a

$$Z_a = \frac{p_a}{u_a} = \frac{\zeta_1 Z_{\text{end}} + \zeta_2}{Z_{\text{end}} + \zeta_3}, \tag{7}$$

where $Z_{\text{end}}=p_b/u_b$ is the surface impedance of the end termination. One can note that Eq. (7) is identical to the general form given in Eq. (2) for a multilayer.

Now, if the excitation comes from the right-hand side (i.e., wave D) and the end termination is placed on the left-hand side, the relation between acoustic pressures and velocities at points a and b is given by (the same coordinate system is used)

$$\begin{Bmatrix} p_b \\ u_b^- \end{Bmatrix} = \begin{bmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{bmatrix}^{-1} \begin{Bmatrix} p_a \\ u_a^- \end{Bmatrix}. \tag{8}$$

Due to the vectorial nature of the velocity, the superscript "$-$" indicates that the excitation is, this time, along the negative x-direction. Consequently, to be consistent with the coordinate system, if the noise control element is inverted instead of inverting excitation and termination, one would find that u^- can be changed by $-u$. Taking this change into account and using the zeta coefficients, Eq. (8) can be rewritten as

$$\begin{Bmatrix} p_b \\ u_b \end{Bmatrix} = \frac{T_{21}}{\det(\mathbf{T})} \begin{bmatrix} \zeta_3 & \zeta_2 \\ 1 & \zeta_1 \end{bmatrix} \begin{Bmatrix} p_a \\ u_a \end{Bmatrix}. \tag{9}$$

In this case, the surface impedance at point b is given by

$$Z_b = \frac{p_b}{u_b} = \frac{\zeta_3 Z_{\text{end}} + \zeta_2}{Z_{\text{end}} + \zeta_1}, \tag{10}$$

where the surface impedance of the end termination placed on the left-hand side is now given by $Z_{\text{end}}=p_a/u_a$. Again, Eq. (10) is of the form given by Eq. (2) for a multilayer. Compared to Eq. (7), ζ_1 and ζ_3 have been permuted. For a noise control element that is symmetrical along its thickness, $T_{11}=T_{22}$ and, consequently, $\zeta_1=\zeta_3$ and $Z_a=Z_b$.

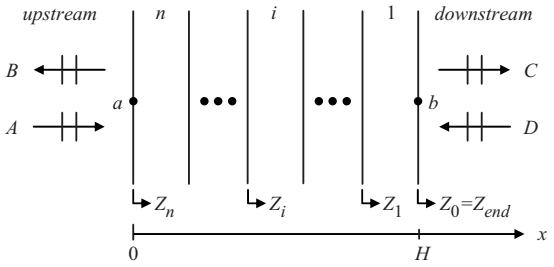


FIG. 2. General multilayered system backed by arbitrary surface impedance (Z_{end}). In the direct mounting configuration, incident wave A impinges the medium on face a.

2. Rigid end termination and meaning of ζ_1 and ζ_3

With a rigid end termination (i.e., $Z_{\text{end}}=\infty$), one easily deduces from Eqs. (7) and (10) that ζ_1 and ζ_3 are the surface impedance of the direct and inverted mounting configura-

tions, respectively. The direct configuration is when face a is facing the excitation (wave A) at $x=0$ and face b is lying on the end termination at $x=H$ (i.e., the configuration shown in Fig. 2). The inverted configuration is when face b is now facing the excitation (wave A) at $x=0$ and face a is lying on the end termination at $x=H$. Both impedances can be directly measured using a typical impedance tube following ASTM E1050-98 or ISO 10534-2:1998 standards.[11,12]

3. Anechoic end termination and transmission loss

With the excitation coming from the left-hand side and an anechoic end termination (i.e., $Z_{\text{end}}=\rho_0 c_0$), wave D vanishes, and the pressures and velocities at points a and b (respectively, at $x=0$ and $x=H$) become

$$p_a = A(1+R_\infty), \quad p_b = A\tau e^{-jk_0H},$$

$$u_a = \frac{A}{\rho_0 c_0}(1-R_\infty), \quad u_b = \frac{A\tau}{\rho_0 c_0}e^{-jk_0H}, \tag{11}$$

where $\tau=C/A$ is the transmission coefficient in the positive x-direction, and $R_\infty=B/A$ is the reflection coefficient at point a. Using Eq. (11), Eq. (6) is rewritten as

$$\begin{Bmatrix} 1+R_\infty \\ \dfrac{1-R_\infty}{\rho_0 c_0} \end{Bmatrix} = T_{21}\begin{bmatrix} \zeta_1 & \zeta_2 \\ 1 & \zeta_3 \end{bmatrix}\begin{Bmatrix} 1 \\ \dfrac{1}{\rho_0 c_0} \end{Bmatrix}\tau e^{-jk_0H}. \tag{12}$$

Similarly, for the inverted configuration, one finds

$$\begin{Bmatrix} 1+R'_\infty \\ \dfrac{1-R'_\infty}{\rho_0 c_0} \end{Bmatrix} = \frac{T_{21}}{\det(\mathbf{T})}\begin{bmatrix} \zeta_3 & \zeta_2 \\ 1 & \zeta_1 \end{bmatrix}\begin{Bmatrix} 1 \\ \dfrac{1}{\rho_0 c_0} \end{Bmatrix}\tau' e^{-jk_0H}, \tag{13}$$

where τ' and R'_∞ are the transmission coefficient and reflection coefficient in the inverted configuration.

From Eqs. (12) and (13), the reflection coefficients of the direct and inverted mounting configurations of the noise control element can be written, respectively, as

$$1+R_\infty = 2\frac{\zeta_1 + \dfrac{\zeta_2}{\rho_0 c_0}}{\zeta_1+\zeta_3+\dfrac{\zeta_2}{\rho_0 c_0}+\rho_0 c_0} \tag{14}$$

and

$$1+R'_\infty = 2\frac{\zeta_3 + \dfrac{\zeta_2}{\rho_0 c_0}}{\zeta_1+\zeta_3+\dfrac{\zeta_2}{\rho_0 c_0}+\rho_0 c_0}. \tag{15}$$

In general, the noise control element is not symmetrical (i.e., $Z_a \neq Z_b$). In this case, $R'_\infty \neq R_\infty$, and $T_{11} \neq T_{22}$. However, invoking the *principle of reciprocity*[14] (if nonlinear acoustic response is not considered), the transmission coefficients are equal, $\tau'=\tau$. Consequently, dividing the first equation in Eq. (12) by the first equation of Eq. (13), and using Eqs. (14) and (15), one shows

$$\det(\mathbf{T}) = 1 \tag{16}$$

and

$$T_{21}^2 = \frac{1}{\zeta_1\zeta_3-\zeta_2}. \tag{17}$$

Taking into account these results, Eq. (12) or Eq. (13) leads to the transmission coefficient

$$\tau = \frac{2\sqrt{\zeta_1\zeta_3-\zeta_2}}{\zeta_1+\dfrac{\zeta_2}{\rho_0 c_0}+\zeta_3+\rho_0 c_0}e^{jk_0H} \tag{18}$$

or in terms of the T_{ij} coefficients

$$\tau = \frac{2e^{jk_0H}}{T_{11}+\dfrac{T_{12}}{\rho_0 c_0}+\rho_0 c_0 T_{21}+T_{22}}. \tag{19}$$

Here, one needs to select the proper sign for the root. This is not straightforward; it depends on the phase to unwrap it correctly. However, using the nSTL, this difficulty is overcome since only the absolute value of the transmission coefficient is used. The nSTL is defined as

$$\text{nSTL} = -10\log_{10}\|\tau\|^2. \tag{20}$$

This equation, together with the definition of τ in Eq. (19), is identical to Eq. (7) in the original paper by Lung and Doige[4] on the two-load four-microphone method.

4. Meaning of ζ_2

To try giving a physical interpretation of ζ_2, Eq. (18) can be rewritten as

$$\tau = 2j\frac{\sqrt{1-\dfrac{\zeta_1\zeta_3}{\zeta_2}}}{\dfrac{\zeta_1+\zeta_3}{\sqrt{\zeta_2}}+\dfrac{\sqrt{\zeta_2}}{\rho_0 c_0}+\dfrac{\rho_0 c_0}{\sqrt{\zeta_2}}}e^{jk_0H}. \tag{21}$$

For a single symmetrical homogeneous isotropic rigid porous layer of thickness H, one can show from Ref. 8 that the transmission coefficient is given by

$$\tau = 2j\frac{\sqrt{1-\coth^2\gamma H}}{2\coth\gamma H+\dfrac{W}{\rho_0 c_0}+\dfrac{\rho_0 c_0}{W}}e^{jk_0H}, \tag{22}$$

where W and γ are the characteristic impedance and propagation constant of the porous layer. Comparing Eqs. (21) and (22), one can conclude that for a general noise control element, ζ_2 can be seen as the square of its "equivalent" characteristic impedance.

III. TEST PROCEDURES

A. One-load procedure

Figure 3 shows a typical test procedure to use with the standard two-microphone impedance tube technique.[11,12] It consists in three surface impedance measurements performed on the element to test. The first measurement is performed on the direct mounting configuration; the element to test is

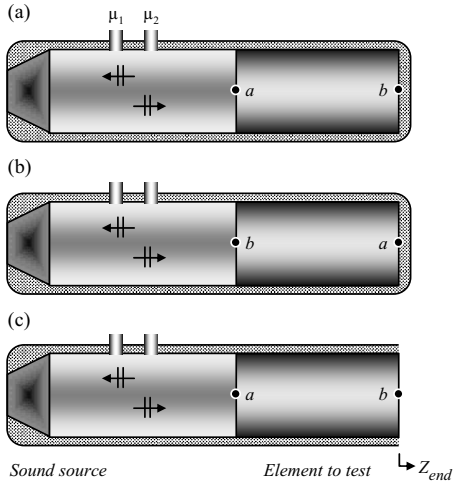


FIG. 3. Test configurations for the USI method. (a) Direct mounting configuration with hard backing. (b) Inverted mounting configuration with hard backing. (c) Direct mounting configuration with arbitrary load Z_{end}.

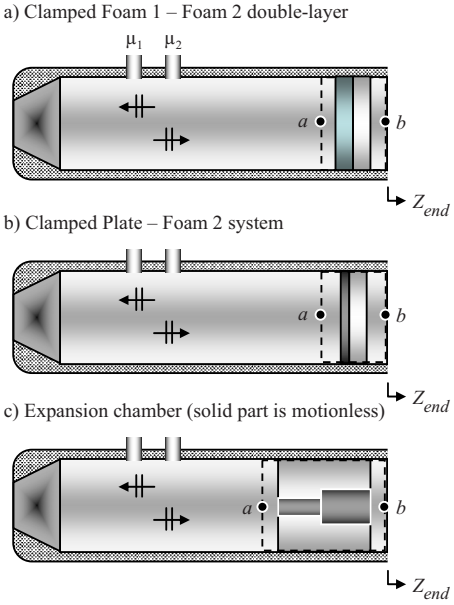


FIG. 4. (Color online) Tested configurations in a circular impedance tube of 45-mm in diameter. The dashed line represents the element to test with its optional air plenums.

backed by the rigid wall, and face a is facing the excitation. The second measurement is performed on the inverted mounting configuration; face b is now facing the excitation. As explained in Sec. II B 2, these two measurements directly give surface impedances ζ_1 and ζ_3, respectively. The third measurement gives the surface impedance Z_s. It is performed on the direct mounting configuration backed by a load of known surface impedance Z_{end}. The load may be an additional plenum of air, an anechoic termination, or another type of nonrigid termination (e.g., termination with an absorbing layer). Its surface impedance may be measured or, for simple terminations, analytically computed; this is the case for a plenum of air and a perfectly anechoic termination, where their Z_{end} are analytically given by Eq. (3) and $\rho_0 c_0$, respectively. From the previously measured surface impedances, the only remaining unknown coefficient, ζ_2, is found using Eq. (2):

$$\zeta_2 = Z_s(Z_{\text{end}} + \zeta_3) - \zeta_1 Z_{\text{end}}. \tag{23}$$

Now, the zeta coefficients being known, Eqs. (18) and (20) can be used to evaluate the nSTL. Also, using Eq. (17) and the relations given in Eq. (5), the global transfer matrix **T** can be built.

B. Three-load procedure

Another method to evaluate the three zeta coefficients, without inverting the element to test, is to measure the normal surface impedance of the element for three different loads of known surface impedances $Z_{\text{end}}^{(1)}$, $Z_{\text{end}}^{(2)}$, and $Z_{\text{end}}^{(3)}$. These three measurements yield $Z_s^{(1)}$, $Z_s^{(2)}$, and $Z_s^{(3)}$, respectively. The loads may be air cavities, a rigid end, an anechoic end, or another type of conditions. Again, as discussed in the previous paragraph, their respective surface impedance may be measured or, for simple terminations, analytically computed. With these results, one builds the linear system

$$\begin{bmatrix} Z_{\text{end}}^{(1)} & 1 & -Z_s^{(1)} \\ Z_{\text{end}}^{(2)} & 1 & -Z_s^{(2)} \\ Z_{\text{end}}^{(3)} & 1 & -Z_s^{(3)} \end{bmatrix} \begin{Bmatrix} \zeta_1 \\ \zeta_2 \\ \zeta_3 \end{Bmatrix} = \begin{Bmatrix} Z_s^{(1)} Z_{\text{end}}^{(1)} \\ Z_s^{(2)} Z_{\text{end}}^{(2)} \\ Z_s^{(3)} Z_{\text{end}}^{(3)} \end{Bmatrix}. \tag{24}$$

Solving this linear system yields the zeta coefficients, and the global transfer matrix and nSTL can be computed. It is worth mentioning that one can also perform more than three tests (i.e., the minimum number of tests to solve for ζ_1, ζ_2, and ζ_3). In this case, the system in Eq. (24) would be overdetermined, and the system could be solved in a least squares sense; this could be helpful to minimize the effects of random noise during measurements.

IV. NUMERICAL APPLICATIONS

To validate the theoretical developments (Sec. II) and the proposed impedance tube test procedures (Sec. III), three numerical applications are investigated, see Fig. 4. The first application deals with a clamped double-layer made up from two poroelastic open-cell foams. The second application deals with a clamped plate-foam system. The third application deals with an expansion chamber. All these elements are non-symmetric along their thickness, circular, and clamped to the wall of the tube (i.e., no axial motion at the wall of the tube). The tube is 45-mm in diameter. The material properties and dimensions of these three elements are given in Table I. Note that due to their clamped conditions, the double poroelastic layer and the plate-foam system could not be analytically modeled using the classical transfer matrix approach.[14] In fact, the classical transfer matrix approach supposes infinite lateral dimensions.[15]

With a view to validate the theoretical developments, the impedance tube applications shown in Fig. 4 will be simulated numerically using an axi-symmetrical finite element method; this will allow eliminating uncontrollable parameters and uncertainties inherent to experimentations. The

TABLE I. Physical and elastic properties of the materials used for the simulations.

	Foam 1	Foam 2	Φ2 mm pore 1	Φ5 mm pore 2	Plate	Units
Open porosity	0.97	0.98	1	1	⋯	
Static airflow resistivity	87 000	11 500	147	24	⋯	$N\,s/m^4$
Tortuosity	2.52	1.03	1	1	⋯	
Viscous characteristic dimension	37×10^{-3}	90×10^{-3}	1	2.5	⋯	mm
Thermal characteristic dimension	119×10^{-3}	180×10^{-3}	1	2.5	⋯	mm
Bulk density	31	9	⋯	⋯	2742	kg/m^3
Young's modulus	143 000	160 000	⋯	⋯	69×10^9	Pa
Poisson's ratio	0.3	0.33	⋯	⋯	0.33	
Loss factor	0.055	0.04	⋯	⋯	0.007	
Thickness	25	50	50	50	0.6	mm

used axi-symmetrical finite element models are shown in Fig. 5. While typical acoustical and solid finite elements will be used to model the air and the plate, respectively, the foams will be modeled using poroelastic finite elements.[16] Also, to take into account the viscous and thermal losses in the narrow tubes of the expansion chamber, the effective properties of the air in the narrow tubes will be computed using a cylindrical pore model. Using commercially available acoustic softwares, a single pore can be modeled as an equivalent fluid[14] with the properties given in Table I (note that we have not taken into account any restriction effects that would virtually increase the tortuosity or effective length of the pores).

To simulate a normal incidence excitation, a unit pressure will be applied on the source side. The end condition will be simulated using an additional fluid layer (for an air cavity backing), a bonded condition (for a rigid termination), or a $\rho_0 c_0$ impedance condition (for an anechoic termination). Once the finite element models will be solved for frequencies between 100 to 4000 Hz, the pressures at microphones μ_1 and μ_2 will be used to apply the standard two-microphone technique[11,12] to deduce the surface impedance at $x=0$. Note that in the finite element model the element to test may include air plenums before and after the basic element to test. The use of such additional air plenums will be discussed below.

Finally, the results obtained with the proposed USI procedures will be compared to expected nSTL results. These expected results are also obtained from axi-symmetrical finite element simulations. The elements to test are modeled as previously discussed. They are placed between two axi-symmetrical waveguides of length equal to their diameter (this ensures only plane waves to propagate at both ends). At one end, a plane wave excitation is imposed. At the other end, an anechoic termination is imposed. The transmission coefficient is computed from the ratio between the transmitted sound power and the incident sound power. The transmitted and incident sound powers are numerically computed at the anechoic end and the excited end, respectively. From the transmission coefficient, Eq. (20) is used to deduce the nSTL.

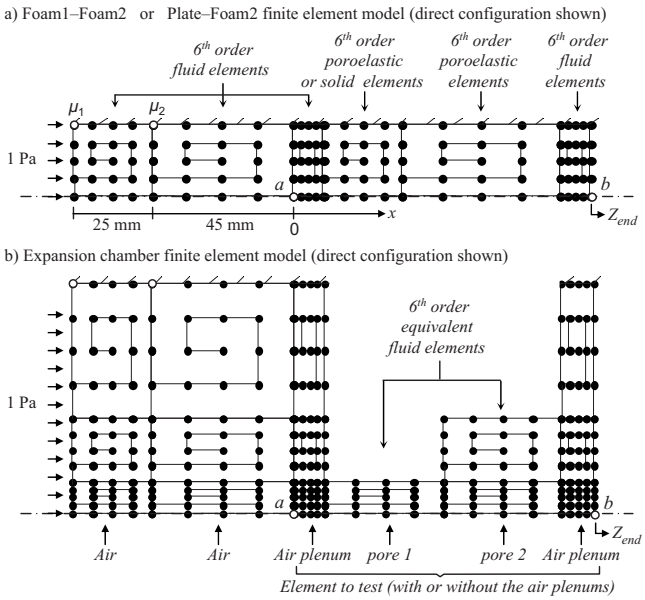


FIG. 5. Axi-symmetrical finite element models for the tested applications.

A. Double-layer poroelastic system

The results for the double-layer poroelastic configuration using the developed USI method are shown in Fig. 6. In Fig. 6(a), the one-load procedure is used on the element to test with and without additional 1-mm-thick air plenums before and after the double-layer. Here, the load is a 20-mm air cavity (note that similar results are obtained with another cavity thickness or with an anechoic load). As shown, when no air plenums are used, some discrepancies occur compared to the expected nSTL. This is due to the fact that the vibrating elastic frame of the material is in contact with the rigid

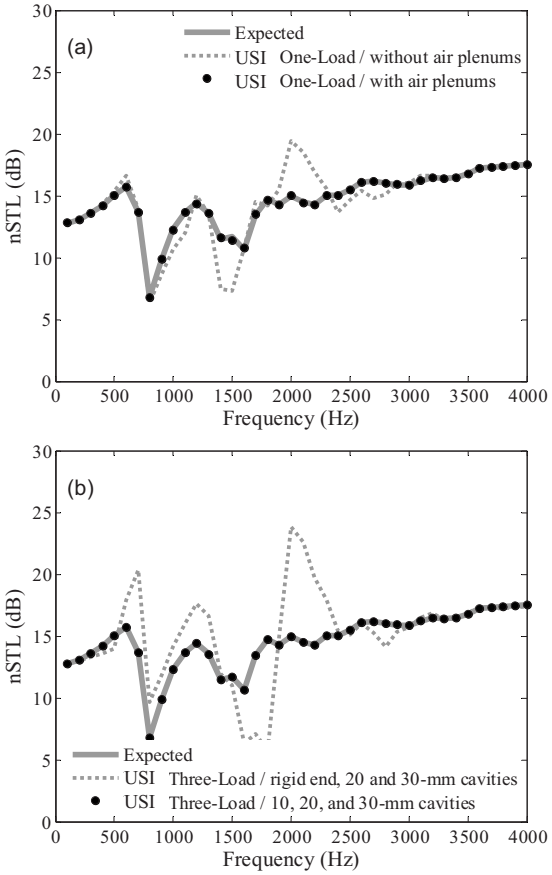


FIG. 6. nSTL of the foam 1-foam 2 system. Comparison between the expected nSTL and the developed method (USI). (a) One-load procedure. (b) Three-load procedure.

wall backing. In this case, the elastic compression wave is more excited than in a transmission loss situation (i.e., with an air backing and not with a rigid backing). Adding only 1-mm-thick air plenums before and after the double-layer solves the problem.

In the case of the three-load procedure, see Fig. 6(b), if one of the three loads is the rigid wall condition directly applied on the double poroelastic layer, the method fails for the same reasons than previously discussed. To solve the problem, each of the three loads should start with air (e.g., air cavities and anechoic termination).

B. Plate-foam 2 system

For the second application, with a view to minimize the effect of the rigid backing on the vibration behavior of the element to test, the 1-mm air plenums are added before and after the plate-foam configuration. For the USI one-load method, using a 20-mm cavity of air or an anechoic termination for the load yields the results shown in Fig. 7(a). Here, two different tests were done. The first one is when the plate-foam sequence is considered to be the direct configuration. The second is when the foam-plate sequence is the direct configuration. Theoretically, similar results should be obtained. This is what is almost observed on Fig. 7(a). However, a zoom in the range 3000–4000 Hz shows a small discrepancy between the foam-plate sequence and the expected result, see Fig. 7(b). This is due to the fact that in this

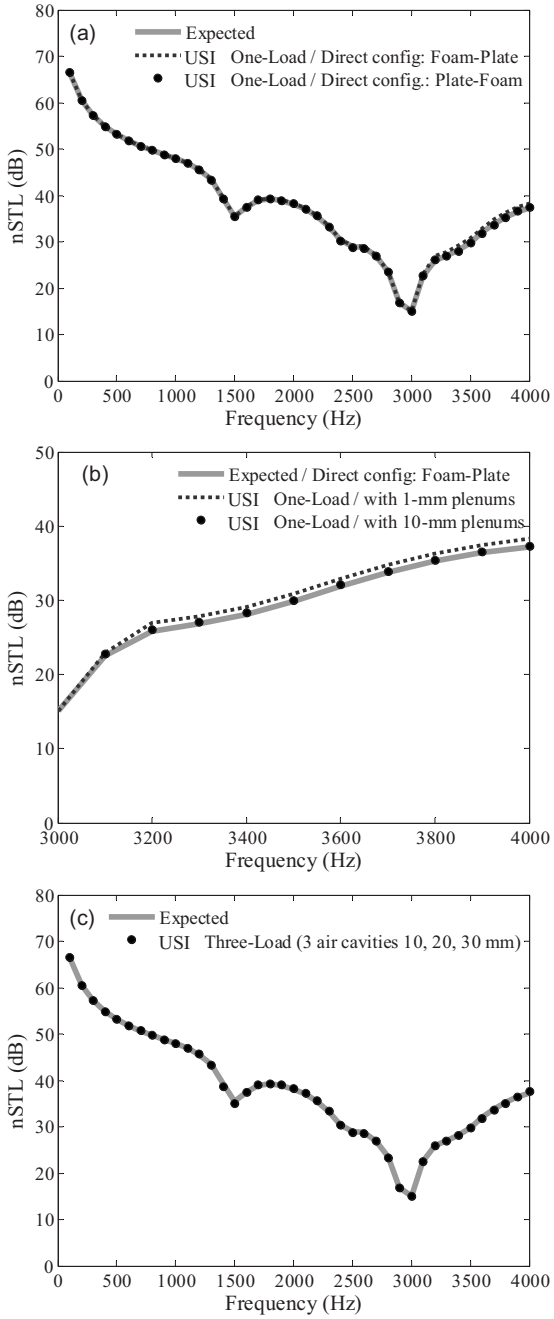


FIG. 7. nSTL of the plate-foam 2 system. Comparison between the expected nSTL and the developed method (USI). [(a) and (b)] One-load procedure. (c) Three-load procedure.

foam-plate direct configuration sequence, the second surface impedance measurement—as per Fig. 3—is done on the foam-plate sequence backed by a rigid end. Even if 1-mm plenums were added before and after the foam-plate system, the effect of the backing on the longitudinal compression elastic wave is still apparent. This may be explained by the high compression stiffness of the thin 1-mm air layer. To reduce this effect, thicker air plenums are suggested. For instance, using 10-mm air plenums before and after the element to test almost eliminates this effect, as shown in Fig. 7(b).

In the case of the USI three-load method, see Fig. 7(c), both sequences (plate-foam or foam-plate) yield similar re-

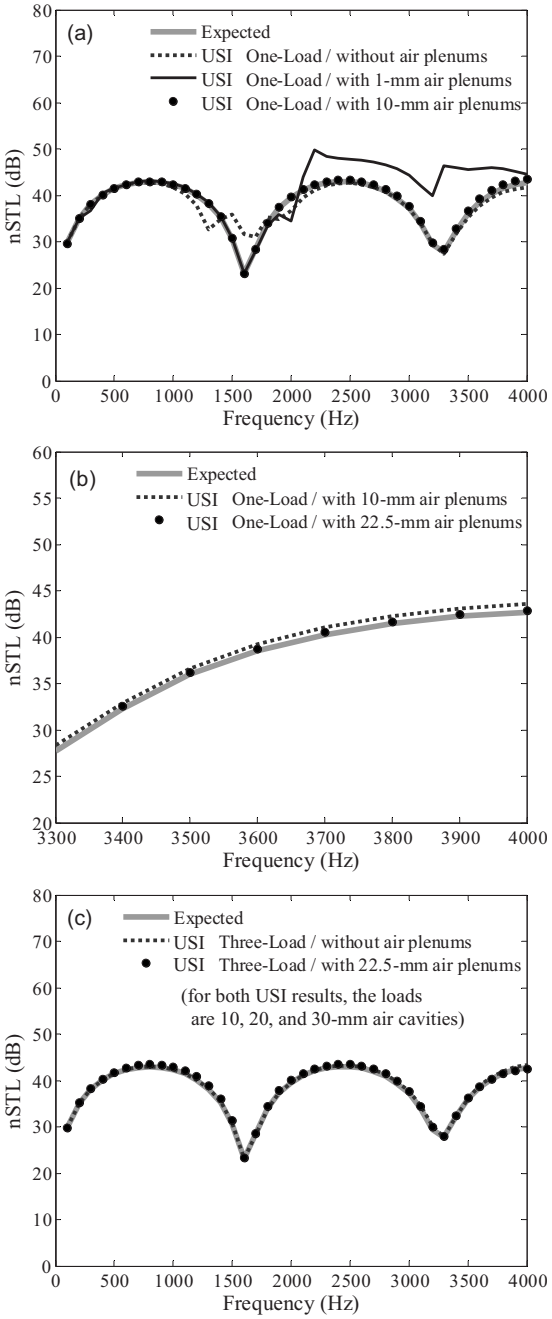


FIG. 8. nSTL of the expansion chamber. Comparison between the expected nSTL and the developed method (USI). [(a) and (b)] One-load procedure. (c) Three-load procedure.

sults. Also, it is worth mentioning that for the three-load method, the additional air plenums are not required since the rigid backing condition is never applied directly on the vibrating element to test; in fact, three cavities of air are used for the three loads (note that one of these cavities can also be replaced by the anechoic load to yield the same results).

C. Expansion chamber

For the third application and for the USI one-load method, the expansion chamber is tested without additional air plenums, with 1-mm air plenums and with 10-mm air plenums. The results are shown in Fig. 8(a). The used load is a 20-mm cavity of air (note that similar results are obtained with another cavity thickness or the anechoic termination). Even if in this case the solid part is motionless (no elastic vibration) compared to the previous applications, one can note that the air plenums before and after the expansion chamber are still necessary, the 10-mm air plenums yielding better results than the 1-mm air plenums. This may be explained by the reasons exposed for the previous applications. To be cautious, one could use air plenums equal to the radius of the impedance tube—see improvements in Fig. 8(b) with 22.5-mm plenums.

In the case of the USI three-load method, the results are presented in Fig. 8(c). Here, the three loads are three cavities of 10-, 20-, and 30-mm-thick. The cases without and with 22.5-mm air plenums before and after the expansion chamber are shown. Both cases yield the expected nSTL.

V. DISCUSSION AND CONCLUSION

A method to measure the nSTL with USI measurements was developed. For this so-called USI method, two procedures were proposed to retrieve the nSTL. The first one, the one-load procedure, requires a maximum of four successive surface impedance measurements: one on the direct mounting configuration backed by a rigid wall, one on the inverted mounting configuration backed by a rigid wall, one on the direct mounting configuration backed by a given load, and one on the load only. The latter measurement is not required if the surface impedance of the load is known (from previous measurements or from an analytical model). The second procedure is the three-load procedure. It requires a maximum of six successive surface impedance measurements: three on the direct mounting configuration backed by three different loads, respectively, and one on each three loads. Again, the latter three measurements are not required if the surface impedances of the loads are known.

To validate the theoretical developments of the USI one-load and three-load methods, three numerical applications based on the finite element method were studied: the double poroelastic layer, the plate-foam system, and the expansion chamber element. The results obtained have shown the validity of the theoretical developments. However, and mainly for the USI one-load method, some precautions have been highlighted to ensure the element to test is placed in a transmission loss situation. It means that the element to test should be sandwiched between air plenums, as shown in Fig. 4, to fall exactly on the expected nSTL. Then, the new element to test is the basic element sandwiched between air plenums. Since the transmission loss of a small plenum of air is approximately equal to zero, the nSTL of the whole configuration is practically equal to the nSTL of the element alone. However, if one needs a very accurate evaluation of the transmission coefficient or the global transfer matrix elements, one could subtract the effect of the plenums by premultiplying and postmultiplying the found transfer matrix by the inverse of the analytically computed transfer matrix of the air plenums. Finally, the air plenums to add may be as thin as 1-mm when dealing with low resistive or permeable elements, but for high resistive or impervious elements, the plenums must be thicker to reduce the rigidity of the air

plenum behind the element. To be cautious, the thickness of the air plenums could be set equal to the radius of the tube.

To ensure the experimental feasibility of the presented method and its optimal use conditions (load conditions, thickness of the air plenums, and digital data processing), a number of experiments on different noise control elements are yet to be investigated. However, the good numerical validations presented in this paper motivate one to go further in this investigation. Also, further works should address the possibility to extend the USI method to free-field measurements and to take into account an airflow which is important for some noise control elements such as mufflers.

ACKNOWLEDGMENTS

The author would like to thank NSERC Canada and FQNRT Quebec for their financial supports.

[1]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).
[2]J. Y. Chung and D. A. Blaser, "Transfer function method of measuring in-duct acoustic properties. I. Theory," J. Acoust. Soc. Am. **68**, 907–913 (1980).
[3]J. Y. Chung and D. A. Blaser, "Transfer function method of measuring in-duct acoustic properties. II. Experiment," J. Acoust. Soc. Am. **68**, 914–921 (1980).
[4]T. Y. Lung and A. G. Doige, "A time-averaging transient testing method for acoustic properties of piping systems and mufflers," J. Acoust. Soc. Am. **73**, 867–876 (1983).
[5]J. S. Bolton, R. J. Yun, J. Pope, and D. Apfel, "Development of a new sound transmission test for automotive sealant materials," Proceedings of SAE 1997, Paper No. 971896.
[6]Z. Tao and A. F. Seybert, "A review of current techniques for measuring muffler transmission loss," SAE 2003 Noise & Vibration Conference and Exhibition (Grand Traverse, MI), Document No. 2003-01-1653, http://www.sae.org/technical/papers/2003–01–1653 (2003).
[7]M. L. Munjal and A. G. Doige, "Theory of a two source-location method for direct experimental evaluation of the four-pole parameters of an aeroacoustic element," J. Sound Vib. **141**, 323–333 (1990).
[8]B. H. Song and J. S. Bolton, "A transfer matrix approach for estimating the characteristic impedance and wave numbers of limp and rigid porous materials," J. Acoust. Soc. Am. **107**, 1131–1152 (2000).
[9]B. H. Song and J. S. Bolton, "Effect of circumferential edge constraint on the acoustical properties of glass fibers," J. Acoust. Soc. Am. **110**, 2902–2916 (2001).
[10]O. Oliviero and J. S. Bolton, "Measurement of transmission loss of materials using standing wave tube," Proceedings of Inter-Noise 2006, Honolulu, HI, 3–6 December 2006.
[11]Anonymous, "Standard test method for impedance and absorption of acoustical materials using a tube, two microphones, and a digital frequency analysis system," ASTM Standard E1050-98.
[12]Anonymous, "Acoustics—Determination of sound absorption coefficient and impedance in impedance tubes—Part 2: Transfer-function method," ISO Standard 10534-2:1998.
[13]I. P. Dun and W. A. Davern, "Calculation of acoustic impedance of multilayer absorbers," Appl. Acoust. **19**, 321–334 (1986).
[14]J.-F. Allard, B. Brouard, D. Lafarge, and W. Lauriks, "Reciprocity and antireciprocity in sound transmission through layered materials including elastic and porous media," Wave Motion **17**, 329–335 (1993).
[15]J.-F. Allard, *Propagation of Sound in Porous Media* (Elsevier Applier Science, New York, 1993).
[16]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).

Quantifying sound-field diffuseness in small rooms using multifractals

S. J. Loutridis[a)]
Department of Electrical Engineering, Technological Institute of Larissa, Sensors and Instrumentation Laboratory, GR 41-110 Larissa, Greece

(Received 24 April 2008; revised 13 December 2008; accepted 15 December 2008)

With the aim of quantifying sound-field uniformity in spaces of relatively low volume, four different configurations in terms of acoustic treatment and sound-field diffuseness were constructed and tested. In a diffuse sound field, reflections are not strongly correlated both to the original sound and to earlier reflections. The degree of correlation is embedded in the impulse response structure, but is not trivial to identify. The room impulse responses exhibit self-similarity and therefore may be treated as multifractal signals characterized by a singularity spectrum. The singularity spectrum contains a wealth of information about the acoustic field established in the room. The width of the singularity spectrum, in particular, is directly related to the complexity of the impulse response structure. The uniformity of the sound field in the room can be evaluated by examining the variation of the singularity spectra with position. A new definition for sound field diffuseness is given that is not based on the concept of the energy density. The proposed method is simple to apply, statistically robust, and provides a measure of diffusion independent of the room reverberation.
 [DOI: 10.1121/1.3075560]

PACS number(s): 43.55.Br [NX] Pages: 1498–1505

I. INTRODUCTION

The role of diffusion in auditorium acoustics has been acknowledged by many authors; see, for example, Ref. 1. A sound field is said to be perfectly diffuse if at any moment the sound energy density is uniform in the whole volume and if at any point the energy flow is the same at all directions.[2] This is an ideal situation, which a sound field may approach more or less closely. The problem with sound-field diffuseness is that no satisfactory test procedure for its measurement exists, although many suggestions and attempts have been made in the past years. Most of the work done was carried out in order to quantify the sound-field uniformity in reverberation chambers for absorption measurements.[3,4] Very little work aimed specifically at examining the sound-field structure in small rooms, such as studio control rooms. In this context, the work of Randal and Ward[5] should be mentioned. It is true that a diffuse field is more difficult to establish in small spaces. However, it has been experimentally demonstrated that even in small rooms, the uniformity of the sound field can be significantly improved with diffusers.[6]

A diffuser is any device that reflects sound in a great number of directions independent of the incident angle. From the perspective of the receiver, the reflection originates from multiple points on the diffuser surface. In addition, the surface roughness produces multiple propagation distances, which in turn result in a temporal distribution of reflected sound, Fig. 1. The effects of diffuse reflections are both spatial broadening and temporal smearing. The inclusion of diffuse reflections is important in room-acoustic modeling in order to enhance prediction accuracy. A review of computerized prediction methods, which can handle both diffuse and specular reflections, is given by Dalenback *et al.*[7] Recently, Jing and Xiang[8] proposed a modified boundary condition that improves the room-acoustic prediction accuracy of a diffusion equation model. A comparison of the theoretical predictions with the experimental results indicates that the proposed diffusion equation model is capable of predicting the non-uniformity of sound energy density and the sound energy decay in disproportionate rooms.

Hodgson[9] examined the conditions under which a diffuse field is expected to occur in a room. He found that diffuse-field theory can be applied in the case of an empty room with quasi-cubic dimensions, specularly reflecting surfaces and uniform surface absorption. Any departure from these conditions seemed to deteriorate theoretical predictions. Theoretically, a diffuse field can be established with only specular reflections provided that there is enough time for the sound waves to blend. In practice, some amount of absorption is necessary in order to control reverberation and therefore a diffuse field must be obtained by modifying the shape of the room surfaces. In principle, the application of absorbent materials reduces sound-field diffuseness because absorption speeds up the decay of sound and as a result there is less time available for the sound waves to blend. The modern practice in control room design is to avoid excessive absorption and use instead diffusive elements.

In this work, the diffuseness of a sound field in a room is assessed by use of the concept of multifractality. The paper is structured as follows: Section II presents a literature review on the methods used to measure diffusion. Section III describes the calculation of the singularity spectrum with the detrended fluctuation analysis (DFA) algorithm. Section IV describes the experimental set-up used. Section V presents

[a)]Author to whom correspondence should be addressed. Electronic mail: loutridi@teilar.gr

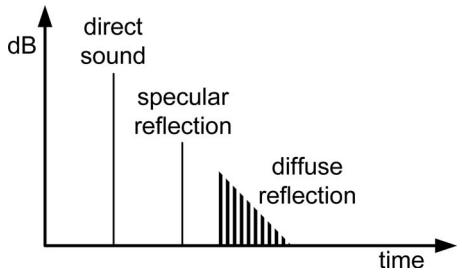


FIG. 1. Specular reflection and diffuse reflection.

the results obtained on the room configurations studied. Section VI concludes the paper.

II. METHODS FOR EVALUATING DIFFUSION

Various methods have been proposed for evaluating the diffuseness of a sound field. A brief discussion of some of them follows.

A. Directional microphone

A directional microphone turned in all possible directions in a diffuse sound field should yield the same output independent of the direction in which it is pointing. An implementation of the method is reported by Kulowski.[10] According to the author the sensitivity of the method depends on the resolving power of the directional microphone. Gover *et al.*[11] used a spherical array of 32 microphones in order to measure the directional properties of the reverberant sound fields in small and medium sized rooms. In the above study, the directional diffusion was defined as the ratio of the incident energy on a single element to the total energy. The time evolution of the sound field has also been investigated by calculating the energy in windows of different sizes. Spring and Randal[12] considered the directional microphone methods to be valueless for small rooms of moderate to heavy absorption.

B. Angular dependence of the reverberant field

Schultz[13] designed and tested an acoustic watt-meter. Today a number of devices capable of measuring sound intensity are commercially available. In principle, an acoustic watt-meter measures vector energy flow and therefore should yield a reading independent of angular orientation. As the author mentions, the method is not very sensitive to deviations from perfect diffuseness even in the extreme case of a plane wave.

C. Frequency irregularity

In this method, it is assumed that the state of sound diffusion in a room can be determined by a measurement of the frequency irregularity, i.e., the steady state frequency response; see Ref. 14. Schroeder and Kuttruff[15] proved that the frequency irregularity is proportional to the reverberation time with a factor of proportionality, which is independent of the room. Furthermore, statistics dictate that the level of a tone in a room may vary on average ± 10 dB depending on the position of measurement. The sound pressure level at any position is mostly determined by the standing wave patterns.

D. Cross-correlation methods

Cook *et al.*[16] measured the cross-correlation coefficients for the sound pressure at any two points in a room, as a means for determining the degree of "randomness" of a sound field. They reported that in a diffuse field the excitations of the two microphones are independent of each other, as soon as a certain distance is exceeded, i.e., the correlation function becomes zero. This result is only a necessary condition and not a sufficient one for the existence of a diffuse sound field. Similar results have been reported by Balachandran[17] and more recently by Nelisse and Nicolas.[18] Jacobsen and Roisin[19] introduced the spatial coherence functions and showed that the experimental spatial coherence in a room is significantly different from that obtained in a reverberation chamber. Schroeder *et al.*[20] proposed interaural cross correlation (IACC), which is a measure of the difference of the signals arriving at a listener's ears. In a study of European concert halls the IACC reduces as the degree of diffuseness increases. The IACC has been found to be a sensitive parameter to sound-field diffuseness in other studies as well.[21,22] In these studies it is also mentioned that another parameter affected by the scattering of the hall surfaces is the initial time delay gap.

E. Uniformity and linearity of decay rate

Randal and Ward[5] statistically processed the decay of sound energy in a small room in which the distribution of absorbers could be changed. They reported that the variation of the reverberation time with position is related to the diffuseness of the sound field. It is also well documented that in a diffuse field the decay of sound energy with time on logarithmic axes approaches a straight line.[6]

F. Subjective methods

Haan and Fricke[23] took a subjective approach on the subject of diffusion evaluation. For estimating the goodness of a hall they defined an acoustic quality index (AQI) ranging from 0 to 1. A surface diffusivity index (SDI) was also proposed based on the visual inspection of the hall surfaces. They found that the correlation between SDI and AQI exceeded $r^2 > 0.6$.

With the exception of the methods described in Sec. II E, no other method has been reported to work well in small spaces. It is therefore not an exaggeration to conclude that no method provides really encouraging results in terms of applicability and shear sensitivity. In Sec. V, the sound-field diffuseness is related to the width of the singularity spectrum of a multifractal impulse response.

III. CALCULATION OF THE SINGULARITY SPECTRUM WITH DETRENDED FLUCTUATION ANALYSIS

In this section, a brief theoretical background is provided on the concept of multifractality and the calculation of the singularity spectrum. The theory presented below is

nowadays a mature theory; however, a brief presentation is provided so that this work is as self-contained as possible.

Time-series analysis has benefited from the introduction of the concept of self-similarity by Mandelbrot and Van Ness.[24] Self-similar or fractal signals have a power spectral density (PSD) of the form

$$S(f) = \frac{S_0}{f^{\beta}}, \tag{1}$$

where S_0 is the direct current (dc) term and β is the scaling exponent, or else the slope of the power spectrum. There are two basic models used to describe self-similarity: the fractional Gaussian noise (FGN) and the fractional Brownian motion (fBm). If β is less than 1, then the signal is described by the FGN model. This would indicate a zero or positive slope in the signal spectrum meaning that the PSD increases at high frequencies. This fact is not at all common in room acoustics. The fBm model is thus a more realistic model for room acoustics signals.

For a fBm type signal the scaling exponent ranges between $1 \leq \beta \leq 3$. A simple fractal, or else a monofractal, is characterized by a parameter H known as the Hurst exponent.[25] The relation between the scaling exponent β and the Hurst exponent H is

$$H = \frac{\beta - 1}{2}, \quad 1 \leq \beta \leq 3. \tag{2}$$

The Hurst exponent ranges between 0 and 1 and quantifies persistence in a time-series. The persistence measures correlations between adjacent or distant samples within the time-series. For $H > 0.5$ (positive persistence) adjacent values are highly correlated and a value larger than the mean is more likely to be followed by an even larger value. For $H < 0.5$ (negative persistence) values are anti-correlated and large and small values are more likely to alternate. For $H = 0.5$ no long-term correlation exists.

The behavior of real time-series is rarely characterized by a single exponent. The multifractal formalism is an expanded theory for the description of more complex types of signals, which require an infinite number of exponents for their characterization. A statistical description of the range of exponents, or else singularities, found in the signal is provided by the singularity spectrum. The singularity spectrum can be computed with a number of methods. In this work, the multifractal detrended fluctuation analysis (MDFA) introduced recently by Kantelhardt *et al.*[26] is used. The method has been applied in a multitude of diverse scientific areas, such as heart beat dynamics,[27] econometrics,[28] and seismic time-series analysis.[29] The MDFA procedure consists of the following steps.

(1) Given a time-series $y(i)$, $i = 1, 2, \ldots, N$ estimate the PSD $S(f)$. In case of self-similarity, $S(f) \sim f^{\beta}$ where the exponent β must have a value $\beta > 1$. If $\beta < 1$, work with the cumulative sum of the original time-series instead.
(2) Divide the time-series into $N_s = \text{integer}(N/s)$ non-overlapping segments of length s (scale s). For each segment the local trend can be calculated by least-squares fitting the polynomial $P_n(i)$ of order n to the data. Then determine the variance as

$$F(s) = \frac{1}{s} \sum_{i=1}^{s} \{y[(n-1)s + i] - P_n(i)\}^2 \tag{3}$$

for each segment $n = 1, 2, \ldots, N_s$. Linear, quadratic, or higher order polynomials can be used for the detrending of each segment in the fitting procedure and this is best determined by trial and error.
(3) Average over all segments to obtain the qth order fluctuation for a range of moments q as

$$F_q(s) = \left\{ \frac{1}{N_s} \sum_{i=1}^{Ns} [F(s)]^{q/2} \right\}^{1/q}. \tag{4}$$

(4) If the analyzed signal features fractal properties the fluctuation obeys a power-law of the form

$$F_q(s) \sim s^{h(q)}. \tag{5}$$

The parameter $h(q)$ can be considered as a generalization of the Hurst exponent with the equivalence $H = h(2)$. If $h(q)$ is independent of q, the time-series in study is monofractal; otherwise it is multifractal.
(5) The singularity strength or Holder exponent α is subsequently computed from the equation

$$\alpha = h(q) + q \frac{dh}{dq}. \tag{6}$$

Finally, the singularity spectrum $f(\alpha)$ is computed by means of the Legendre transform

$$f(\alpha) = q[\alpha - h(q)] + 1. \tag{7}$$

IV. EXPERIMENTAL SET-UP

All tests were carried out in a room with dimensions $6.5 \times 4.2 \times 2.7$ m^3. The side walls were made of gypsum board, all other surfaces being painted concrete. The floor was covered by a light carpet. The reverberation times for the empty room are given in Table I. Four different acoustic configurations were created and tested:

(i) A very absorbent environment is referred to as the "absorbent" configuration throughout this work. 14 strips (50×2300 mm^2) and 48 patches (50×300 mm^2) of foam were distributed equally over the side walls, see Fig. 2(a). A heavy curtain covered the rear wall. The average RT in the frequency range 250–4000 Hz was 0.34 s.
(ii) A less absorbent environment with only 24 patches of foam distributed on the side walls is referred to as the "live" configuration, see Fig. 2(b). The average RT was 0.7 s.
(iii) Twenty-eight chipboard panels that covered the total surface of the side walls and a quadratic residue diffuser mounted on the rear wall. The panel's width and maximum distance from the wall are shown in Fig. 3.

TABLE I. Reverberation times and average absorption coefficients for the room configurations tested. Data for the empty room are also provided.

Frequency (Hz)		63	125	250	500	1000	2000	4000	8000
	RT	0.63s	0.55s	0.66s	0.45s	0.29s	0.26s	0.20s	0.20s
Absorbent	α	0.154	0.175	0.148	0.209	0.305	0.334	0.410	0.410
	RT	0.90	0.73	0.87	0.73	0.54	0.48	0.44	0.44
Live	α	0.111	0.135	0.114	0.135	0.178	0.197	0.213	0.213
	RT	0.67s	0.57s	0.50s	0.48s	0.51s	0.52s	0.54s	0.46s
Panels	α	0.146	0.169	0.190	0.197	0.187	0.184	0.178	0.205
	RT	0.71s	0.55s	0.46s	0.40s	0.33s	0.35s	0.31s	0.33s
Panels+foam	α	0.138	0.175	0.205	0.232	0.274	0.260	0.289	0.274
	RT	0.90s	0.78s	0.96s	0.85s	0.79s	0.74s	0.64s	0.64s
Empty	α	0.111	0.127	0.104	0.117	0.125	0.133	0.152	0.152

The diffuser was placed at a height of 1.2 m and covered approximately 30% of the area of the rear wall. The average RT was 0.50 s.

(iv) The same as (iii) with the exception that some foam was added to reduce the RT down to 0.32 s. The diffuser had been removed from the rear wall.

A wide-band Maximum Length Sequence (MLS) signal (0–10 kHz) was used as the excitation. The MLS signal was fed to a power amplifier that drove a pair of omnidirectional loudspeakers, Fig. 4. The loudspeaker frequency response was reasonably flat for the frequency range from 100 Hz up to 10 kHz. The room impulse response was recorded at 12 positions with an omnidirectional microphone. In positions 1, 3, 5, 7, 9, and 11 the microphone was set 1.5 m above the floor. In positions 2, 4, 6, 8, 10, and 12 the microphone was placed at a height of 1 m above the room floor. The sampling frequency was 30 kHz and a total of 16 384 samples were recorded providing 0.54 s of data. To reduce the effect of random noise, 20 averages have been taken in each measurement. Table I presents the RT values and average absorption coefficients for all room configurations in the octave bands from 63 Hz to 8 kHz.

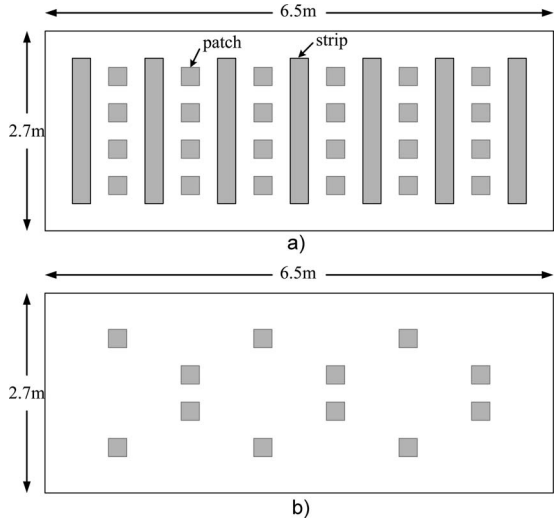


FIG. 2. (a) The absorbent configuration (side wall) and (b) the live configuration (side wall).

V. RESULTS

In Fig. 5, the room impulse responses for the absorbent and live configurations, recorded at position No. 8, are presented. The corresponding PSDs, plotted on logarithmic axes are shown in Fig. 6. It can be seen that in both cases the PSDs obey a power-law with $\beta=1.98$ and $\beta=1.61$, respectively. The Hurst exponents according to Eq. (2) are $H=0.49$ for the absorbent room and $H=0.31$ for the live room. The configurations with the panels installed show a more or less similar picture, the only difference being that the scaling exponents are lower in value. Typical values for the room with panels at position No. 8 are $\beta=1.18$ ($H=0.09$) and for the room with panels plus foam $\beta=1.36$ ($H=0.18$) at the same position. The main conclusions from the PSD graphs are that (i) the impulse responses portray self-similarity and (ii) absorption increases with frequency in the configurations treated with foam. The absorption introduced by the panels at low frequencies gives an almost uniform absorption of energy across the audio spectrum, thus resulting in a lower scaling exponent value.

It should be noted that due to its great variability, the PSD is not a reliable estimator for the Hurst exponent. The DFA provides a statistically robust estimation in the range of scales s from $10^{1.5}$ to $10^{2.8}$ (50–1000 Hz), Fig. 7. The Hurst exponents calculated from linear regression are $H=0.33$ for the absorbent configuration [Fig. 7(a)] and $H=0.09$ for the panels' configuration [Fig. 7(b)]. Although statistically stable, the DFA does not yield meaningful results if the signal is multifractal. In this case, the procedure detailed in Sec. IV should be applied and the relation between $h(q)$ and q should be examined, in particular. This procedure is detailed in Sec. V A.

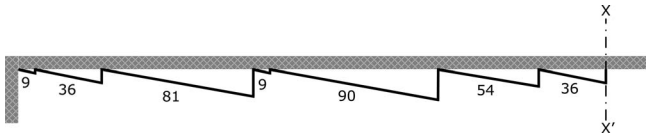


FIG. 3. The panels' configuration. Depth and width profiles of one-half of the side wall. The other half is symmetrical with respect to the X-X' axis. Dimensions are in centimeters.

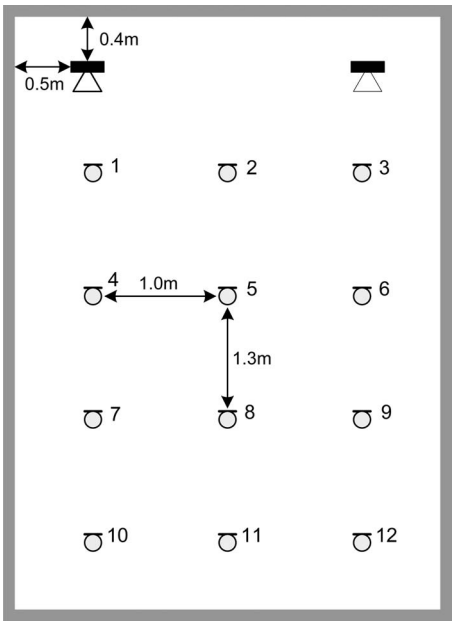


FIG. 4. Grid of measurement and loudspeaker positions.

A. Relating singularity spectra to sound-field diffuseness

Following the steps described in Sec. III the fluctuation $F_q(s)$ for a number of moments q is estimated. Then the singularity spectrum is calculated according to Eqs. (6) and (7). The fluctuation $F_q(s)$ is computed on a dense enough distribution of moments and the values obtained are interpolated so that a smooth graph of generalized exponent $h(q)$ vs moment q is obtained. This is necessary in order to carry out discrete differentiation for the estimation of the singularity spectrum. Figure 8 presents the generalized exponent $h(q)$ and the singularity spectrum $f(a)$ for the "panels'" configuration computed at position No. 12. The generalized exponent $h(q)$ shows a strong nonlinear dependence on the moment q. This behavior is typical of multifractal signals and indicates that the Hurst exponent alone is not a suitable descriptor for the structure of the room impulse responses. The singularity spectrum [Fig. 8(b)] depicts both the range of singularities and their relative strength. The singularity spectrum is convex by construction and could be interpreted as a probability density function. The peak in the spectrum corresponds to the most probable Holder exponent, which in this case is $\alpha_{peak}=0.113$. The low value is a direct result of the more or less uniform absorption across the frequency spectrum. The total range of exponents in Fig. 8 is $-0.12 < \alpha < 0.30$.

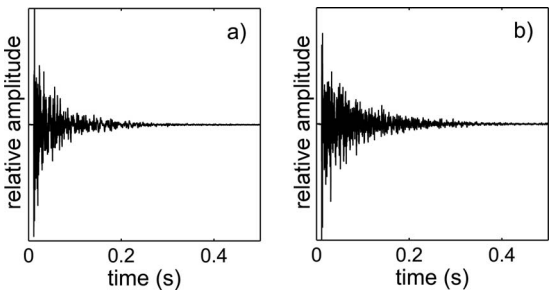


FIG. 5. Impulse responses: (a) absorbent configuration and (b) live configuration (position No. 8).

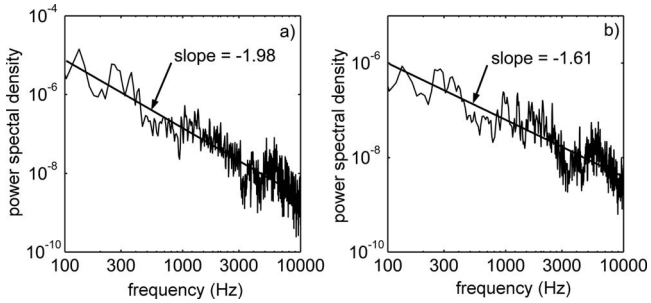


FIG. 6. PSDs and slope determined by least squares fit: (a) absorbent configuration and (b) live configuration (position No 8).

The experimental results of Fig. 8 should be compared to those for the case of the absorbent configuration, presented in Fig. 9. Again, the dependence of $h(q)$ on q is nonlinear implying a multifractal signal structure. The peak occurs at $\alpha_{peak}=0.655$. The higher value of peak exponent is to be expected, as in the room treated with foam, absorption increases with frequency. The range of exponents in the singularity spectrum is $-0.17 < \alpha < 1.34$.

The most striking difference between the spectra of Figs. 8 and 9 is the range of exponents needed to characterize the impulse response structure. In the room with the panels installed, the range of exponents is relatively small. This means that the structure of the impulse response from the sound field build up throughout its decay is relatively uniform with a low degree of correlation between samples. In the room treated with foam the correlation between samples increases with time because of the higher number of specular reflections. In the beginning, the correlation is quite low (exponent below the peak value), but as the steady state is reached the degree of correlation increases (exponent close to 1.0). The non-uniformity of the sound field in the absorbent configuration results in a broader range of exponents.

In Fig. 10, a comparison between the singularity spectra for all four configurations tested is presented (position No. 8). From left to right, the spectra correspond to configurations that change from a high degree of diffuseness to a low degree of diffuseness. The panels' configuration with the most uniform sound field portrays a narrow singularity spectrum. The introduction of foam to the room limits the total

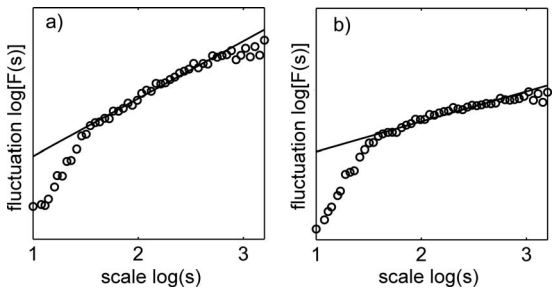


FIG. 7. Fluctuation vs scale obtained with DFA and best linear fit: (a) absorbent configuration and (b) panels' configuration. Very low frequencies are ignored in the slope determination.

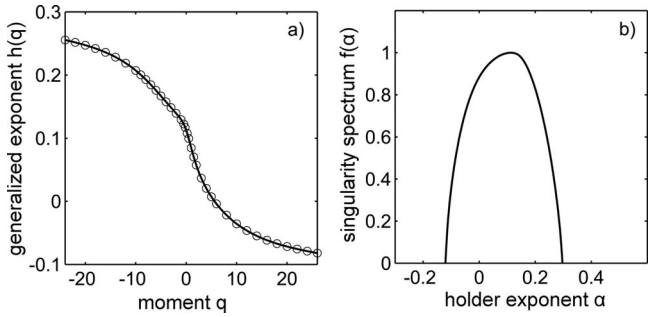


FIG. 8. (a) Generalized exponent h vs moment q and (b) singularity spectrum for the panels' configuration.

number of reflections and in addition provides a number of almost specular reflections as some of the panel area is replaced with foam. For this reason, the singularity spectrum for the "panels+foam" configuration is broader. Moving on further, we notice that in a room of small dimensions with a minimum amount of absorptive material the sound field is far from uniform (live room). Whatever amount of sound diffusion is achieved is due to the longer reverberation, or else the increased number of mostly specular reflections that take place. Finally, in the absorbent configuration, the absorbent material severely limits the total number of reflections from the buildup of the sound field to its final decay. The nonuniformity of the sound field results in a wide range of exponents and a very broad singularity spectrum.

B. Variation of singularity spectrum with position

In Sec. V A, the structure of the impulse response was examined at a particular position. The variation of the singularity spectra with position is studied in this paragraph.

Figure 11 presents the strongest Holder exponent (peak in the singularity spectrum) for all room configurations. The extreme values for the Holder exponent are also given in the form of error bars. It is evident that the singularity spectra depend to some extent on the position at which the impulse response was recorded. However, the variation due to position is small compared to the variation due to a change in room characteristics. This, in principle, allows us to make an estimate for the diffuseness of the sound field in the room just from a single measurement. The justification for this is that diffuse reflections do not only produce spatial scattering, but also a smearing of the reflections in time. This statement seems to be valid for the particular room examined, but it

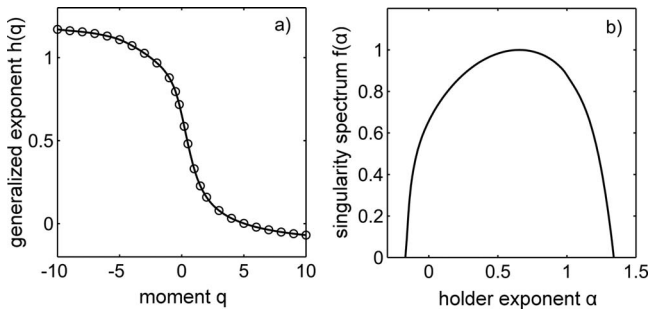


FIG. 9. (a) Generalized exponent h vs moment q and (b) singularity spectrum for the absorbent configuration.

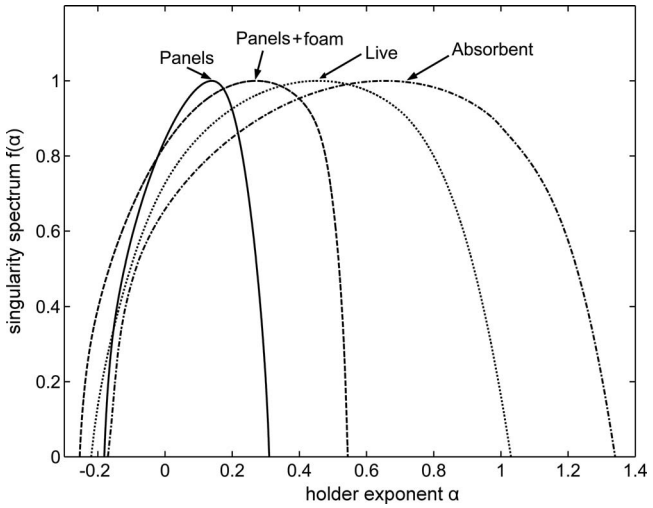


FIG. 10. A comparison of the singularity spectra for all four configurations tested (position No. 8). A narrow spectrum indicates increased diffuseness.

may not hold up in bigger or non-mixing spaces. Clearly, further work is needed in order to examine the conditions under which sound-field diffuseness estimation from a single measurement is possible.

Figure 11 demonstrates that there is a progressive broadening of spectrum toward higher exponents when moving from a state of high sound-field diffuseness (panels) to a state of lower diffuseness (foam absorption). An additional observation is that irrespective of the configuration, the Holder exponent is higher at position Nos. 2, 5, 8, and 11. This in turn means that absorption is higher at the room median plane and lower at positions close to the side walls. As there is an inverse relationship between absorption and diffuseness, the sound field is expected to be less correlated in positions close to the side walls.

Based on the results of Sec. V B, it is now possible to provide a definition for a diffuse field that is not based on energy density. A perfectly diffuse field is one that has a

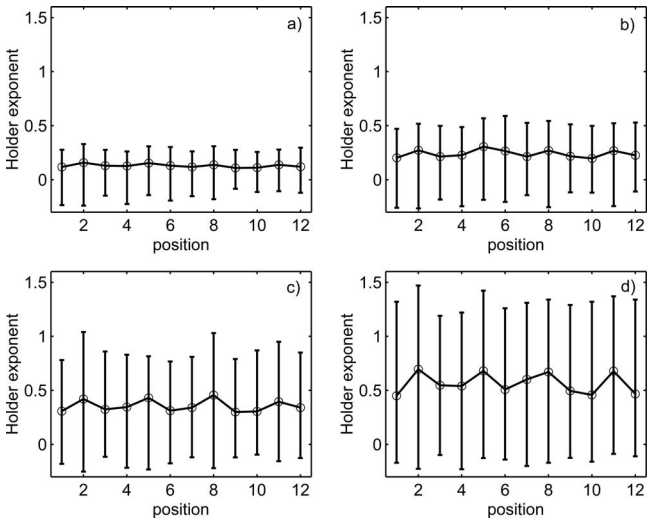


FIG. 11. Variation of singularity spectra with position where the circle denotes peak exponent value: (a) panels, (b) panels plus foam, (c) live room, and (d) absorbent room.

TABLE II. Mean values of minimum, peak, and maximum Holder exponent for the room configurations tested and corresponding diffuseness measure.

Room configuration	α_{min}	α_{peak}	α_{max}	d_m
Panels	−0.162	0.129	0.286	2.23
Panels+foam	−0.194	0.240	0.522	1.39
Live	−0.167	0.356	0.866	0.97
Absorbent	−0.154	0.566	1.321	0.68

singularity spectrum of zero width at every position within the room volume. In order to formally quantify sound-field diffuseness, the *diffuseness measure* is introduced here, defined as

$$d_m = \frac{1}{\alpha_{max} - \alpha_{min}}, \tag{8}$$

where α_{max} and α_{min} are the maximum and minimum exponents in the spectrum calculated as the average for all positions of measurement. Table II presents the basic statistics for all four configurations tested. The extreme values for parameter d_m are 2.23 for the panels' room and 0.68 for the absorbent room.

VI. CONCLUDING REMARKS

With the aim of quantifying sound-field diffuseness in spaces of relatively low volume, four different configurations in terms of acoustic treatment were constructed and tested. Two of the configurations achieved a higher degree of diffuseness due to the scattering provided by specially constructed panels. The other two configurations were treated with absorptive material (foam), which resulted in a less uniform sound field. The sound field in the room was sampled at 12 positions and the corresponding impulse responses were analyzed with the MDFA method in order to obtain the singularity spectrum. It has been shown that the singularity spectra contain a wealth of information for the acoustic field established in the room. The width of the singularity spectrum, in particular, is directly related to the impulse response complexity. An ideally diffuse field yields a singularity spectrum of zero width, or in other words a spectrum that corresponds to a monofractal signal. From there on, the less diffuse the sound field is, the more the singularity spectrum width is increasing.

The proposed method has been proved to work in small rooms and is therefore suitable for the estimation of the sound-field structure in control rooms and listening rooms. For the cases under study, an estimate about sound-field diffuseness can be made just from a single measurement. This, however, may not be possible in bigger spaces or non-mixing environments. A diffuseness measure has been proposed in order to quantify sound-field diffuseness. It has been found that rooms with approximately the same reverberation time can vary significantly in terms of the complexity of their sound field.

The experimental work reported in this study was carried out in a room of relatively small dimensions; however, there is no specific reason why this methodology would not work in bigger spaces, such as concert halls or multipurpose rooms. This, however, remains to be determined in future effort. Further work could be also carried out in the direction of quantifying diffuseness in specific frequency bands. This could be done, for example, by filtering the impulse response in 1/3 octave bands and subsequently deriving the diffuseness measure for each band.

ACKNOWLEDGMENTS

The author is grateful to the anonymous reviewers whose constructive comments greatly helped in improving the clarity of the original manuscript.

[1] L. L. Beranek, *Music, Acoustics and Architecture* (Wiley, New York, 1962).
[2] L. Beranek, *Acoustics* (Acoustical Society of America, New York, 1981).
[3] J. L. Davy, W. A. Davern, and P. Dubout, "Qualification of room diffusion for absorption measurements," Appl. Acoust. **28**, 177–185 (1989).
[4] K. Bondlund, "A study of diffusion in reverberation chambers provided with special devices," J. Sound Vib. **50**, 253–283 (1977).
[5] K. Randal and F. Ward, "Diffusion of sound in small rooms," Proc. IEEE **107B**, 439–450 (1960).
[6] P. D'Antonio and J. Konnert, "The RPG reflection phase grating acoustical diffusor: Design theory and application," J. Audio Eng. Soc. **32**, 228–238 (1984).
[7] B. L. Dalenback, M. Kleiner, and P. Svensson, "A macroscopic view of diffuse reflections," J. Audio Eng. Soc. **42**, 793–807 (1994).
[8] Y. Jing and N. Xiang, "On boundary conditions for the diffusion equation in room-acoustic prediction: Theory, simulations and experiments," J. Acoust. Soc. Am. **123**, 145–153 (2008).
[9] M. Hodgson, "When is diffuse-field theory applicable?," Appl. Acoust. **49**, 197–207 (1996).
[10] A. Kulowski, "Remarks on a limit value of the sound directional diffusion coefficient in rooms," Appl. Acoust. **32**, 93–105 (1991).
[11] B. N. Gover, J. G. Ryan, and M. R. Stinson, "Measurements of directional properties of reverberant sound fields in rooms using a spherical microphone," J. Acoust. Soc. Am. **116**, 2138–2148 (2004).
[12] N. Spring and K. Randall, "The measurement of the sound diffusion in small rooms," BBC Research Department Report No. 1969/16, British Broadcasting Corp., London, 1969.
[13] T. Schultz, "An acoustic watt-meter," J. Acoust. Soc. Am. **28**, 693–697 (1956).
[14] R. Bolt and R. Roop, "Frequency response fluctuations in rooms," J. Acoust. Soc. Am. **22**, 280–289 (1950).
[15] M. Schroeder and K. Kuttruff, "Frequency correlation functions of frequency responses in rooms," J. Acoust. Soc. Am. **34**, 1819–1824 (1962).
[16] R. Cook, R. Waterhouse, R. Berent, and M. Thompson, "Measurement of correlation coefficients in reverberant sound fields," J. Acoust. Soc. Am. **27**, 1072–1077 (1955).
[17] C. G. Balachandran, "Random sound field in reverberation chambers," J. Acoust. Soc. Am. **31**, 1319–1321 (1959).
[18] H. Nelisse and J. Nicolas, "Characterization of a diffuse field in a reverberation chamber," J. Acoust. Soc. Am. **101**, 3517–3524 (1997).
[19] F. Jacobsen and T. Roisin, "The coherence of reverberant sound fields," J. Acoust. Soc. Am. **108**, 204–210 (2000).
[20] M. R. Schroeder, D. Gottlob, and K. F. Siebrasse, "Comparative study of European concert halls: Correlation of subjective preference with geometric and acoustic parameters," J. Acoust. Soc. Am. **56**, 1195–1201 (1974).
[21] Y. Suzumura, M. Sakurai, and Y. Ando, "An evaluation of the effects of scattered reflections in a sound field," J. Sound Vib. **232**, 303–308 (2000).
[22] J. K. Ryu and J. Y. Jeon, "Subjective and objective evaluations of a scattered sound field in a scale opera house," J. Acoust. Soc. Am. **124**, 1538–1549 (2008).
[23] C. Haan and F. R. Fricke, "An evaluation of the importance of surface diffusivity in concert halls," Appl. Acoust. **51**, 53–69 (1997).
[24] B. B. Mandelbrot and J. W. Van Ness, "Fractional Brownian motions, fractional noises and applications," SIAM Rev. **10**, 422–437 (1968).
[25] H. E. Hurst, "Long-term storage capacity of reservoirs," Trans. Am. Soc. Civ. Eng. **116**, 770–808 (1951).
[26] J. W. Kantelhardt, S. A. Zschiegner, E. Koscielny-Bunde, S. Havlin, A. Bunde, and H. E. Stanley, "Multifractal detrended fluctuation analysis of nonstationary time series," Physica A **316**, 87–114 (2002).

[27]D. Makowiec, R. Gała¸skab, A. Dudkowskaa, A. Rynkiewiczb, and M. Zwierza, "Long-range dependencies in heart rate signals—revisited," Physica A **369**, 632–644 (2006).
[28]P. Oświe¸cimka, J. Kwapień, and S. Drożdż, "Multifractality in the stock market: price increments versus waiting times," Physica A **347**, 626–638 (2005).
[29]L. Telesca and V. Lapenaa, "Measuring multifractality in seismic sequences," Tectonophysics **423**, 115–123 (2006).

Analytical and experimental investigation on transmission loss of clamped double panels: Implication of boundary effects

F. X. Xin and T. J. Lu[a)]
MOE Key Laboratory for Strength and Vibration, School of Aerospace, Xi'an Jiaotong University, Xi'an 710049, People's Republic of China

(Received 28 August 2008; revised 5 January 2009; accepted 6 January 2009)

The air-borne sound insulation performance of a rectangular double-panel partition clamp mounted on an infinite acoustic rigid baffle is investigated both analytically and experimentally and compared with that of a simply supported one. With the clamped (or simply supported) boundary accounted for by using the method of modal function, a double series solution for the sound transmission loss (STL) of the structure is obtained by employing the weighted residual (Galerkin) method. Experimental measurements with Al double-panel partitions having air cavity are subsequently carried out to validate the theoretical model for both types of the boundary condition, and good overall agreement is achieved. A consistency check of the two different models (based separately on clamped modal function and simply supported modal function) is performed by extending the panel dimensions to infinite where no boundaries exist. The significant discrepancies between the two different boundary conditions are demonstrated in terms of the STL versus frequency plots as well as the panel deflection mode shapes. © *2009 Acoustical Society of America.*
[DOI: 10.1121/1.3075766]

PACS number(s): 43.55.Ti, 43.20.Tb, 43.55.Rg [AE: NX] Pages: 1506–1517

I. INTRODUCTION

With superior sound insulation properties over single-panel configurations, double-panel partitions have found a wide range of important applications in modern buildings, transportation vehicles, aerospace and aeronautical structures, etc.[1–6] To gain a fundamental understanding of the sound insulation mechanisms of double-panel partitions, the frequency characteristics of sound transmission loss (STL) are usually needed. In particular, how these are affected by different boundary conditions (i.e., clamped and simply supported) is of great theoretical and practical interest. The present investigation aims to address the significant differences between the two different boundary conditions for double-panel partitions containing air cavities, both theoretically and experimentally.

For decades the vibration responses of single- and double-panel constructions interacting with the surrounding fluid have been an attractive research topic. The approaches adopted to study the vibroacoustic behavior of both structures are similar, although there are more difficulties associated with the latter. Traditionally, the method of statistical energy analysis (SEA) advanced significantly by Maidanik[7] has been widely used to analyze the vibration response of a complex structure under force or sound excitation. However, the SEA method is less effective at relatively low frequencies on account of its pre-assumption that enough structural modes need to be excited. This is usually difficult to satisfy, causing statistical uncertainties that prevail in low frequency force and sound excitations.[8] A finite element method model was developed by Ruzzene[9] to evaluate the acoustic characteristics of sandwich beams in terms of structural response and sound transmission reduction index, which is more effective for low frequencies but requires high computational cost for high frequencies.[10] For relatively simple structures, analytical solutions suited for a wide frequency range have been developed by various researchers.[11–13]

As for the sound insulation properties of double-panel structures, the classical work of London[14] addressed an infinite double-panel structure, and hence the influence of boundary conditions was not considered. Similarly, the study of Kropp and Rebillard[15] and Antonio *et al.*[16] focused on the air-borne sound insulation capability of infinite double wall constructions. The sound insulation property and radiation efficiency of an infinite double-plate connected by periodical studs have been investigated with the Fourier transform technique[17,18] and with the space-harmonic expansion method,[19,20] respectively. Brunskog[21] examined the influence of finite cavities on the sound insulation properties of periodically framed infinite double-plate structures. To take into account the finite size of a double-panel structure, Villot *et al.*[22] developed an approximate technique based on the spatial windowing of plane waves. More recently, the problem of sound transmission through double-panel structures of finite extent was solved by considering simple boundary conditions[8,23] on the basis of modal superposition theory from different viewpoints. In addition, to improve the sound isolation properties of double-panel partitions, various active control strategies have been proposed, both experimentally[1,24] and theoretically,[1,25] which are significant from the viewpoint of practical noise control.

Although persistent effort has been devoted to the study of sound transmission through finite or infinite single- and double-leaf panels, many physical details remain an indistinct matter, especially for *clamped* double-panel configurations under oblique sound excitation. Often, the experimental

[a)]Electronic mail: tjlu@mail.xjtu.edu.cn

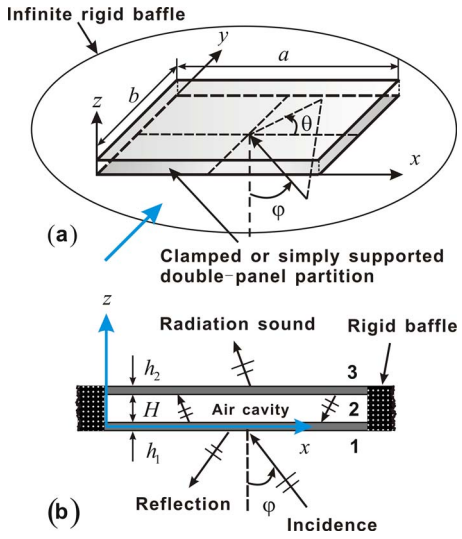


FIG. 1. (Color online) Schematic illustration of sound transmission through a baffled double-panel partition, which is clamped (or simply supported) on its edges: (a) global view; (b) side view in the arrow direction in (a).

measurements do not possess reproducibility, and different STL curves from different laboratories were obtained even though the same panels were compared.[26] According to Kim *et al.*,[27] this may be attributed to the so-called tunneling effect, and the different mounting conditions adopted in different laboratories should be another key factor. On the other hand, to the authors' best knowledge, the problem of sound transmission through a double-panel partition with air cavity fully clamped (different from being simply supported) on its edges has not been analytically solved. The existing method[1,12] for the influence of the clamped boundary condition is to modify the analytical solution for simply supported boundary conditions, which is only approximate for predicting the sound insulation properties of fully clamped finite single- or double-panel constructions. The present paper squarely addresses this deficiency. A step by step analysis for the vibroacoustic performance of a finite double-panel partition fully clamped on its edges under sound excitation is presented (note that aspects of the current theoretical formulations have been presented as a simplified outline in Ref. 6). To gain full insight into the influence of the boundary condition on sound transmission across the structure, STL values obtained with the clamped boundary condition are compared with those obtained with the simply supported boundary condition. Experimental measurements for both types of boundary condition are subsequently carried out to validate the model predictions. The remarkable difference between the two different boundary conditions is highlighted. As sound transmission through a double-panel partition simply supported on its edges has been extensively studied, the present focus is placed upon exploring how this is different from that of a fully clamped double-panel partition.

II. DESCRIPTION OF THE VIBROACOUSTIC STRUCTURE

Consider a rectangular double-panel partition with air cavity, which is baffled and fully clamped (or simply supported) along its edges, as shown in Fig. 1(a). The two panels are taken as homogenous, isotropic, and sufficiently thin. The geometrical dimensions of the structure are width of panel a, length of panel b, depth of air cavity H, thicknesses of incidence panel h_1, and radiating panel h_2 [Fig. 1(b)]. The whole configuration is fully clamped (or simply supported) on an infinite acoustic rigid baffle, which separates the space into two fields: sound incidence field ($z<0$) and sound radiating field ($z>H$). Cartesian coordinates (x,y,z) are selected, as shown in Fig. 1.

An oblique plane sound wave varying harmonically in time is incident on the bottom panel with elevation angle φ and azimuth angle θ. The vibration of the bottom panel induced by the incident sound is transmitted through the hermetical air cavity to the upper panel, which radiates sound pressure waves into the upper acoustic domain [Fig. 1(b)]. The model proposed below describes the vibroacoustic behavior of the double-panel structure, either fully clamped or simply supported, and its STL characteristics.

III. MODELING OF THE VIBROACOUSTIC COUPLED SYSTEM

A. Theoretical formulation and solution

The acoustic velocity potential for a plane sound wave varying harmonically in time can be expressed as

$$\phi = Ie^{-j(\mathbf{k}\cdot\mathbf{r}-\omega t)}, \tag{1}$$

where I is the sound amplitude, ω is the angular frequency, $\mathbf{r}(=x\hat{\mathbf{e}}_x+y\hat{\mathbf{e}}_y+z\hat{\mathbf{e}}_z)$ is the position vector (Fig. 1) with $\hat{\mathbf{e}}_x$, $\hat{\mathbf{e}}_y$, and $\hat{\mathbf{e}}_z$ representing separately the unit vectors along x-, y-, and z-directions, and $\mathbf{k}(=k_x\hat{\mathbf{e}}_x+k_y\hat{\mathbf{e}}_y+k_z\hat{\mathbf{e}}_z)$ is the wave vector with components k_x, k_y, and k_z. These wave numbers are determined by the elevation angle φ and azimuth angle θ of the incidence sound wave as

$$k_x = k_0 \sin\varphi\cos\theta, \quad k_y = k_0 \sin\varphi\sin\theta, \quad k_z = k_0\cos\varphi, \tag{2}$$

where $k_0=\omega/c_0$ is the acoustic wave number in air and c_0 is the acoustic speed in air.

The flexural motions of a double-panel partition with air cavity induced by sound excitation [Fig. 1(b)] are governed by

$$D_1\nabla^4 w_1 + m_1\frac{\partial^2 w_1}{\partial t^2} - j\omega\rho_0(\Phi_1-\Phi_2) = 0, \tag{3}$$

$$D_2\nabla^4 w_2 + m_2\frac{\partial^2 w_2}{\partial t^2} - j\omega\rho_0(\Phi_2-\Phi_3) = 0, \tag{4}$$

where $\nabla^4=(\partial^2/\partial x^2+\partial^2/\partial y^2)^2$, ρ_0 is the air density, $j=\sqrt{-1}$, w_1 and w_2 are the transverse displacements, m_1 and m_2 are the mass per unit area, and D_1 and D_2 are the flexural rigidity of the bottom panel (panel 1) and the upper panel (panel 2) located at $z=0$ and $z=H$, respectively. Damping of the panel material is taken into account by introducing the complex Young's modulus, $E_i(1+j\eta_i)$, where η is the loss factor. The flexural rigidity of the panel $D_i(i=1,2)$ can thence be written as

$$D_i = \frac{E_i h_i^3 (1 + j\eta_i)}{12(1 - v_i^2)}. \tag{5}$$

Let $\Phi_i (i=1,2,3)$ denote the velocity potentials for the acoustic fields in the proximity of the two panels, corresponding to the sound incidence, the air cavity, and the structure radiating field [fields 1, 2, and 3 in Fig. 1(b)], respectively. The acoustic velocity potential in the incidence field [field 1, Fig. 1(b)] is defined as

$$\Phi_1(x,y,z;t) = Ie^{-j(k_x x + k_y y + k_z z - \omega t)} + \beta e^{-j(k_x x + k_y y - k_z z - \omega t)}, \tag{6}$$

where the first term represents the velocity potential of the incident acoustic wave and the second term represents the sum velocity potential of the reflected and the radiating acoustic waves, and I and β are the amplitudes of the incident (i.e., positive-going) and the reflected plus radiating (i.e., negative-going) waves, respectively. Similarly, the velocity potential in the air cavity [field 2, Fig. 1(b)] can be written as

$$\Phi_2(x,y,z;t) = \varepsilon e^{-j(k_x x + k_y y + k_z z - \omega t)} + \zeta e^{-j(k_x x + k_y y - k_z z - \omega t)}, \tag{7}$$

where ε is the amplitude of the positive-going wave and ζ is the amplitude of the negative-going wave. In the transmitting field [field 3, Fig. 1(b)] adjacent to the radiating upper panel, there exist no reflected waves, and therefore the velocity potential is only for the transmitting (or radiating) waves, given as

$$\Phi_3(x,y,z;t) = \xi e^{-j(k_x x + k_y y + k_z z - \omega t)}, \tag{8}$$

where ξ is the amplitude of the radiating (i.e., positive-going) wave. These velocity potentials are related to the acoustic particle velocities by $\hat{\mathbf{u}}_i = -\nabla \Phi_i$ and to the sound pressure by

$$p_i = \rho_0 \frac{\partial \Phi_i}{\partial t} = j\omega\rho_0 \Phi_i \quad (i=1,2,3). \tag{9}$$

With the double-panel partition fully clamped onto a rigid baffle, the transverse deflection and the moment rotation of each panel are constrained to be zero along the edges. In view of the rectangular geometry of the double-panel structure, the boundary conditions can be expressed as

$$x = 0, a, \quad w_1 = w_2 = 0, \quad \frac{\partial w_1}{\partial x} = \frac{\partial w_2}{\partial x} = 0, \tag{10}$$

$$y = 0, b, \quad w_1 = w_2 = 0, \quad \frac{\partial w_1}{\partial y} = \frac{\partial w_2}{\partial y} = 0. \tag{11}$$

At the air-panel interface the normal velocity is continuous, yielding the corresponding velocity compatibility condition equations

$$z = 0, \quad -\frac{\partial \Phi_1}{\partial z} = j\omega w_1, \quad -\frac{\partial \Phi_2}{\partial z} = j\omega w_1, \tag{12}$$

$$z = H, \quad -\frac{\partial \Phi_2}{\partial z} = j\omega w_2, \quad -\frac{\partial \Phi_3}{\partial z} = j\omega w_2. \tag{13}$$

Since the two panels are excited by a harmonic sound wave, their transverse displacements can be written as

$$w_1(x,y,t) = \sum_{m,n} \phi_{mn}(x,y) q_{1,mn}(t), \tag{14}$$

$$w_2(x,y,t) = \sum_{m,n} \phi_{mn}(x,y) q_{2,mn}(t), \tag{15}$$

where the modal functions (or, more strictly speaking, the basic functions) ϕ_{mn} and the modal displacements $q_{i,mn}$ take the following forms:

$$\phi_{mn}(x,y) = \left(1 - \cos\frac{2m\pi x}{a}\right)\left(1 - \cos\frac{2n\pi y}{b}\right), \tag{16}$$

$$q_{1,mn}(t) = \alpha_{1,mn} e^{j\omega t}, \quad q_{2,mn}(t) = \alpha_{2,mn} e^{j\omega t}, \tag{17}$$

where $\alpha_{1,mn}$ and $\alpha_{2,mn}$ are the modal coefficients of the bottom panel and the upper panel, respectively. Note that the clamped modal function of Eq. (16) is different from the simply supported modal function $\phi^s_{mn} = \sin(m\pi x/a) \times \sin(n\pi y/b)$ used by previous researchers[2,8,23,28–31] because the former satisfies the boundary condition of zero moment rotation while the latter does not. Moreover, the clamped double-panel can transmit rotation on its edges, whereas the simply supported one cannot, which has been confirmed experimentally by Utley and Fletcher.[32]

By applying the modal functions for the clamped double-panel structure, the velocity potentials for the acoustic fields 1–3 [Fig. 1(b)] can be expressed as

$$\Phi_1(x,y,z,t) = \sum_{m,n} I_{mn}\phi_{mn} e^{-j(k_z z - \omega t)} + \sum_{m,n} \beta_{mn}\phi_{mn} e^{-j(-k_z z - \omega t)}, \tag{18}$$

$$\Phi_2(x,y,z,t) = \sum_{m,n} \varepsilon_{mn}\phi_{mn} e^{-j(k_z z - \omega t)} + \sum_{m,n} \zeta_{mn}\phi_{mn} e^{-j(-k_z z - \omega t)}, \tag{19}$$

$$\Phi_3(x,y,z,t) = \sum_{m,n} \xi_{mn}\phi_{mn} e^{-j(k_z z - \omega t)}, \tag{20}$$

where the coefficients I_{mn}, β_{mn}, ε_{mn}, ζ_{mn}, and ξ_{mn} are determined by the Fourier cosine transform

$$\chi_{mn} = \frac{4}{ab}\int_0^b \int_0^a \chi e^{-j(k_x x + k_y y)} \cos\frac{2m\pi x}{a}\cos\frac{2n\pi y}{b} dx dy. \tag{21}$$

Here, the symbol χ can be referred to any of the coefficients I, β, ε, ζ, and ξ.

Substituting Eqs. (18)–(20) into Eqs. (12) and (13) for the continuity of velocity at the air-panel interface and omitting the time factor $e^{j\omega t}$, one obtains

$$-Ie^{-j(k_x x + k_y y)} + \sum_{m,n}\left[\beta_{mn} + \frac{\omega}{k_z}\alpha_{1,mn}\right] \times \phi_{mn}(x,y) = 0, \tag{22}$$

$$\sum_{m,n}[k_z(-\varepsilon_{mn}+\zeta_{mn})+\omega\alpha_{1,mn}]\times\phi_{mn}(x,y)=0, \tag{23}$$

$$\sum_{m,n}[k_z(-\varepsilon_{mn}e^{-jk_zH}+\zeta_{mn}e^{jk_zH})+\omega\alpha_{2,mn}]\times\phi_{mn}(x,y)=0, \tag{24}$$

$$\sum_{m,n}[-k_z\xi_{mn}e^{-jk_zH}+\omega\alpha_{2,mn}]\times\phi_{mn}(x,y)=0. \tag{25}$$

According to the weighted residual (Galerkin) method, by setting the integral of a weighted residual of the modal function to zero, an arbitrarily accurate double series solution can be obtained. For the current double-leaf partition system, the integral equations are

$$\int_0^b\int_0^a\left[D_1\nabla^4w_1+m_1\frac{\partial^2w_1}{\partial t^2}-j\omega\rho_0(\Phi_1-\Phi_2)\right]\times\phi_{mn}(x,y)dxdy=0, \tag{26}$$

$$\int_0^b\int_0^a\left[D_2\nabla^4w_2+m_2\frac{\partial^2w_2}{\partial t^2}-j\omega\rho_0(\Phi_2-\Phi_3)\right]\times\phi_{mn}(x,y)dxdy=0. \tag{27}$$

Substituting Eqs. (14), (15), and (18)–(20) into Eqs. (26) and (27) and then performing laborious but straightforward algebraic manipulations, one gets

$$4D_1\pi^4ab\left\{\left[3\left(\frac{m}{a}\right)^4+3\left(\frac{n}{b}\right)^4+2\left(\frac{m}{a}\right)^2\left(\frac{n}{b}\right)^2\right]\alpha_{1,mn}+\sum_k 2\left(\frac{n}{b}\right)^4\alpha_{1,kn}+\sum_l 2\left(\frac{m}{a}\right)^4\alpha_{1,ml}\right\}+\frac{9ab}{4}Q_{1,mn}+\frac{3ab}{2}\sum_k Q_{1,kn}+\frac{3ab}{2}\sum_l Q_{1,ml}+ab\sum_{k,l}Q_{1,kl}=2j\omega\rho_0If_{mn}(k_x,k_y)\quad\text{at}\quad z=0\ (k\neq m,l\neq n), \tag{28}$$

$$4D_2\pi^4ab\left\{\left[3\left(\frac{m}{a}\right)^4+3\left(\frac{n}{b}\right)^4+2\left(\frac{m}{a}\right)^2\left(\frac{n}{b}\right)^2\right]\alpha_{2,mn}+\sum_k 2\left(\frac{n}{b}\right)^4\alpha_{2,kn}+\sum_l 2\left(\frac{m}{a}\right)^4\alpha_{2,ml}\right\}+\frac{9ab}{4}Q_{2,mn}+\frac{3ab}{2}\sum_k Q_{2,kn}+\frac{3ab}{2}\sum_l Q_{2,ml}+ab\sum_{k,l}Q_{2,kl}=0\quad\text{at}\quad z=H\ (k\neq m,l\neq n), \tag{29}$$

where

$$Q_{1,mn}=-m_1\omega^2\alpha_{1,mn}+j\omega\rho_0\left(\frac{\omega}{k_z}\alpha_{1,mn}e^{jk_zz}+\varepsilon_{mn}e^{-jk_zz}+\zeta_{mn}e^{jk_zz}\right), \tag{30}$$

$$Q_{2,mn}=-m_2\omega^2\alpha_{2,mn}-j\omega\rho_0[(\varepsilon_{mn}-\xi_{mn})e^{-jk_zz}+\zeta_{mn}e^{jk_zz}]. \tag{31}$$

In the above expressions, the abbreviated symbols $\Sigma_{k,l}$, Σ_k, and Σ_l denote separately $\Sigma_{k=1}^{\infty}\Sigma_{l=1}^{\infty}$, $\Sigma_{k=1}^{\infty}$, and $\Sigma_{l=1}^{\infty}$ (the same can be said of $\Sigma_{m,n}$), and $f_{mn}(k_x,k_y)$ is a constant generated in the process of integration, expressed as

$$f_{mn}(k_x,k_y)=\begin{cases}ab & \text{for } k_x=0 \text{ and } k_y=0\\ \dfrac{4jn^2\pi^2a(1-e^{-jk_yb})}{k_y(k_y^2b^2-4n^2\pi^2)} & \text{for } k_x=0 \text{ and } k_y\neq 0\\ \dfrac{4jm^2\pi^2b(1-e^{-jk_xa})}{k_x(k_x^2a^2-4m^2\pi^2)} & \text{for } k_x\neq 0 \text{ and } k_y=0\\ -\dfrac{16m^2n^2\pi^4(1-e^{-jk_xa})(1-e^{-jk_yb})}{k_xk_y(k_x^2a^2-4m^2\pi^2)(k_y^2b^2-4n^2\pi^2)} & \text{for } k_x\neq 0 \text{ and } k_y\neq 0\end{cases}. \tag{32}$$

Together with Eqs. (22)–(25), Eqs. (28) and (29) form a set of infinite algebraic simultaneous equations for the unknown coefficients $\alpha_{1,mn}$ and $\alpha_{2,mn}$. For numerical calculation, it is necessary to take truncation at $1\leq m\leq M$ and $1\leq n\leq N$, leading to $2MN$ algebraic simultaneous equations. In matrix form, these can be grouped into

$$\begin{bmatrix}T_{11,kl} & T_{12,kl}\\ T_{21,kl} & T_{22,kl}\end{bmatrix}_{2MN\times 2MN}\begin{Bmatrix}\alpha_{1,kl}\\ \alpha_{2,kl}\end{Bmatrix}_{2MN\times 1}=\begin{Bmatrix}F_{kl}\\ 0\end{Bmatrix}_{2MN\times 1}, \tag{33}$$

where k and l take values from 1 to M and from 1 to N, respectively. Detailed derivations of Eq. (33) can be found in the Appendix. Once the unknowns $\alpha_{1,mn}$ and $\alpha_{2,mn}$ are determined by solving Eq. (33), the deflections (w_1,w_2) of the bottom and upper panels and the relevant parameters $(\beta_{mn},\varepsilon_{mn},\zeta_{mn},\xi_{mn})$ are also determined, enabling thence the analysis of sound transmission across the double-leaf configuration.

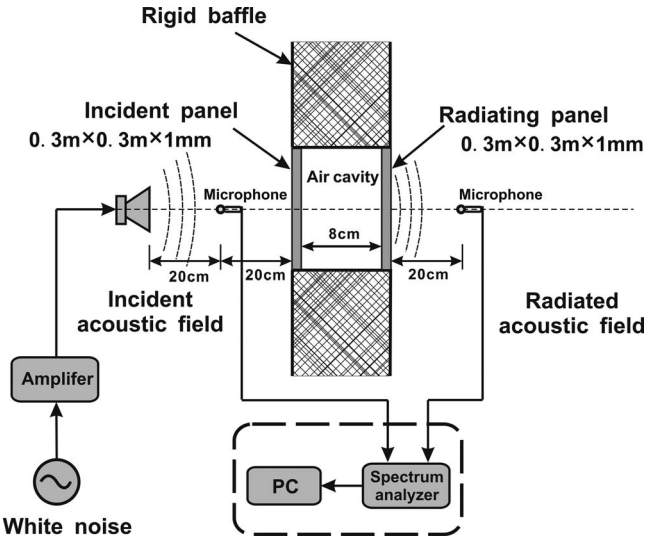


FIG. 2. Schematic of experimental setup for STL measurements of fully simply supported or fully clamped double-panel partition.

B. Definition of sound transmission loss

The sound power of the relevant acoustic field can be defined as[6,23,33]

$$\prod_i = \frac{1}{2}\mathrm{Re}\int\int_A p_i \cdot v_i^* dA \quad (i=1,2,3), \tag{34}$$

where the local volume velocity is associated with the sound pressure through the impendence of air as $v_i = p_i/(\rho_0 c_0)$. The superscript asterisk denotes the complex conjugate.

The power transmission coefficient that is a function of the incident angle (φ and θ) can be given by the ratio of the transmitted sound power to the incident sound power:

$$\tau(\varphi,\theta) = \frac{\Pi_3}{\Pi_1}. \tag{35}$$

Then the STL is defined as the inverse of the power transmission coefficient in decibel scale, given by

$$\mathrm{STL} = 10\log_{10}\left(\frac{1}{\tau}\right). \tag{36}$$

The STL index is commonly used as a measure of the effectiveness of the double-panel structure in isolating the incident sound.

IV. SOUND TRANSMISSION MEASUREMENTS

A. Experimental setup

To validate the proposed theoretical model, STL measurements for fully clamped and simply supported double-panel partitions are separately carried out. The experimental setup is schematically illustrated in Fig. 2, while Fig. 3 presents more details for the clamped case. The transmission loss measurements are performed by utilizing a pressure method.[34,35] Two condenser microphones (Knowles: FG-23742-150, diameter $d=2.59$ mm) are located on the incident side and the radiated side, respectively, with the same

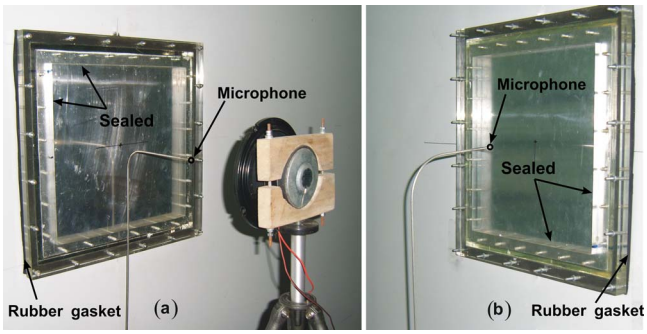


FIG. 3. (Color online) Experimental setup for STL measurements of fully clamped double-panel partition: (a) incident side; (b) transmitted side.

distance 20 cm from the corresponding panel at the centerline of the structure. The microphone located on the radiated side can give a representative sound pressure level of the whole radiated field only when the microphone is placed normal to the planar structure.[36] The tested Al double panels are parallel and clamp (or simply) mounted on a large sandwich panel composed of two steel panels with thickness 2 mm each and heavy asbestos blanket in between as the core. As the steel sandwich panel has much superior sound insulation capability than the tested double-panel partition and is significantly lager than the Al double-panel in size, it may be regarded as an infinite acoustic rigid baffle. The source room and the receiving room are both semianechoic, enabling as ideal a normal incident sound as possible in the source room and the measuring of the sound pressure completely radiated from the tested structure in the receiving room.

The practical implementation of the two different boundary conditions is demonstrated in Fig. 4. Both the simply supported and the clamped fixtures for mounting the test panels are Plexiglass made and are firmly fastened onto the rigid sandwich panel with bolts around the perimeters; rubber gaskets are used as intermediates to minimize sound leaking during the measurements (see Fig. 3). Figures 4(a) and 4(b) give the front view and the side view of the simply supported fixture, while Figs. 4(c) and 4(d) present the front view and the side view of the clamped fixture, respectively.

To implement the simply supported boundary condition, the Al panel is mounted around its perimeter on both sides by a set of steel spheres backed by elastic springs, and the two

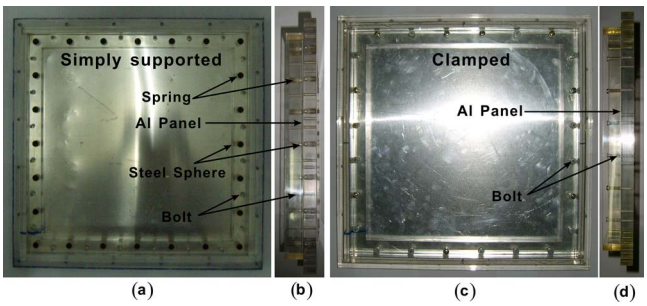


FIG. 4. (Color online) Two typical mounted cases: simply supported case and clamped case. (a) Front view of the simply supported fixture; (b) side view of (a); (c) front view of the clamped fixture; (d) side view of (c).

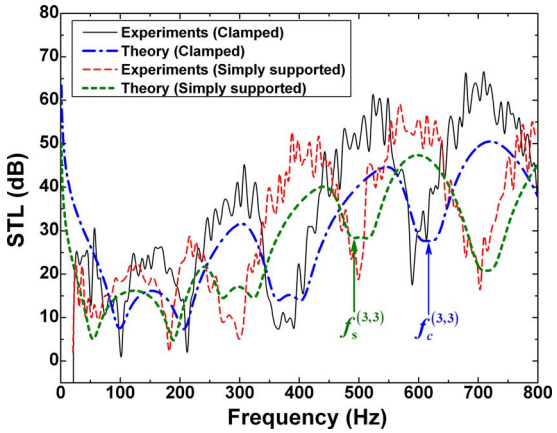


FIG. 5. (Color online) STL plotted as a function of frequency for fully clamped and fully simply supported Al double-panel partitions in the case of normal sound incidence. The (3,3) mode resonance dips for the simply supported double-panel and the clamped double-panel are particularly denoted by arrows at $f_s^{(3,3)}=489$ Hz and $f_c^{(3,3)}=619$ Hz, respectively.

Plexiglass frames holding the panel in between are directly bolted, as shown in Figs. 4(a) and 4(b). Since the steel spheres can freely rotate without any restraint, the edges of the panel can also freely rotate but are restrained in the radial direction of the spheres, which represents the simply supported boundary condition. For the clamped boundary condition, each Al panel is directly held in between the two Plexiglass frames, with each frame firmly fastened by bolts. All the cracks (if any) around the perimeter are sealed using adhesive glass cement (which for narrow cracks provides a seal equivalent to caulking), as shown in Fig. 3.

STL of the double-panel partition system is measured for two different cases, i.e., fully clamped and simply supported. The two identical Al panels with length $a=0.3$ m, width $b=0.3$ m, and thickness $h=1$ mm are separated by an air cavity of depth $H=8$ cm, as shown in Fig. 2. With a loudspeaker (diameter of 20 cm) generating a white noise normally incident on the incident panel, the sound pressures on both sides of the partition system are measured by condenser microphones and analyzed by a dual-channel dynamical signal analyzer (HP: 35670A).

B. Experimental results and model validation

Figure 5 presents the measured STL value as a function of incident frequency (0–800 Hz) for both fully clamped and fully simply supported boundary conditions; for comparison, theoretical predictions obtained with the present model for each type of boundary condition are also included.

It is seen in Fig. 5 that as far as the STL tendency with varying frequency is concerned, the agreement between the theoretical predictions and experimental measurements is good. The discernible discrepancies between the theory and experiment can be attributed to a number of factors, such as the imperfect normal plane sound wave, the uneven panel thickness, and the inevitable structural flanking transmission paths.[1] Note also that the experimental results at frequencies below 50 Hz are not reliable because the flanking transmission paths of the test facility play a prominent role in this frequency range.[1,3,4]

The results of Fig. 5 clearly demonstrate the significant influence of boundary conditions on the transmission loss of a double-panel partition. The intense peaks and dips in the STL versus frequency curve reflect the inherent modal behaviors of the double-panel system. The STL dips in the simply supported case are shifted to lower frequencies in comparison with those of the clamped case, implying the fact that the natural frequencies of the simply supported system are lower than their counterparts of the clamped system. It should be pointed out that the STL dips (apart from the second dips in the two theoretical curves) are dominated by the modal behavior of the radiating panel. It has been established that the second dips are associated with the "plate-cavity-plate" resonance,[6,33] which is insensitive to the imposed boundary conditions.

The research by Carneal and Fuller[1] on transmission loss across double-panel partitions should be mentioned here. The proposed theoretical model by Carneal and Fuller[1] was established on simply supported boundary condition, whereas their experimental measurements were performed on two clamped plates. In order to experimentally verify the model, the stiffness of the simply supported plate was increased artificially by a factor of $\sqrt{2}$ for each boundary to approximate the clamped boundary condition. This assumption was also used to predict the natural frequencies of the clamped double-panel system.

Although the simply supported mode shapes may be a reasonable approximation of the clamped mode shapes,[37] the natural frequencies associated with the simply supported boundary condition are lower, as the clamped condition provides a more rigorous constraint on panel vibration. To account for this increased constraint, simply increasing the panel stiffness by a factor of $\sqrt{2}$, as suggested by Carneal and Fuller,[1] may not be widely feasible because the increased panel stiffness is determined by many parameters, such as panel dimensions, material properties, and incident sound frequency. In order to give a straightforward understanding of the distinctions between the different boundary conditions (i.e., clamped, simply supported, and modified simply supported by increasing panel stiffness), a comparison of the STL versus frequency curves obtained theoretically using the three different boundary conditions is presented in Fig. 6. The STL values predicted by the modified simply supported model are closer to the clamped model predictions when $f>250$ Hz, while they agree better with the simply supported model predictions when $f<250$ Hz. This suggests that by increasing the panel stiffness based on simply supported boundary conditions to emulate the clamped boundary condition may be feasible only when the frequency is sufficiently high.

Therefore, in subsequent studies of sound transmission through double-panel partitions, the modal functions of Eq. (16) that directly satisfy the clamped boundary condition (i.e., $w=0$ and $\partial w/\partial \mathbf{n}=0$, which is different from the simply supported condition $w=0$ and $\partial^2 w/\partial \mathbf{n}^2=0$, where $\mathbf{n}$ denotes the outward vector on the edges) are applied. In a closely related study on sound transmission across double-panel partitions of finite extent,[6] we have demonstrated that the correct implementation of the clamped boundary condition has

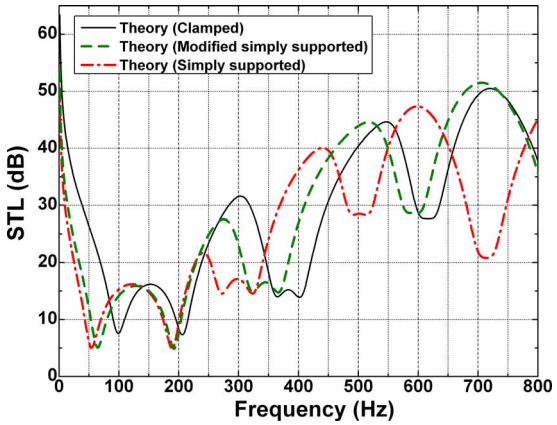


FIG. 6. (Color online) Comparison of STL values obtained theoretically for three theoretical boundary conditions (i.e., clamped, simply supported, and modified simply supported) in the case of normal sound incidence.

noticeable superiority than the approximate approach adopted by Carneal and Fuller[1] in view of the accurate prediction of the STL versus frequency curves and the natural frequencies.

To explore the boundary effects further, the typical (3,3) mode behavior of a fully clamped double-panel partition is compared in Fig. 7 with that of a simply supported one. The (3,3) mode natural frequency of the fully simply supported double-panel system occurs at $f_s^{(3,3)}=489$ Hz, while it shifts to $f_c^{(3,3)}=619$ Hz when fully clamped. The corresponding (3,3) mode shapes of the incident and radiating panels are presented in Figs. 7(a1) and (a2) for the fully simply supported case, while those for the fully clamped case are shown in Figs. 7(b1) and (b2). First, it is observed that the incident panel and the radiating panel vibrate in a symmetrical way (out-of-phase) for both cases; see Figs. 7(a1) and (b1) or (b1) and (b2). Symmetric motion (with respect to the symmetry plane running through the center of the partition) means that the panels move in breathing motion, both in or both out at a

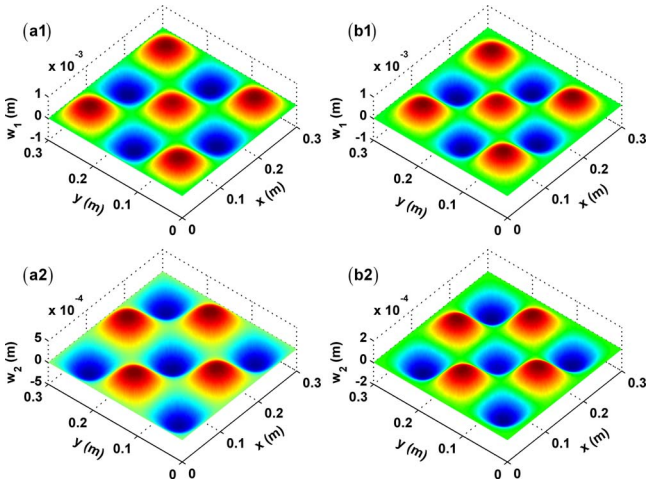


FIG. 7. (Color online) Panel deflection mode shapes under normal sound excitation at frequency $f_s^{(3,3)}=489$ Hz for fully simply supported case and $f_c^{(3,3)}=619$ Hz for fully clamped case, where the responses are controlled by the (3,3) natural mode: (a1) simply supported incident panel; (a2) simply supported radiating panel; (b1) clamped incident panel; (b2) clamped radiating panel.

TABLE I. Comparison of the fundamental frequency coefficient $\Omega_{00}=\sqrt{\rho h/D\omega_{00}a^2}$ for a fully clamped single-leaf panel with that of fully simply supported.

Boundary condition	$\lambda=a/b$	Leissa (Ref. 38)	Laura and Grossi (Ref. 39)	Present study
S-S-S-S[a]	1.0	19.73	19.74	19.74
C-C-C-C[a]	1.0	35.99	35.99	35.99

[a]"S-S-S-S" and "C-C-C-C" represent fully simply supported and fully clamped on four edges of the panel, respectively.

given position and a given time. Second, although the panel mode shapes at different boundary conditions exhibit similar forms [see Figs. 7(a1) and (b1) or (a2) and (b2)], discernible discrepancies can be observed at panel edges, especially for the counterparts of Figs. 7(a2) and (b2). These differences at the panel edges reflect the boundary effects, i.e., the requirement that $\partial w/\partial \mathbf{n}=0$ for the clamped condition and that $\partial^2 w/\partial \mathbf{n}^2=0$ for the simply supported condition.

To better differentiate the clamped case from the simply supported one, the predicted fundamental frequency coefficients $\Omega_{00}=\sqrt{\rho h/D\omega_{00}a^2}$ of a single-leaf panel either fully clamped or simply supported around its edges in vacuum are compared in Table I. The significant alteration of the fundamental frequency coefficient from one case to another confirms once more the remarkable differences of the two types of boundary condition.

V. RELATIONSHIPS BETWEEN CLAMPED AND SIMPLY SUPPORTED BOUNDARY CONDITIONS

Numerical studies are performed in this section to explore further the relationships between the two different boundary conditions and the significant influence of the boundary condition on the sound insulation properties of double-panel structures in terms of frequency characteristic curves (i.e., STL versus frequency curves) and panel vibration behaviors. The panel material properties and structural dimensions are taken as follows: The two panels are both made of aluminum, with Young's modulus $E=70$ GPa, density $\rho=2700$ kg/m^3, Poisson ratio $\nu=0.33$, and loss factor (introduced with the complex Young's modulus) $\eta=0.01$. The two rectangular panels have identical geometrical dimensions: length $a=0.5$ m in the x-direction, width $b=0.5$ m in the y-direction, and thickness $h=2$ mm in the z-direction (Fig. 1). The depth of the air cavity is $H=0.08$ m. Furthermore, the density of air is $\rho_0=1.21$ kg/m^3, the speed of sound in air is $c_0=343$ m/s, and the amplitude of the acoustic velocity potential for the incident sound field is fixed at $I_0=1$ m^2/s.

A. Consistency of clamped model and simply supported model

To check the consistency of the clamped model and the simply supported model, the panel dimensions are extended to infinite so that the constraint effect of the boundary conditions vanishes. The two models based on the clamped modal function and the simply supported modal function, respectively, are also applicable for infinitely large structures

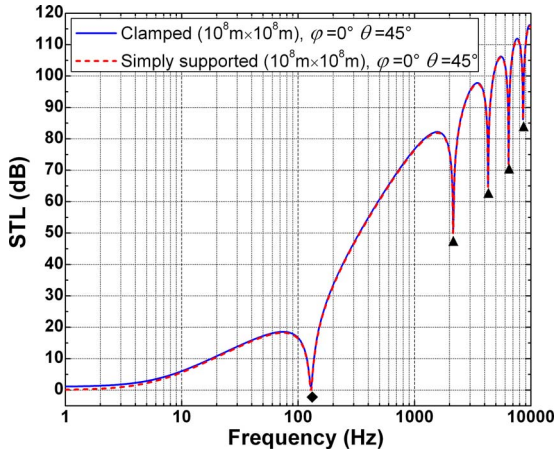


FIG. 8. (Color online) STL of infinitely large double-panel partition theoretically obtained separately with clamped boundary and simply supported boundary. ◆: mass-air-mass resonance; ▲: standing-wave resonance.

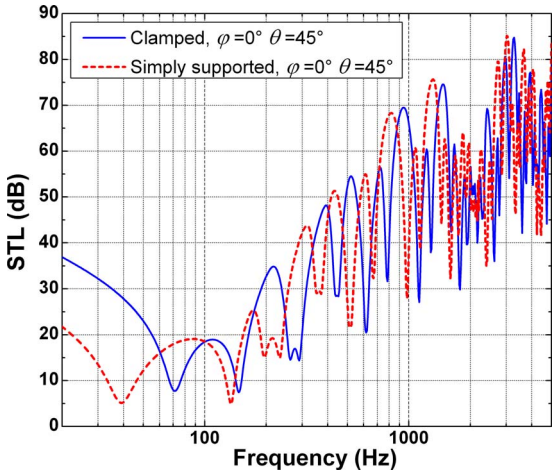


FIG. 9. (Color online) Predicted STL of finite (0.5×0.5 m^2) double-panel partition plotted as a function of frequency for incident sound with elevation angle φ=0° and azimuth angle θ=45°.

if sufficiently large values of panel length a and panel width b are taken in numerical calculations. The corresponding STL results for the two models with $a=10^8$ m and $b=10^8$ m assumed are shown in Fig. 8. As anticipated, an excellent agreement is achieved between the two models, as the effect of boundary condition becomes negligible (in other words, no boundaries exist) when the structure is extended to infinite.

The mass-air-mass resonance dip in Fig. 8 is marked by the symbol ◆, which is a unique phenomenon owned by the double-panel system and can be approximately predicted by the formula[6]

$$f_\alpha = \frac{1}{2\pi\cos\varphi}\sqrt{\frac{\rho_0 c_0^2}{H}\frac{(m_1+m_2)}{m_1 m_2}}. \tag{37}$$

The standing-wave resonance dips labeled by the symbol ▲ occur when the depths of the air gap in between the two panels are integer numbers of half wavelength of the incident sound. The corresponding resonance frequencies can be obtained by[19]

$$f_{s,n} = \frac{nc_0}{2H} \quad (n=1,2,3,\ldots). \tag{38}$$

The theoretical predictions given in Fig. 8 agree excellently well with Eqs. (37) and (38).

B. Effects of different boundary conditions on STL

In this section, the STL behaviors of the rectangular double-panel partition predicted by the two models based separately on the clamped boundary condition and the simply supported boundary condition are compared for different cases, e.g., sound incident with different elevation angles (φ=0°,30°,60°) and fixed azimuth angle of θ=45°. The results are presented in Figs. 9, 11, and 13. Here, the incident azimuth angle is fixed at 45° for the purpose of fully exciting the relevant panel vibration modes.

The typical (3,3) modal shapes of panel deflection associated with the cases of Figs. 9, 11, and 13 are presented in Figs. 10, 12, and 14, respectively. Note that although the mode shapes in Fig. 10 look similar to those shown in Fig. 7, different natural frequencies ($f_s^{(3,3)}$=352 Hz for the simply supported case and $f_c^{(3,3)}$=446 Hz for the clamped case) are obtained due to the different panel dimensions considered here (i.e., 0.5 m×0.5 m×2 mm).

As mentioned above, except for the second dip (i.e., the plate-cavity-plate resonance), which is insensitive to boundary conditions as seen in Figs. 9, 11, and 13, other dips determined mainly by the natural vibration of the radiating panel are significantly shifted when the boundary conditions are changed. Additionally, for the three considered cases, it can be seen that the STL values of the clamped system are distinctly higher than those of the simply supported system in the lower frequency range. For the higher frequency range, however, all the three cases show different trends. For example, for the case of elevation angle φ=0°, the STL values obtained with the two different boundary conditions have overall the same order of magnitude, although the resonance

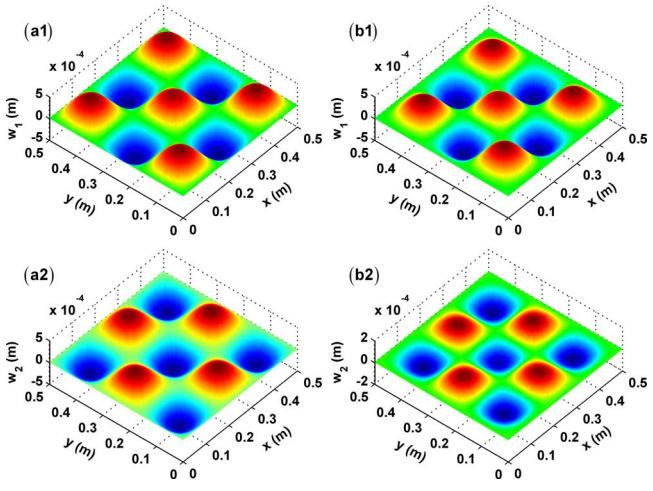


FIG. 10. (Color online) Panel deflection mode shapes under normal sound excitation (φ=0° and θ=45°) at frequency $f_s^{(3,3)}$=352 Hz for fully simply supported case and $f_c^{(3,3)}$=446 Hz for fully clamped case, where the responses are controlled by the (3,3) natural mode: (a1) simply supported incident panel; (a2) simply supported radiating panel; (b1) clamped incident panel; (b2) clamped radiating panel.

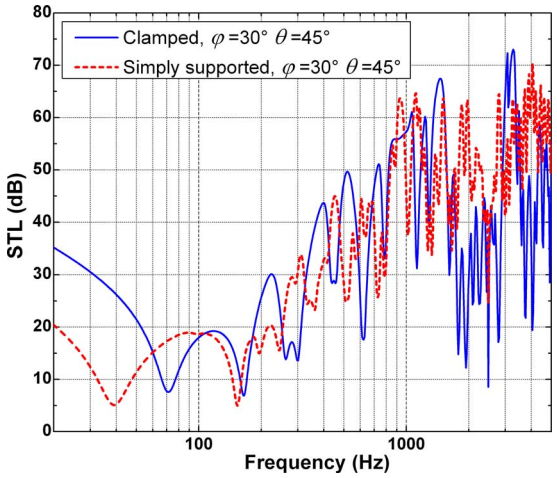


FIG. 11. (Color online) Predicted STL of finite (0.5×0.5 m^2) double-panel partition plotted as a function of frequency for incident sound with elevation angle $\varphi=30°$ and azimuth angle $\theta=45°$.

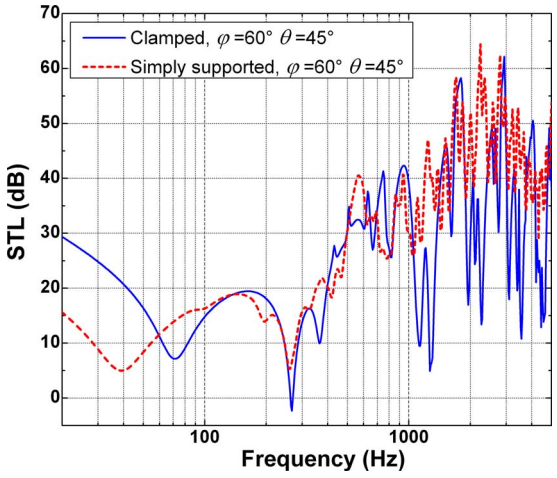


FIG. 13. (Color online) Predicted STL of finite (0.5×0.5 m^2) double-panel partition plotted as a function of frequency for incident sound with elevation angle $\varphi=60°$ and azimuth angle $\theta=45°$.

dips are not in accord with each other. By comparing the three plots in succession, it can be seen that as the elevation angle is increased, the discrepancies between the STL values obtained with different boundary conditions increase.

Taking a whole view of the results shown in Figs. 10, 12, and 14, one can see that the mode shapes of the simply supported panels can be approximated as the corresponding clamped panel mode shapes only for the case of normal sound incidence (see Fig. 10). For oblique sound incidence, the mode shapes of the simply supported panels are dramatically different from those of the clamped panels (see Figs. 12 and 14). The asymmetric sound incidence (i.e., oblique sound incidence, $0°<\varphi<90°$) induces asymmetric mode shapes of the simply supported panels (symmetric only about the incident plane), while the mode shapes of the clamped panels remain highly symmetric, which confirms the more rigorous restraint of the fully clamped condition on the movement of panel edges than that of the simply supported condition.

VI. CONCLUSIONS

An analytical approach has been developed to investigate the influence of boundary constraints (fully clamped versus fully simply supported) on the sound insulation performance of a finite double-panel structure containing an air cavity. The theory is built upon the vibration responses of the two panels coupled by the air cavity. Experimental measurements are subsequently performed to validate the theoretical predictions, with good overall agreement achieved for both types of the boundary condition. The inherent consistency of the two models based separately on the clamped boundary condition and the simply supported boundary condition is confirmed when the panel dimensions become infinitely large. The model is then used to systematically explore the

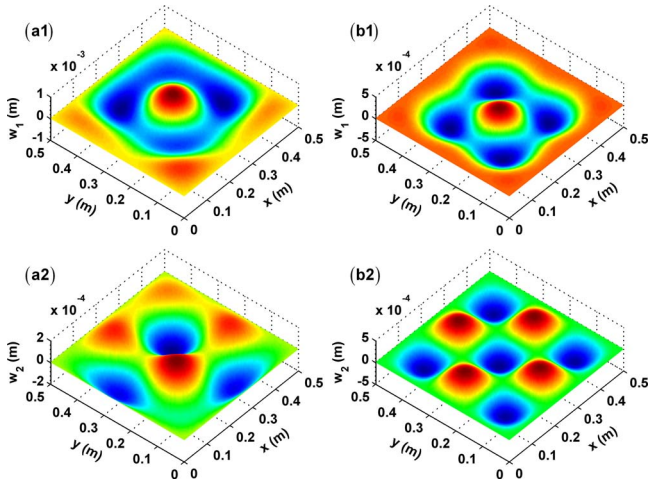


FIG. 12. (Color online) Panel deflection mode shapes under oblique sound excitation ($\varphi=30°$ and $\theta=45°$) at frequency $f_s^{(3,3)}=352$ Hz for fully simply supported case and $f_c^{(3,3)}=446$ Hz for fully clamped case, where the responses are controlled by the (3,3) natural mode: (a1) simply supported incident panel; (a2) simply supported radiating panel; (b1) clamped incident panel; (b2) clamped radiating panel.

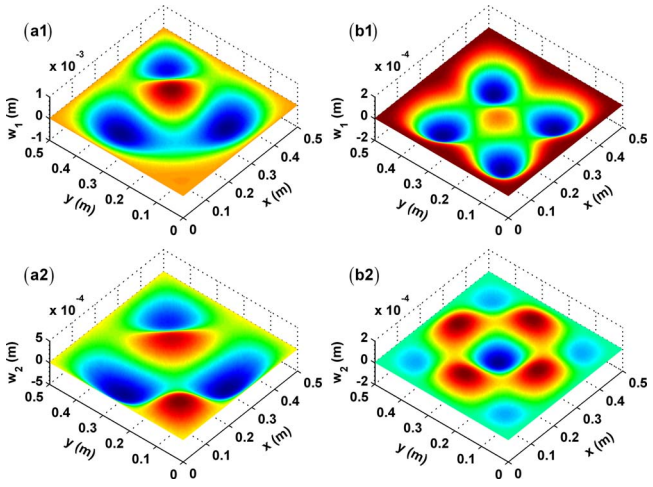


FIG. 14. (Color online) Panel deflection mode shapes under oblique sound excitation ($\varphi=60°$ and $\theta=45°$) at frequency $f_s^{(3,3)}=352$ Hz for fully simply supported case and $f_c^{(3,3)}=446$ Hz for fully clamped case, where the responses are controlled by the (3,3) natural mode: (a1) simply supported incident panel; (a2) simply supported radiating panel; (b1) clamped incident panel; (b2) clamped radiating panel.

effects of different boundary conditions on the sound isolation capability of double-panel partitions in terms of both the STL versus frequency plots and the (3,3) mode shapes of panel vibration.

The comparison of the STL versus frequency plots obtained with the two different boundary conditions suggests that the natural frequencies of a fully clamped double-panel partition are higher than those of a fully simply supported one (except for the plate-cavity-plate resonance frequency). This is attributed to the more rigorous constraint provided by the clamped condition than that by the simply supported condition, which is equivalent to increasing the panel stiffness. However, to account for this increased panel stiffness by an artificial factor (e.g., $\sqrt{2}$) is not widely feasible, as the increased panel stiffness is determined by many parameters such as panel dimensions, material properties, and incident sound frequency.

Obtained results at oblique sound incidence for the two different boundary condition cases suggest that at the lower frequency range, the STL values show noticeable discrepancies for the two cases, while at the higher frequency range the discrepancies depend on the incident elevation angle. While the vibration mode shapes of a simply supported panel can be approximated as those of its clamped counterpart only in the case of normal sound incidence, dramatic distinctions exist between them in the case of oblique sound incidence.

ACKNOWLEDGMENTS

The authors would like to appreciate the valuable suggestions from Professor Ning Xiang and the equally valuable critical comments from the reviewers. This work is supported by the National Basic Research Program of China (Grant No. 2006CB601202), the National 111 Project of China (Grant No. B06024), the National Natural Science Foundation of China (Grant Nos. 10632060 and 10825210), and the National High Technology Research and Development Program of China (Grant No. 2006AA03Z519).

APPENDIX: DERIVATION OF EQUATION (33)

The deflection coefficients of the two panels are

$$\{\alpha_{1,kl}\} = [\alpha_{1,11} \quad \alpha_{1,21} \quad \cdots \quad \alpha_{1,M1} \quad \alpha_{1,12} \quad \alpha_{1,22} \quad \cdots \quad \alpha_{1,M2} \quad \cdots \quad \alpha_{1,MN}]^T_{MN\times 1}, \tag{A1}$$

$$\{\alpha_{2,kl}\} = [\alpha_{2,11} \quad \alpha_{2,21} \quad \cdots \quad \alpha_{2,M1} \quad \alpha_{2,12} \quad \alpha_{2,22} \quad \cdots \quad \alpha_{2,M2} \quad \cdots \quad \alpha_{2,MN}]^T_{MN\times 1}. \tag{A2}$$

The left-hand side of Eq. (33) represents the generalized force, where

$$\{F_{kl}\} = 2j\omega\rho_0 I[f_{11} \quad f_{21} \quad \cdots \quad f_{M1} \quad f_{12} \quad f_{22} \quad \cdots \quad f_{M2} \quad \cdots \quad f_{MN}]^T_{MN\times 1}, \tag{A3}$$

$$\lambda^{*1}_{1,mn} = 3\left(\frac{m}{a}\right)^4 + 3\left(\frac{n}{b}\right)^4 + 2\left(\frac{m}{a}\right)^2\left(\frac{n}{b}\right)^2, \tag{A4}$$

$$\Delta^{*1}_1 = \begin{bmatrix} \lambda^{*1}_{1,11} & & & & & & & & & \\ & \lambda^{*1}_{1,21} & & & & & & & & \\ & & \ddots & & & & & & & \\ & & & \lambda^{*1}_{1,M1} & & & & & & \\ & & & & \lambda^{*1}_{1,12} & & & & & \\ & & & & & \lambda^{*1}_{1,22} & & & & \\ & & & & & & \ddots & & & \\ & & & & & & & \lambda^{*1}_{1,M2} & & \\ & & & & & & & & \ddots & \\ & & & & & & & & & \lambda^{*1}_{1,MN} \end{bmatrix}_{MN\times MN}, \tag{A5}$$

$$\lambda^{*2}_{1,n} = \frac{2n^4}{b^4}\begin{bmatrix} 0 & 1 & 1 & \cdots & 1 \\ 1 & 0 & 1 & \cdots & 1 \\ 1 & 1 & 0 & \cdots & \cdots \\ \cdots & \cdots & \cdots & \ddots & 1 \\ 1 & 1 & \cdots & 1 & 0 \end{bmatrix}_{M\times N}, \tag{A6}$$

$$\Delta^{*2}_1 = \begin{bmatrix} \lambda^{*2}_{1,1} & & & \\ & \lambda^{*2}_{1,2} & & \\ & & \ddots & \\ & & & \lambda^{*2}_{1,N} \end{bmatrix}_{MN\times MN}, \tag{A7}$$

$$\lambda_1^{*3} = \frac{2}{a^4}\begin{bmatrix} 1^4 & & & \\ & 2^4 & & \\ & & \ddots & \\ & & & M^4 \end{bmatrix}_{M\times N}, \tag{A8}$$

$$\Delta_1^{*3} = \begin{bmatrix} 0 & \lambda_1^{*3} & \lambda_1^{*3} & \cdots & \lambda_1^{*3} \\ \lambda_1^{*3} & 0 & \lambda_1^{*3} & \cdots & \lambda_1^{*3} \\ \lambda_1^{*3} & \lambda_1^{*3} & 0 & \cdots & \cdots \\ \cdots & \cdots & \cdots & \ddots & \lambda_1^{*3} \\ \lambda_1^{*3} & \lambda_1^{*3} & \cdots & \lambda_1^{*3} & 0 \end{bmatrix}_{MN\times MN}, \tag{A9}$$

$$\Delta_2^{*1} = \frac{9ab}{4}\begin{bmatrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{bmatrix}_{MN\times MN}, \tag{A10}$$

$$\lambda_2^{*2} = \frac{3ab}{2}\begin{bmatrix} 0 & 1 & 1 & \cdots & 1 \\ 1 & 0 & 1 & \cdots & 1 \\ 1 & 1 & 0 & \cdots & \cdots \\ \cdots & \cdots & \cdots & \ddots & 1 \\ 1 & 1 & \cdots & 1 & 0 \end{bmatrix}_{M\times N}, \tag{A11}$$

$$\Delta_2^{*2} = \begin{bmatrix} \lambda_2^{*2} & & & \\ & \lambda_2^{*2} & & \\ & & \ddots & \\ & & & \lambda_2^{*2} \end{bmatrix}_{MN\times MN}, \tag{A12}$$

$$\lambda_2^{*3} = \frac{3ab}{2}\begin{bmatrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{bmatrix}_{M\times N}, \tag{A13}$$

$$\Delta_2^{*3} = \begin{bmatrix} 0 & \lambda_2^{*3} & \lambda_2^{*3} & \cdots & \lambda_2^{*3} \\ \lambda_2^{*3} & 0 & \lambda_2^{*3} & \cdots & \lambda_2^{*3} \\ \lambda_2^{*3} & \lambda_2^{*3} & 0 & \cdots & \cdots \\ \cdots & \cdots & \cdots & \ddots & \lambda_2^{*3} \\ \lambda_2^{*3} & \lambda_2^{*3} & \cdots & \lambda_2^{*3} & 0 \end{bmatrix}_{MN\times MN}, \tag{A14}$$

$$\lambda_2^{*4} = ab\begin{bmatrix} 0 & 1 & 1 & \cdots & 1 \\ 1 & 0 & 1 & \cdots & 1 \\ 1 & 1 & 0 & \cdots & \cdots \\ \cdots & \cdots & \cdots & \ddots & 1 \\ 1 & 1 & \cdots & 1 & 0 \end{bmatrix}_{M\times N}, \tag{A15}$$

$$\Delta_2^{*4} = \begin{bmatrix} 0 & \lambda_2^{*4} & \lambda_2^{*4} & \cdots & \lambda_2^{*4} \\ \lambda_2^{*4} & 0 & \lambda_2^{*4} & \cdots & \lambda_2^{*4} \\ \lambda_2^{*4} & \lambda_2^{*4} & 0 & \cdots & \cdots \\ \cdots & \cdots & \cdots & \ddots & \lambda_2^{*4} \\ \lambda_2^{*4} & \lambda_2^{*4} & \cdots & \lambda_2^{*4} & 0 \end{bmatrix}_{MN\times MN}. \tag{A16}$$

Using the definition of the sub-matrices presented above, one obtains

$$[T_{11,kl}]_{MN\times MN} = 4D_1\pi^4 ab(\Delta_1^{*1} + \Delta_1^{*2} + \Delta_1^{*3}) - \left(m_1\omega^2 + j\omega\rho_0\frac{2\omega e^{2jk_zH}}{k_z(1-e^{2jk_zH})}\right)\cdot(\Delta_2^{*1} + \Delta_2^{*2} + \Delta_2^{*3} + \Delta_2^{*4}), \tag{A17}$$

$$[T_{12,kl}]_{MN\times MN} = j\omega\rho_0\frac{2\omega e^{jk_zH}}{k_z(1-e^{2jk_zH})}\times(\Delta_2^{*1} + \Delta_2^{*2} + \Delta_2^{*3} + \Delta_2^{*4}), \tag{A18}$$

$$[T_{21,kl}]_{MN\times MN} = j\omega\rho_0\frac{2\omega e^{jk_zH}}{k_z(1-e^{2jk_zH})}\times(\Delta_2^{*1} + \Delta_2^{*2} + \Delta_2^{*3} + \Delta_2^{*4}), \tag{A19}$$

$$[T_{22,kl}]_{MN\times MN} = 4D_2\pi^4 ab(\Delta_1^{*1} + \Delta_1^{*2} + \Delta_1^{*3}) - \left(m_2\omega^2 + j\omega\rho_0\frac{2\omega e^{2jk_zH}}{k_z(1-e^{2jk_zH})}\right)\cdot(\Delta_2^{*1} + \Delta_2^{*2} + \Delta_2^{*3} + \Delta_2^{*4}). \tag{A20}$$

[1] J. P. Carneal and C. R. Fuller, "An analytical and experimental investigation of active structural acoustic control of noise transmission through double panel systems," J. Sound Vib. **272**, 749–771 (2004).
[2] S. J. Pietrzko and Q. Mao, "New results in active and passive control of sound transmission through double wall structures," Aerosp. Sci. Technol. **12**, 42–53 (2008).
[3] J. D. Quirt, "Sound transmission through windows I. Single and double glazing," J. Acoust. Soc. Am. **72**, 834–844 (1982).
[4] J. D. Quirt, "Sound transmission through windows II. Double and triple glazing," J. Acoust. Soc. Am. **74**, 534–542 (1983).
[5] K. Lyle and J. Mixson, "Laboratory study of sidewall noise transmission and treatment for a light aircraft fuselage," J. Aircr. **24**, 660–665 (1987).
[6] F. X. Xin, T. J. Lu, and C. Q. Chen, "Vibroacoustic behavior of clamp mounted double-panel partition with enclosure air cavity," J. Acoust. Soc. Am. **124**, 3604–3612 (2008).
[7] G. Maidanik, "Response of ribbed panels to reverberant acoustic fields," J. Acoust. Soc. Am. **34**, 809–826 (1962).
[8] F. G. Leppington, E. G. Broadbent, and G. F. Butler, "Transmission of sound through a pair of rectangular elastic plates," IMA J. Appl. Math. **71**, 940–955 (2006).
[9] M. Ruzzene, "Vibration and sound radiation of sandwich beams with honeycomb truss core," J. Sound Vib. **277**, 741–763 (2004).
[10] A. Pellicier and N. Trompette, "A review of analytical methods, based on the wave approach, to compute partitions transmission loss," Appl. Acoust. **68**, 1192–1212 (2007).
[11] J. H. Lee and J. Kim, "Analysis of sound transmission through periodically stiffened panels by space-harmonic expansion method," J. Sound Vib. **251**, 349–366 (2002).
[12] B. Liu, L. Feng, and A. Nilsson, "Sound transmission through curved aircraft panels with stringer and ring frame attachments," J. Sound Vib. **300**, 949–973 (2007).
[13] D. Takahashi, "Effects of panel boundedness on sound transmission problems," J. Acoust. Soc. Am. **98**, 2598–2606 (1995).
[14] A. London, "Transmission of reverberant sound through double walls," J. Acoust. Soc. Am. **22**, 270–279 (1950).
[15] W. Kropp and E. Rebillard, "On the air-borne sound insulation of double wall constructions," Acta. Acust. Acust. **85**, 707–720 (1999).
[16] J. M. P. Antonio, A. Tadeu, and L. Godinho, "Analytical evaluation of the acoustic insulation provided by double infinite walls," J. Sound Vib. **263**, 113–129, (2003).
[17] G.-F. Lin and J. M. Garrelick, "Sound transmission through periodically framed parallel plates," J. Acoust. Soc. Am. **61**, 1014–1018 (1977).
[18] D. Takahashi, "Sound radiation from periodically connected double-plate

structures," J. Sound Vib. **90**, 541–557 (1983).
[19]J. Wang, T. J. Lu, J. Woodhouse, R. S. Langley, and J. Evans, "Sound transmission through lightweight double-leaf partitions: Theoretical modelling," J. Sound Vib. **286**, 817–847 (2005).
[20]F. X. Xin and T. J. Lu, "Effects of core topology on sound insulation performance of lightweight all-metallic sandwich panels," Mater. Manuf. Processes (2009), In press.
[21]J. Brunskog, "The influence of finite cavities on the sound insulation of double-plate structures," J. Acoust. Soc. Am. **117**, 3727–3739 (2005).
[22]M. Villot, C. Guigou, and L. Gagliardini, "Predicting the acoustical radiation of finite size multi-layered structures by applying spatial windowing on infinite structures," J. Sound Vib. **245**, 433–455 (2001).
[23]J. D. Chazot and J. L. Guyader, "Prediction of transmission loss of double panels with a patch-mobility method," J. Acoust. Soc. Am. **121**, 267–278 (2007).
[24]C. Bao and J. Pan, "Experimental study of different approaches for active control of sound transmission through double walls," J. Acoust. Soc. Am. **102**, 1664–1670 (1997).
[25]J. Pan and C. Bao, "Analytical study of different approaches for active control of sound transmission through double walls," J. Acoust. Soc. Am. **103**, 1916–1922 (1998).
[26]R. E. Jones, "Intercomparisons of laboratory determinations of airborne sound transmission loss," J. Acoust. Soc. Am. **66**, 148–164 (1979).
[27]B. K. Kim, H. J. Kang, J. S. Kim, H. S. Kim, and S. R. Kim, "Tunneling effect in sound transmission loss determination: Theoretical approach," J. Acoust. Soc. Am. **115**, 2100–2109 (2004).
[28]N. H. Farag and J. Pan, "Free and forced in-plane vibration of rectangular plates," J. Acoust. Soc. Am. **103**, 408–413 (1998).
[29]I. Bosmans, P. Mees, and G. Vermeir, "Structure-borne sound transmission between thin orthotropic plates: Analytical solutions," J. Sound Vib. **191**, 75–90 (1996).
[30]W. R. Graham, "Analytical approximations for the modal acoustic impedances of simply supported, rectangular plates," J. Acoust. Soc. Am. **122**, 719–730 (2007).
[31]C. Maury, P. Gardonio, and S. J. Elliott, "Model for active control of flow-induced noise transmitted through double partitions," AIAA J. **40**, 1113–1121 (2002).
[32]W. A. Utley and B. L. Fletcher, "The effect of edge conditions on the sound insulation of double windows," J. Sound Vib. **26**, 63–72 (1973).
[33]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).
[34]V. Hongisto, M. Lindgren, and J. Keranen, "Enhancing maximum measurable sound reduction index using sound intensity method and strong receiving room absorption," J. Acoust. Soc. Am. **109**, 254–265 (2001).
[35]ASTM E 90-04, "Standard test method for laboratory measurement of airborne sound transmission loss of building partitions and elements," 2004.
[36]C. Guigou, Z. Li, and C. R. Fuller, "The relationship between volume velocity and far-field radiated pressure of a planar structure," J. Sound Vib. **197**, 252–254 (1996).
[37]A. W. Leissa, *Vibrations of Plates* (Acoustical Society of America, New York, 1993).
[38]A. W. Leissa, "The free vibration of rectangular plates," J. Sound Vib. **31**, 257–293 (1973).
[39]P. A. A. Laura and R. O. Grossi, "Transverse vibrations of rectangular plates with edges elastically restrained against translation and rotation," J. Sound Vib. **75**, 101–107 (1981).

Internal noise reduction from dependent signal mixtures using microphones and a piezoelectric device under blind condition

Mitsuharu Matsumoto[a)]
The Education and Research Center for Frontier Science, The University of Electro-Communications, 1-5-1 Chofugaoka, Chofu-shi, Tokyo 182-8585, Japan

Shuji Hashimoto[b)]
Department of Applied Physics, Waseda University, 55N-4F-10A, 3-4-1 Okubo, Shinjuku-ku, Tokyo 169-8555, Japan

(Received 16 May 2008; revised 7 January 2009; accepted 9 January 2009)

The acoustical array composed of microphones and piezoelectric devices can reduce internally generated noise from the system such as a machine, a vehicle, or a robot. However, when the acoustical array combining microphones and piezoelectric devices is employed, it is necessary to estimate the gains of microphones and piezoelectric devices concerning the system that generates the noise in advance. The aim of this paper is to simplify this process, that is, to reduce internally generated noise without preknowledge concerning the gains of microphones and piezoelectric devices. Although the proposed method handles an anechoic mixing, it has some merits that complement the other typical blind source separation algorithms. In this paper, the problem is first formulated. The assumptions and the methodology of the proposed method are then given with some experimental results. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3077218]

PACS number(s): 43.60.Fg, 43.60.Ac, 43.60.Pt [EJS] Pages: 1518–1528

I. INTRODUCTION

Microphone array is widely employed in various scenes.[1–3] The applications include hearing aids,[4] speech enhancement,[5] robot audition,[6,7] and so on. There are many methods utilizing the typical microphone array such as the delay-sum type microphone array,[8] the adaptive microphone array,[9–11] and direction of arrival (DOA) estimation.[12–14] Although it can effectively reduce the external noise coming from the environment, it is difficult to reduce internally generated noise from the system such as machines, vehicles, and robots, because, in those cases, the noise source is very close to the microphones. The number of noise cannot always be known. Moreover, the noise intensity is also often larger than the signal, because the main sound propagation path is not the air but the solid frame of the machine.

On the other hand, the authors proposed another type of acoustical array composed not only of microphones but also of piezoelectric devices to reduce the internally generated noise.[15,16] In our method, we use piezoelectric devices as vibration sensors. In mechatronics, there has been a lot of research into controlling vibration with piezoelectric devices as actuators.[17–19] However, there is little research into applying piezoelectric devices to a microphone array to reduce the internally generated noise. Our aim is to reduce the internally generated noise by combining microphones and piezoelectric devices that have been attached to a noise source. The typical microphone array requires a precise estimation of the time difference or gains from the sounds to each microphone. In a similar fashion, when the acoustical array combining microphones and piezoelectric devices is employed, the gains of the microphones and the piezoelectric devices to the noise source should be estimated in advance. The objective of our research is to simplify this process, that is, to reduce the internally generated noise without the preknowledge concerning the gains of the microphones and the piezoelectric devices.

There are some useful blind source separation (BSS) algorithms such as independent component analysis[20–23] (ICA) and sparseness approaches.[24–26] ICA is based on the unitary diagonalization of the whitened data covariance matrix. There also exist some BSS algorithms utilizing higher order statistics. Sparseness means that most of the frequency components of a signal are zero, so that the sources rarely overlap in frequency domain. Under this assumption, it is possible to extract each signal using time-frequency binary masks. However, it is also difficult to employ these BSS algorithms to reduce the internally generated noise, because the internally generated noise is not always stationary, statistically independent, and sparse.

In this paper, we propose a method for reducing the internally generated noise by the acoustical array combining microphones and a piezoelectric device under the blind condition. The proposed method can reduce the internally generated noise from each microphone even when the number of the sounds and noise is larger than that of the acoustical sensors. The sounds may not only be statistically independent but may also be statistically dependent unlike ICA. The sounds and noise may overlap in frequency domain unlike the sparseness approach. They may also overlap in most of time domain. In Sec. II, we formulate the problem of the

[a)]Electronic mail: mitsuharu.matsumoto@ieee.org
[b)]Electronic mail: shuji@waseda.jp

 0001-4966/2009/125(3)/1518/11/$25.00

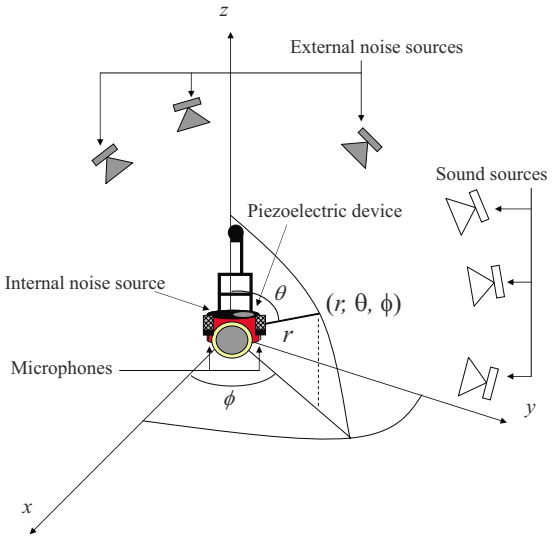


FIG. 1. (Color online) Basic concept.

acoustical array combining microphones and a piezoelectric device. In Sec. III, we state the assumptions and describe the algorithm of the proposed method. We show the experimental results to confirm the performance of the proposed method in Sec. IV. Conclusions are given in Sec. V.

II. PROBLEM FORMULATION

Let us consider a three-dimensional space where the coordinate is represented as (r, θ, ϕ) by a spherical polar coordinate system as illustrated in Fig. 1. O represents the origin. The relations among r, θ, ϕ, x, y, and z can be described as

$$x = r \sin\theta \cos\phi, \tag{1}$$

$$y = r \sin\theta \sin\phi, \tag{2}$$

and

$$z = r \cos\theta. \tag{3}$$

Let us consider J sounds, K external noise sources, and a microphone array with N microphones mounted on a system such as a machine, a vehicle, or a robot, which generates the internally generated noise, as illustrated in Fig. 1. $s_j(t)$ is the jth sound generated at the position $c_j=(r_j, \theta_j, \phi_j)$ at time t. $n^I(t)$ and $n_k^E(t)$ are the internally generated noise at the origin and the kth external noise at the position $\mathbf{c_k^E}=(r_k^E, \theta_k^E, \phi_k^E)$, respectively. To detect the object vibration which generates the noise, we also consider a piezoelectric device attached to the system. The system is located at the origin. $\mathbf{r_i^M}=(r_i^M, \theta_i^M, \phi_i^M)$ represents the coordinate of the ith microphone position. According to the survey by O'Grady *et al.*, signal mixtures of an acoustical array are categorized as the instantaneous, anechoic, and echoic mixings.[27] In this paper, we discuss an anechoic mixing. Under the above situation, it is considered that the signal $x_i^M(t)$ obtained by the ith microphone is expressed as

$$x_i^M(t) = \sum_{j=1}^{J} m_{ij} s_j(t - \tau_{ij}) + \sum_{k=1}^{K} m_{ik}^E n_k^E(t - \tau_{ik}^E) + m_i^I n^I(t), \tag{4}$$

where m_{ij}, m_{ik}^E, and m_i^I are the gains of the ith microphone to the jth signal source, the kth external noise, and the system which generates the internally generated noise, respectively. τ_{ij} and τ_{ik}^E represent the time delay from the jth sound and the kth external noise to the ith microphone, respectively. They can be described as

$$\tau_{ij} = \frac{|\mathbf{c_j}| - |\mathbf{c_j} - \mathbf{r}^{\mathbf{M}_i}|}{c}, \tag{5}$$

and

$$\tau_{ik}^E = \frac{|\mathbf{c}^{\mathbf{E}_k}| - |\mathbf{c}^{\mathbf{E}_k} - \mathbf{r}^{\mathbf{M}_i}|}{c}, \tag{6}$$

where c is the speed of sound, which is 343 m/s in air. As the piezoelectric device directly detects the object vibration, the signal $x^P(t)$ obtained by the piezoelectric device can be described as follows:

$$x^p(t) = p^I n^I(t), \tag{7}$$

where the p^I represents the gain of the piezoelectric device to the system which generates the internally generated noise. Note that Eq. (7) is valid as the piezoelectric device is collocated with the microphone. If there is no noise source, we may separate the sounds and the external noise by employing the typical microphone array and BSS algorithms such as ICA or sparseness approaches. However, the internally generated noise prevents us from employing these useful methods. Our aim is to reduce the internally generated noise without preknowledge of the amplitude gain of the microphones and the piezoelectric devices in spite of the existence of multiple sounds. To simplify the explanation, we rewrite $x_i^M(t)$ and $x^P(t)$ as follows:

$$x_i^M(t) = \sum_{j=1}^{J} d_{ij} s_j'(t - \tau_{ij}) + \sum_{k=1}^{K} d_{ik}^E n_k'^E(t - \tau_{ik}^E) + d_i^I n'^I(t), \tag{8}$$

and

$$x^P(t) = n'^I(t), \tag{9}$$

where d_{ij}, d_{ik}^E, and d_i^I represent the amplitude gain ratio of the ith microphone to the piezoelectric device concerning the jth sound, the kth external noise, and the internally generated noise, respectively. $s_j'(t)$, $n_k'^E(t)$, and $n'^I(t)$ represent the jth signal, the kth external noise, and the internally generated noise amplified by p^I, respectively, and can be expressed as

$$s_i'(t) = p^I s_i(t), \tag{10}$$

$$n_i'^E(t) = p^I n_i^E(t), \tag{11}$$

and

$$n'^I(t) = p^I n^I(t). \tag{12}$$

In our method, we estimate all d_i^I from the obtained $x_i^M(t)$ and $x^P(t)$, and reduce the internally generated noise from $x_i^M(t)$ by utilizing the estimated d_i^I.

III. PROPOSED METHOD

A. Assumptions regarding the sounds

Let us define $\Psi_{ii}^{M}(q,t)$ as the short-time auto-correlation function of the received signal $x_i^M(t)$ of the ith microphone at time t as follows:[28]

$$\Psi_{ii}^{M}(q,t) = \frac{1}{L}\sum_{l=0}^{L-1} x_i^M(l+t)x_i^M(q+l+t), \tag{13}$$

where q and L represent the time deviation and the number of the data used for calculating averages, respectively. Let us also define $\Psi^P(q,t)$ as the short-time auto-correlation function of the received signal $x^P(t)$ of the piezoelectric device at time t as follows:

$$\Psi^{P}(q,t) = \frac{1}{L}\sum_{l=0}^{L-1} x^P(l+t)x^P(q+l+t). \tag{14}$$

We define the ratio of the inter-channel power difference $\Delta A_i(t)$ as follows:

$$\Delta A_i(t) = \frac{\Psi_{ii}^{M}(0,t)}{\Psi^{P}(0,t)} \quad (i=1,2,\ldots,N). \tag{15}$$

We assume the following conditions regarding the sound sources.

Assumption 1. The signals are partially disjoint.

Let us define $\Phi_{jj}(q,t)$, $\Phi_{kk}^{E}(q,t)$, and $\Phi^{I}(q,t)$ as the short-time auto-correlation function of $s_j(t)$, $n_k^E(t)$, and $n^I(t)$ at time t as follows:

$$\Phi_{jj}(q,t) = \frac{1}{L}\sum_{l=0}^{L-1} s_j(l+t)s_j(q+l+t), \tag{16}$$

$$\Phi_{kk}^{E}(q,t) = \frac{1}{L}\sum_{l=0}^{L-1} n_k^E(l+t)n_k^E(q+l+t), \tag{17}$$

and

$$\Phi^{I}(q,T) = \frac{1}{L}\sum_{l=0}^{L-1} n^I(l+t)n^I(q+l+t). \tag{18}$$

This condition is described as follows.

There exists T_α for all $t \in [T_\alpha, T_\alpha + \max(\tau_{ij}, \tau_{ik}^{E})]$ such that

$$\Phi_{jj}(0,t) = 0, \tag{19}$$

$$\Phi_{kk}^{E}(0,t) = 0, \tag{20}$$

and

$$\Phi^{I}(0,t) \neq 0, \tag{21}$$

where $\Phi_{jj}(0,t)$, $\Phi_{kk}^{E}(0,t)$, and $\Phi^{I}(0,t)$ are the short-time mean power of $s_j(t)$, $n_k^E(t)$, and $n^I(t)$ at time t. Let us define the set Λ whose elements are t_l satisfying Eqs. (19)–(21). Let us also define r as the number of the elements of Λ. For instance, when we consider ten $t=1,2,3,\ldots$, and 10, and only $t=1$, 2, and 3 are satisfied with the Eqs. (19)–(21), the elements of Λ become 1, 2, and 3.

Assumption 2. Nonuniformity of $\Delta A_i(t)$.

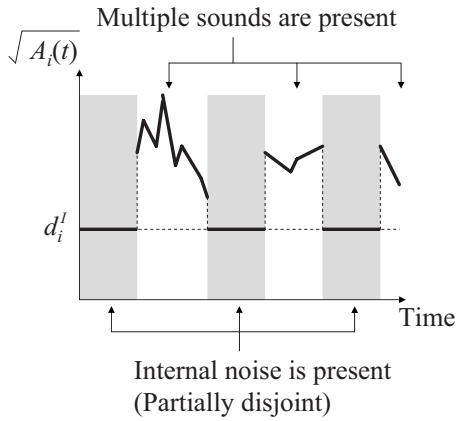


FIG. 2. Basic concept of $\Delta A_i(t)$.

We assume that the multiple signals should vary when they occur simultaneously, so that the number of $\Delta A_i(t)$ values appearing the most frequently in the ovarlapping intervals is less than the number of the disjoint points. Overlapping intervals are the intervals when some or all of the signals overlap in the time domain. In general, $\Psi_{ii}^{M}(q,t)$ can be expressed as follows:

$$\Psi_{ii}^{M}(q,t) = \sum_{j=1}^{J} m_{ij}^2 \Phi_{jj}(q,t) + \sum_{k=1}^{K} m_{ik}^{E2}\Phi_{kk}^{E}(q,t) + m_i^{I2}\Phi^{I}(q,t) + \Xi(q,t), \tag{22}$$

where $\Xi(q,t)$ represents the summation of the cross-correlations of the sounds, the external noise, and the internally generated noise. On the other hand, $\Psi^P(q,t)$ can be expressed as follows:

$$\Psi^{P}(q,t) = p^{I2}\Phi^{I}(q,t). \tag{23}$$

In Eqs. (22) and (23), the 2 in the superscript is not a label but a square operation. Let us define q_i as the maximal number such that $\Delta A_i(t)$ has the same value in spite of the mixtures of multiple sounds. Assumption 2 can be expressed as follows:

$$r > q_i. \tag{24}$$

We will verify that speech signals satisfy this assumption enough to allow for noise reduction concerning the internally generated noise from the microphones under the blind condition later. In the binary mask, it should be assumed that the frequency components of the mixed sounds hardly overlap in any time. On the other hand, in our method, the frequency components of mixed sounds may overlap in the frequency domain. The mixed sounds may also overlap in most of the time domain. Moreover, the sounds may not only be statistically independent but may also be statistically dependent. It should be noted that we do not need to know the intervals where the sounds are disjoint in the time domain through all the procedure.

B. Algorithm of the proposed method

Figure 2 shows the basic concept of the time fluctuation regarding $\Delta A_i(t)$. As shown in Fig. 2, $\Delta A_i(t)$ is constant when only one sound is present, while $\Delta A_i(t)$ varies when multiple

sounds are present under Assumptions 1 and 2. According to Assumption 1, if we consider the time t_l at which only one sound is present, $\Delta A_i(t_l)$ can be expressed as follows:

$$\Delta A_i(t_l) = \frac{m_i^{I2}\Phi^I(0,t_l)}{p^{I2}\Phi^I(0,t_l)} = \frac{d_i^{I2}\Phi'^I(0,t_l)}{\Phi'^I(0,t_l)} = d_i^{I2}. \tag{25}$$

As shown in Eq. (25), $\Delta A_i(t_l)$ becomes d_i^{I2} if only the internally generated noise is present. Equations (19)–(21) are satisfied by t_l. Hence, we can obtain the estimated d_i^I by utilizing the value of R_i that is value of the $\Delta A_i(t)$ ranked in the top as follows:

$$\overline{d_i^I} = \sqrt{R_i}, \tag{26}$$

where $\overline{d_i^I}$ represents the estimated value of d_i^I.

It should be noted that this procedure is executed independently for each i; that is, d_i^I can be identified even if the number of the acoustical sensors is less than the number of the sounds and noise. In this procedure, we do not need to estimate even the intervals when the sounds are disjoint in order to identify d_i^I. We only have to check the values of $\Delta A_i(t)$ for a certain time, which include the intervals where the signal is disjoint.

After the estimation of d_i^I, we can ideally reduce the internally generated noise in the time domain as follows:

$$\overline{x_i^M}(t) = x_i^M(t) - \overline{d_i^I}x^P(t), \tag{27}$$

where $\overline{x_i^M}(t)$ corresponds to the output $x_i^M(t)$ of the ith microphone without the internally generated noise, because $\overline{x_i^M}(t)$ can be rewritten as

$$\overline{x_i^M}(t) = x_i^M(t) - \overline{d_i^I}x^P(t) = \sum_{j=1}^{J} m_{ij}s_j(t-\tau_{ij}) + m_i^I n^I(t) - \frac{\overline{m_i^I}}{\overline{p^I}}p^I n^I(t) \approx \sum_{j=1}^{J} m_{ij}s_j(t-\tau_{ij}). \tag{28}$$

Due to a tiny phase difference among sounds recorded by the microphones and the piezoelectric device in the real world, the results of the noise reduction concerning the internally generated noise in the time-frequency domain like spectral subtraction may be better than those in the time domain. It should be noted that $\overline{x_i^M}(t)$ includes all the physical information between the microphones and the signals without the internally generated noise. Hence, we can employ various methods based on the microphone array such as beamforming or adaptive microphone array, ICA, or sparseness approaches. It should also be noted that the number of the sound sources and noise may be larger than the number of the acoustical sensors in our method although it is an ill-posed problem in the typical microphone array.

IV. EXPERIMENT

A. Examples of the waveform of the sound source and noise

To show that the piezoelectric device only detects the object vibration, we show the waveforms of the signal and noise. Figure 3 illustrates the experimental setup. The sound was generated from 0° and 50 cm apart from the mobile personal computer (PC) as illustrated in Fig. 3. We attached a piezoelectric device (SBT-C, Fishman) to the mobile PC. A microphone (ECM-TS125, Sony) was also set on the mobile PC. The recording room was a silent studio. The reverberation time was 72 ms. Figures 4 and 5 show the waveform of the output from the microphone and the piezoelectric device, respectively, when the speech signal was generated from the speaker. As shown in Figs. 4 and 5, although the microphone captures the sound from the speaker, the piezoelectric device does not capture the sound from the speaker because the sound was generated from the place apart from the mobile PC. To show that the piezoelectric device detects the object vibration, we also show the waveform of the output from the microphone and the piezoelectric device, respectively, when the mobile PC was powered. Figures 6 and 7 show the waveform of the output from the microphone and the piezoelectric

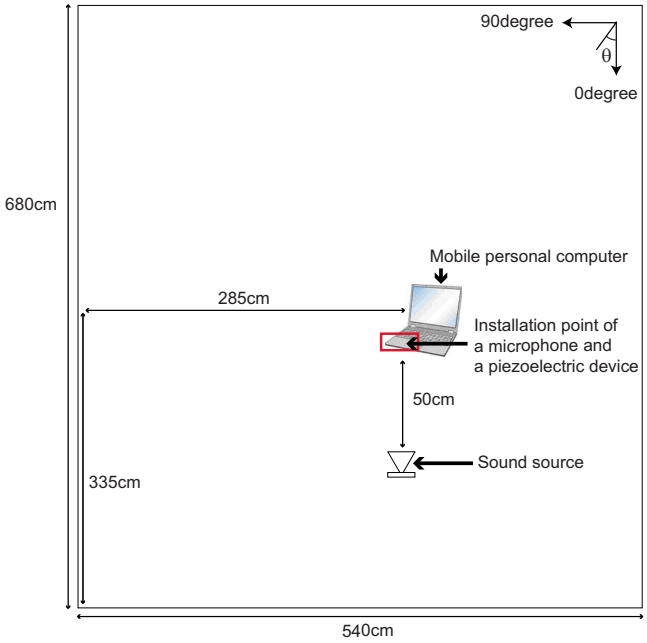


FIG. 3. (Color online) Experimental setup.

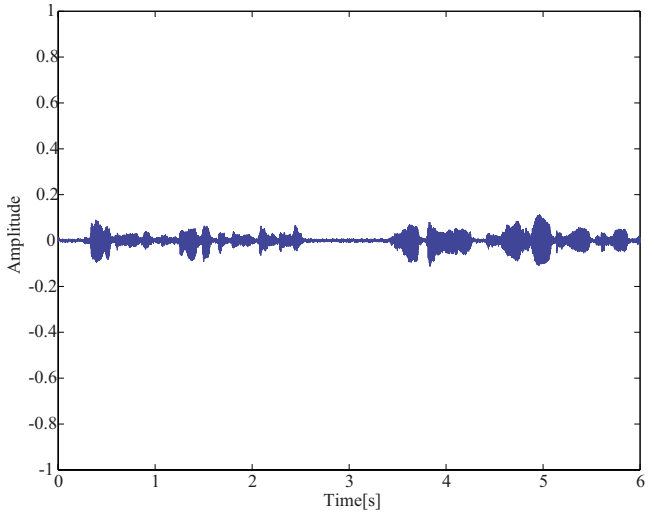


FIG. 4. (Color online) Output from the microphone when the sound was generated from the speaker.

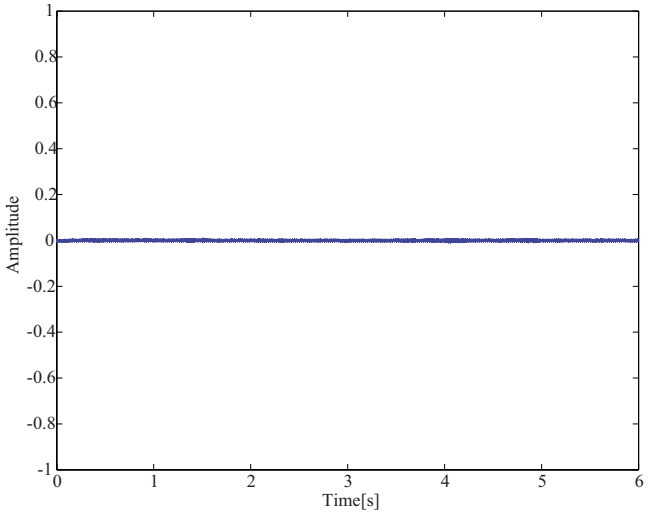


FIG. 5. (Color online) Output from the piezoelectric device when the sound was generated from the speaker.

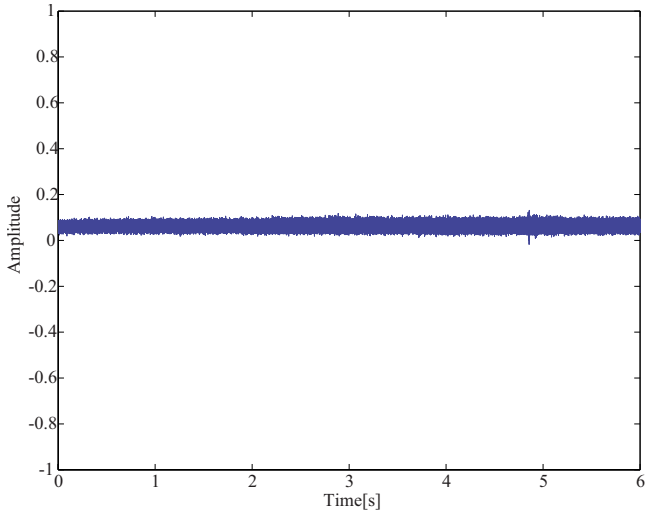


FIG. 7. (Color online) Output from the piezoelectric device when the noise was generated from the mobile PC.

device, respectively, when the noise was generated from the mobile PC. As shown in Figs. 6 and 7, not only the microphone but also the piezoelectric device capture the noise from the mobile PC. These examples show that the piezoelectric device only detect the object vibration regardless of the existence of sound sources apart from the object. In other words, we can assume that the piezoelectric device only detects the object vibration.

B. Experiment to confirm the robustness on the position of the sound sources

We then conducted the experiment to confirm whether we can reduce the noise of the microphone regardless of the sound position. We tried to reduce the noise from the mobile PC such as fan noise and hard disk drive (HDD) noise in the experiment. We used the same microphone, piezoelectric device, and mobile PC. The layout of the piezoelectric device, the microphone, and the mobile PC were also the same as that in Sec. IV A. L was set to 256. The directionalities of a microphone and a piezoelectric device are unknown in the experiments. A sound source was set 50 cm apart from the mobile PC. To confirm the robustness concerning the direction of the speaker, the sound was generated from 0° and 90°. As the sound sources, we selected five male voices and five female voices from "Japanese Newspaper Article Sentences" edited by the Acoustical Society of Japan. To confirm the performance quantitatively, we calculate signal-to-noise ratio (SNR) as follows:

$$\mathrm{SNR} = 10 \log_{10} \frac{\Sigma_{t=0}^{T} s_1^2(t)}{\Sigma_{t=0}^{T} n^{I2}(t)}, \tag{29}$$

where T represents the time length. To calculate SNR, we separately sampled $s(t)$ and $n^I(t)$ in the room. Noise reduction ratio (NRR) is defined as follows:

$$\mathrm{NRR} = \mathrm{SNR}_{\mathrm{after}} - \mathrm{SNR}_{\mathrm{before}}, \tag{30}$$

where $\mathrm{SNR}_{\mathrm{before}}$ and $\mathrm{SNR}_{\mathrm{after}}$ are the SNRs before and after noise reduction, respectively. Tables I and II show the results of $\mathrm{SNR}_{\mathrm{before}}$, $\mathrm{SNR}_{\mathrm{after}}$, and NRR regarding the microphone when the sound was generated from 0° and 90°, respectively. As shown in Tables I and II, the proposed method could reduce the noise over 7 dB throughout all the experiments

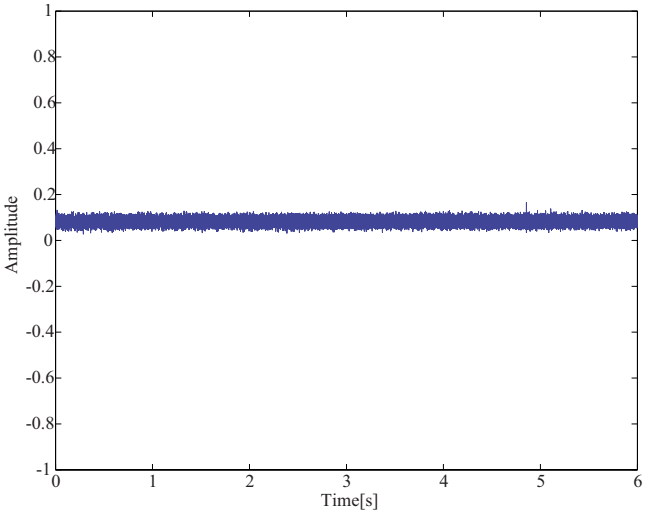


FIG. 6. (Color online) Output from the microphone when the noise was generated from the mobile PC.

TABLE I. NRR results when the sound was generated from 0°. The noise was from mobile PC, which include a fan noise and HDD noise.

	Gender	$\mathrm{SNR}_{\mathrm{before}}$ (dB)	$\mathrm{SNR}_{\mathrm{after}}$ (dB)	NRR (dB)
Sound 1	Male	−2.9227	5.2024	8.1252
Sound 2		1.0569	8.6452	7.5883
Sound 3		1.2398	8.7879	7.5482
Sound 4		5.7757	13.9037	8.128
Sound 5		4.5225	12.1693	7.6468
Sound 6	Female	4.3661	12.5574	8.1913
Sound 7		3.3001	10.8347	7.5347
Sound 8		8.012	15.4832	7.4712
Sound 9		5.8976	13.8739	7.9763
Sound 10		7.3457	15.2349	7.8892

TABLE II. NRR results when the sound was generated from 90°. The noise was from mobile PC, which include a fan noise and HDD noise.

	Gender	SNR_{before} (dB)	SNR_{after} (dB)	NRR (dB)
Sound 1	Male	−2.0951	5.8063	7.9014
Sound 2		−0.8544	7.0146	7.869
Sound 3		1.8961	9.7865	7.8904
Sound 4		7.7779	15.2013	7.4234
Sound 5		4.5006	12.1275	7.6269
Sound 6	Female	5.9733	14.2331	8.2597
Sound 7		5.2074	12.7976	7.5893
Sound 8		9.4863	17.5601	8.0738
Sound 9		7.1737	15.2243	8.0506
Sound 10		8.7106	16.2123	7.5018

although the SNR_{before} was various and the noise was sometimes very big.

C. Experiment to confirm the applicability when the number of sensors is less than that of the signals and noise

To show that the proposed method can reduce the internally generated noise even when the number of the sound sources and the noise is larger than that of the acoustical sensors, we conducted another experiment. In this experiment, we only utilized two acoustical sensors, that is, a microphone and a piezoelectric device, while three sound sources, two signals and an internally generated noise, were generated. In the typical microphone array, it is an ill-posed problem. However, when we consider the acoustical array combining microphones and piezoelectric devices, we can reduce the internally generated noise due to its feature in such a case. Sound 1 and sound 2 were generated from 0° and 90°, respectively. We selected the same five male and five female voices utilized in Sec. IV B. In this experiment, to estimate how the method reduced the internally generated noise, SNR is redefined as follows:

$$\mathrm{SNR} = 10 \log_{10} \frac{\sum_{t=0}^{T} s_1^2(t) + s_2^2(t)}{\sum_{t=0}^{T} n^{I2}(t)}, \tag{31}$$

where $s_1(t)$ and $s_2(t)$ represent two signals from 0° and 90°, respectively. $n^I(t)$ represents the internally generated noise. NRR is described as follows:

$$\mathrm{NRR} = \mathrm{SNR}_{after} - \mathrm{SNR}_{before}, \tag{32}$$

where SNR_{before} and SNR_{after} are the SNRs before and after noise reduction, respectively. Table III shows NRR results when the sound 1 (male) and sound 2 (female) were generated from 0° and 90°, respectively. Table IV shows NRR results when the sound 1 (female) and sound 2 (male) were generated from 0° and 90°, respectively. As shown in Tables III and IV, we could reduce the internally generated noise about 7 dB throughout all the experiments although the number of the sound sources and noise was larger than that of the microphone and the piezoelectric device.

TABLE III. NRR results when the sound 1 (male) and sound 2 (female) were generated from 0° and 90°, respectively. The noise was from mobile PC, which include a fan noise and IDD noise.

Female	Male	SNR_{before} (dB)	SNR_{after} (dB)	NRR (dB)
Sound 1	Sound 6	3.1291	11.0603	7.9312
Sound 2	Sound 7	1.051	9.0584	8.0074
Sound 3	Sound 8	6.6582	14.7139	8.0557
Sound 4	Sound 9	6.2487	13.4759	7.2272
Sound 5	Sound 10	6.2586	14.21	7.9514

TABLE IV. NRR results when the sound 1 (female) and sound 2 (male) were generated from 0° and 90°, respectively. The noise was from mobile PC, which include a fan noise and IDD noise.

Female	Male	SNR_{before} (dB)	SNR_{after} (dB)	NRR (dB)
Sound 6	Sound 1	1.8725	9.2062	7.3337
Sound 7	Sound 2	1.5219	9.5398	8.0179
Sound 8	Sound 3	5.2865	13.0687	7.7821
Sound 9	Sound 4	6.4214	13.9774	7.5559
Sound 10	Sound 5	4.911	13.0167	8.1058

D. Experiment to reduce the noise when the sounds are dependent

To evaluate the performance of the proposed method concerning dependent mixtures, we conducted other experiments. We selected the same five male and five female voices utilized in Sec. IV B. The same sounds were generated from 0° and 90°, that is, they were strongly dependent. Table V shows the results of SNR_{before}, SNR_{after}, and NRR regarding the microphone when the same sounds were generated from 0° and 90°. The proposed method could also reduce the noise about 7 dB throughout all the experiments in spite of dependent mixtures and an ill-posed problem in the typical microphone array as shown in Table V.

E. Experiment to reduce the noise when the sound and the noise are dependent

We also conducted the experiments when the sound and the noise are dependent. The sound was generated from 0°. The noise was from mobile PC, which includes not only a fan noise and HDD noise but also the same sound as the

TABLE V. NRR results when the same sounds were generated from 0° and 90°. The noise was from mobile PC, which include a fan noise and HDD noise.

	Gender	SNR_{before} (dB)	SNR_{after} (dB)	NRR (dB)
Sound 1	Male	−2.8021	5.0933	7.8954
Sound 2		−0.0071	8.2029	8.21
Sound 3		1.0766	9.049	7.9724
Sound 4		6.3053	13.869	7.5637
Sound 5		3.0237	11.2656	8.242
Sound 6	Female	4.4725	12.4389	7.9664
Sound 7		4.0269	12.0554	8.0286
Sound 8		8.1689	16.1161	7.9472
Sound 9		6.0253	13.7956	7.7702
Sound 10		7.4798	15.2653	7.7855

TABLE VI. NRR results when the sound was generated from 0°. The noise was from mobile PC, which include not only a fan noise and HDD noise but also the same sound as the sound source.

	Gender	SNR_{before} (dB)	SNR_{after} (dB)	NRR (dB)
Sound 1	Male	−3.7528	0.30322	4.056
Sound 2		−3.3434	0.55565	3.8991
Sound 3		−3.3658	0.6258	3.9916
Sound 4		−7.7146	−4.1349	3.5797
Sound 5		−9.0115	−5.1443	3.8672
Sound 6	Female	2.1359	5.9366	3.8007
Sound 7		−1.7221	2.3778	4.0999
Sound 8		−0.81226	3.3153	4.1276
Sound 9		2.7897	6.7566	3.9669
Sound 10		−1.4438	2.4947	3.9385

sound source. Table VI shows the experimental results. As shown in Table VI, we also could obtain NRR improvements throughout all the experiments.

F. Experiment to evaluate directional patterns

Although NRRs are shown in Secs. IV A–IV E, it is entirely possible that noise is reduced at some frequencies and increased at others, where the intelligibility of the signal is damaged. For instance, if a 100 Hz noise were reduced and a 1000 Hz signal increased, the proposed method would show an improvement, but for a listener, the result would be poor, since the ear is more sensitive at 1000 Hz than at 100 Hz. To show that the proposed method can reduce the noise at wide range of frequencies, we estimated directional patterns concerning 500, 1000, 2000, and 4000 Hz. In all the directional patterns, we used the estimated parameter when the signal with 500 Hz was used. To measure the directivity, we set for the angle θ in the range of 0°–350° at 10° intervals. Figures 8–11 show the directional patterns of the acoustical array combining the microphone and the piezoelectric device with frequencies of 500, 1000, 2000, and 4000 Hz, respectively. In these figures, the points on the circle represent the gain of the proposed array to that of the microphone for each direction. The center point of the circle represents the gain of the proposed array to that of the microphone for the piezoelectric device zone, that is, the gain for the mobile PC position. As shown in Figs. 8–11, the array is able to steer the null point at the noise source and could obtain NRR improvement at the position of the mobile PC around 7 dB with a tiny reduction in the gain on the other areas. These results show that the proposed method can reduce the noise at relatively wide range of frequencies where a normal user is sensitive.

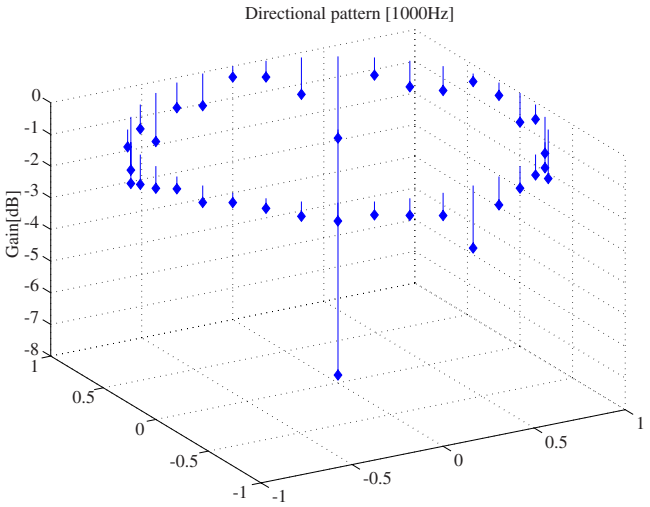


FIG. 9. (Color online) Directional pattern (1000 Hz).

G. Experiment to reduce the noise of an autonomous robot

We next conducted the experiments to apply the proposed method to an actual robot. For the experiments, we utilized an autonomous mobile robot called UBIRO, which was developed in our laboratory. Figure 12 shows UBIRO's appearance. Table VII shows the UBIRO's specifications. As shown in Fig. 12, we set a microphone and a piezoelectric device on the front side of UBIRO. The piezoelectric device (BGT-2000, Belcat) was set on a photoelectronic sensor that had a motor for scanning. A microphone (RP-VC200, Panasonic) was set 3 cm apart from the sensor. The experimental

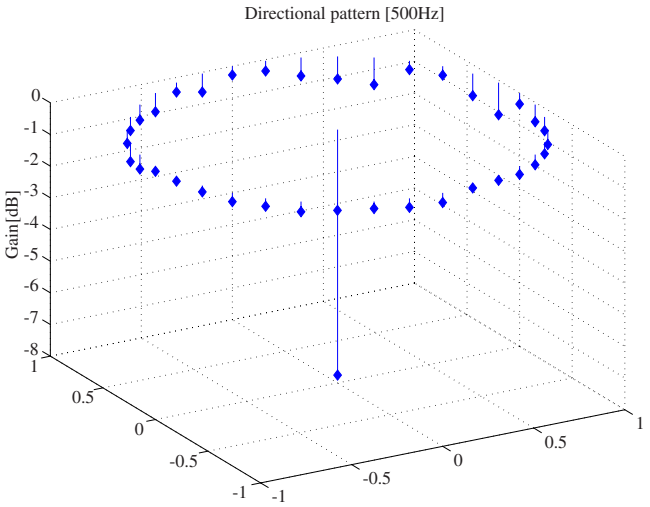


FIG. 8. (Color online) Directional pattern (500 Hz).

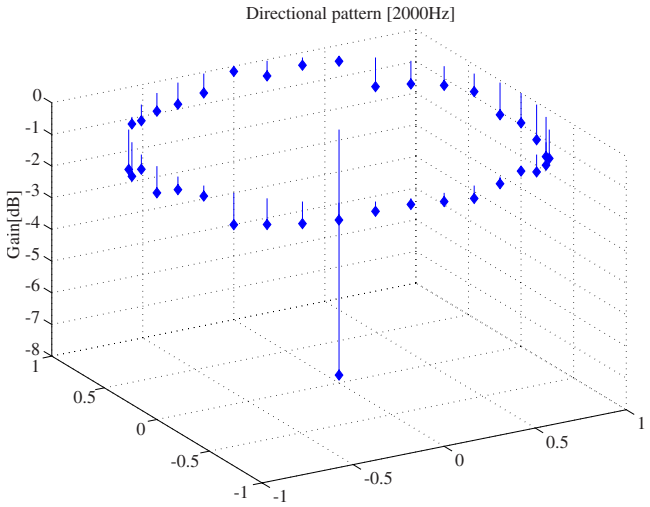


FIG. 10. (Color online) Directional pattern (2000 Hz).

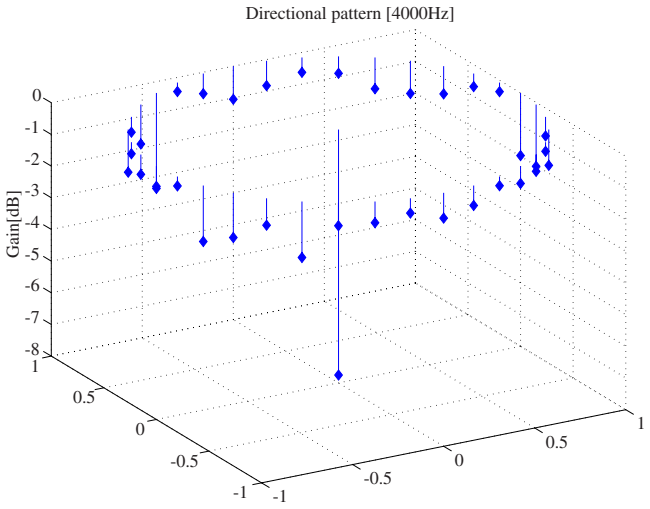


FIG. 11. (Color online) Directional pattern (4000 Hz).

setup is illustrated in Fig. 13. A sound was generated from 0° or 90° and 40 cm apart from the motor. We conducted the experiment not in a special room but in an ordinary conference room. Reverberation time was 387 ms. As sound sources, we selected the same five male and five female voices that were used in the previous experiment. Tables VIII and IX show the results of SNR_{before}, SNR_{after}, and NRR from the microphone when a sound was generated from 0° or 90°, respectively. As shown in Tables VIII and IX, the proposed method could reduce the noise about 4 dB throughout all the experiments even if we applied the proposed method to a real robot. To evaluate the performance of the proposed method concerning dependent mixtures, we also utilized the same voices generated from 0° and 90°; that is, they were strongly dependent. Table X shows the results of SNR_{before} and NRR when two sounds were generated from 0° and 90°, respectively. The proposed method could also reduce the

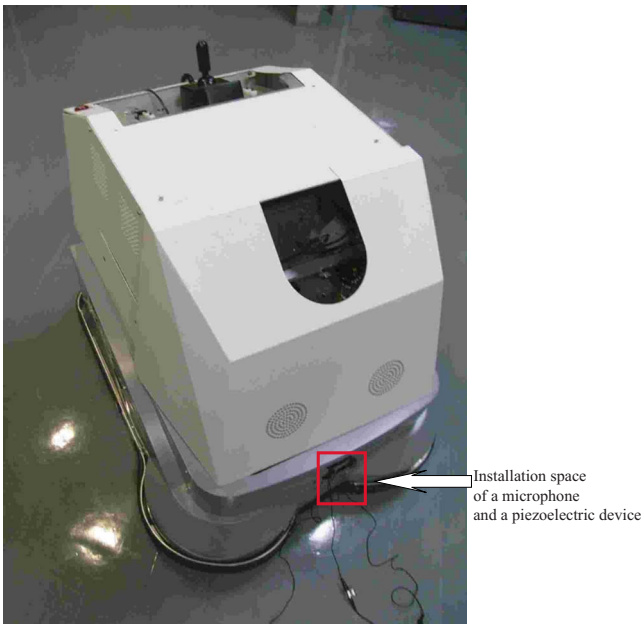


FIG. 12. (Color online) An autonomous mobile robot: UBIRO.

TABLE VII. Specification of UBIRO.

Feature	Description
Size	$800\times900\times900$ mm^3
Weight	90 kg
Speed	Max 6.0 km/h
Operating system	Windows XP (Pentium III 850 MHz)
Communications	Wireless LAN and SR-232C
Batteries	Two 12 V 35 Ah

noise about 4 dB throughout all the experiments in spite of dependent mixtures and an ill-posed problem, as shown in Table X.

H. Experiment to evaluate the performance when the microphone and the piezoelectric device are not collocated

The algorithm is basically based on the same location of the microphones and the piezoelectric device. Since the method does not include a path model between piezoelectric device and microphone, echoes of the internally generated noise could still become injected into the signal. With the same location of the microphones and the piezoelectric device, this is not significant. However, if the piezoelectric device and microphone were not located at the same location, this may become a significant problem. To confirm the effectiveness of the proposed method when the piezoelectric device and microphone were not collocated, we conducted the experiments when the microphone and piezoelectric device are located at the different place. To confirm the above points, we set the microphone 10 cm apart from the mobile PC, while the piezoelectric device is set onto the mobile PC in this experiments. Table XI shows the NRR results. As shown in Table XI, although the microphone is apart from the mobile PC; that is, we could obtain NRR improvement around 4 dB. Hence, the algorithm is still effective even when the microphone and the piezoelectric device are not

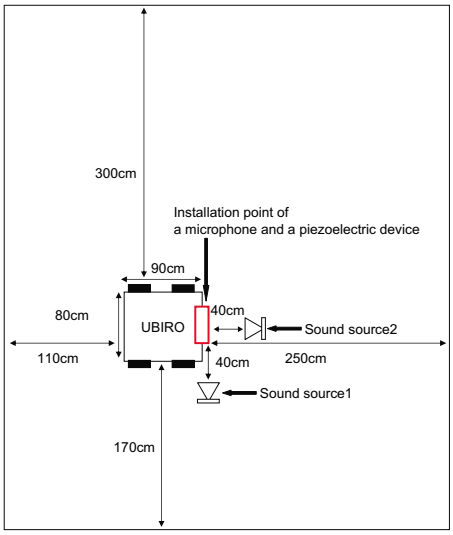


FIG. 13. (Color online) Experimental setup when we utilized a microphone and a piezoelectric device on UBIRO.

TABLE VIII. NRR results when the sound was generated from 0°. The noise source was a photoelectric sensor, which has a motor for scanning on UBIRO.

	Gender	SNR_{before} (dB)	SNR_{after} (dB)	NRR (dB)
Sound 1	Male	−1.7054	−5.6705	3.9651
Sound 2		2.0894	−2.8926	4.982
Sound 3		1.9375	−2.9019	4.8394
Sound 4		4.5473	0.4564	4.0908
Sound 5		1.0416	−3.4544	4.4961
Sound 6	Female	5.9054	0.854	5.0513
Sound 7		7.7625	3.6592	4.1033
Sound 8		9.2579	4.1394	5.1186
Sound 9		7.3163	3.4364	3.8799
Sound 10		8.2059	4.1283	4.0775

TABLE X. NRR results when the same sounds were generated from 0° and 90°. The noise source was a photoelectronic sensor, which has a motor for scanning on UBIRO.

	Gender	SNR_{before} (dB)	SNR_{after} (dB)	NRR (dB)
Sound1	Male	−4.319	7.8604	3.5414
Sound2		−1.3098	5.0072	3.6974
Sound3		−2.0901	7.384	5.2939
Sound4		1.8539	2.1113	3.9652
Sound5		−1.6219	5.1512	3.5293
Sound6	Female	2.1709	2.6648	4.8357
Sound7		1.5686	3.6377	5.2063
Sound8		5.1394	−0.1104	5.029
Sound9		−0.6658	5.0069	4.3411
Sound10		5.2774	−1.093	4.1844

located at the same position although the effectiveness becomes weak when the microphone and the piezoelectric device are located at the different position.

V. DISCUSSION AND CONCLUSION

In this paper, we proposed a method to reduce the internally generated noise based on the gain difference between the microphones and the piezoelectric device attached to the system that generates the internally generated noise. Although we used a piezoelectric device as a vibration sensor, we can employ any vibration sensor instead of the piezoelectric device. The system can reduce the internally generated noise under the blind condition. It can reduce the internally generated noise not only when the number of the acoustical sensors is larger than that of the sound sources and the noise source but also when the number of the acoustical sensors is smaller than that of the sound sources and the noise source. We do not need to know in advance the directionality of the microphones and the piezoelectric device or the interval when the sound is disjoint. Throughout the experiments, we could reduce the noises of mobile PC and the autonomous robot. However, the proposed method also has some limitations summarized as follows:

(1) In our method, we assume that the piezoelectric device is located with the microphone at the same position. Hence, theoretically, the algorithm does not work when the microphone and the piezoelectric device are located at the different positions. Although the algorithm is still empirically effective even when the microphone and the piezoelectric device are not located at the same position, it remains a matter of research.
(2) In our method, we assume that the piezoelectric device is sensitive only to the vibration signal. If there are signal components in the piezoelectric device which are not coupled to the acoustic path between the machine and the microphone, then these components would be added to the microphone signal, degrading performance rather than improving it. Although the assumption is well-satisfied throughout the experiments, this method would not work very much without the piezoelectric device's insensitivity to the speech signal.

Due to these limitations, in some cases, we may also be able to reduce the internally generated noise by using other methods. We consider that we would like to use our proposed method as the pre-processing to combine other methods rather than using our method by itself. For instance, in filtered-X LMS algorithm, the transfer path should be identified in advance to reduce the noise. It also requires the reference signal to reduce the noise. However, in our method, we aim to reduce the internally generated noise

TABLE IX. NRR results when the sound was generated from 90°. The noise source was a photoelectric sensor, which has a motor for scanning on UBIRO.

	Gender	SNR_{before} (dB)	SNR_{before} (dB)	NRR (dB)
Sound 1	Male	−8.2295	4.5379	3.6919
Sound 2		−6.6035	−3.1631	3.4404
Sound 3		−5.5491	−1.4801	4.0691
Sound 4		−3.864	−0.439	3.425
Sound 5		−7.7637	−4.2764	3.4873
Sound 6	Female	−4.6203	−0.4862	4.1341
Sound 7		−1.0484	3.6921	4.7405
Sound 8		−0.4681	4.2458	4.7139
Sound 9		−3.7064	0.6	4.3064
Sound 10		−1.2534	3.1924	4.4458

TABLE XI. NRR results when the microphone is 10 cm apart from the mobile PC. The noise source is from the mobile PC, which includes HDD noise and fan noise.

	Gender	SNR_{before} (dB)	SNR_{after} (dB)	NRR (dB)
Sound1	Male	−1.7086	2.4774	4.1861
Sound2		−2.1546	1.9626	4.1172
Sound3		2.0095	6.4091	4.3997
Sound4		0.76224	5.0896	4.3273
Sound5		−1.7915	2.31	4.1015
Sound6	Female	4.1986	8.5637	4.365
Sound7		3.2478	7.6226	4.3748
Sound8		7.1802	11.724	4.5443
Sound9		6.1213	10.688	4.5667
Sound10		6.4318	10.984	4.5518

TABLE XII. Comparison table of the proposed method with other typical BSS algorithms.

	ICA	Sparseness	Proposed
Can it handle the case when the signals are apart from the noise source?	○	○	△
Can it handle the case when the number of the signal is more than that of the sensor?	×	○	○
Can it handle the case when the frequency bins overlap with each other?	○	×	○
Can it handle the case when the signals are dependent?	×	○	○
Can it handle the case when we employ only microphones?	○	○	×
Can it handle the case when the signals are echoic?	○	○	×

without the identification of the transfer path from the noise source to the sensor. We also do not use the reference signal to reduce the internally generated noise. It is also applicable not only to the stationary noise but also to nonstationary noise. These are examples of the merits of the proposed methods. Due to these features, we could roughly estimate the characteristics of the microphones and the piezoelectric device regarding the noise source under the situations such that the sounds and the noise are present. We think that the proposed method and other methods does not conflict but can coexist. For instance, the proposed method may be used for pre-determining the characteristics of the microphones and piezoelectric device regarding the noise source. And we may next employ some other algorithms to determine the characteristics more precisely.

To clarify the feature of the proposed method, we give a comparison table to show the difference between our method and other BSS algorithms, as shown in Table XII. For future works, we aim to develop a method combining the proposed method and BSS algorithms such as ICA and sparseness approaches. Noise reduction combining the proposed method and vibration control based on piezoelectric devices will also be developed.

ACKNOWLEDGMENT

This research was supported by the research grant of Support Center for Advanced Telecommunications Technology Research (SCAT), by the research grant of Tateisi Science and Technology Foundation, and by the Ministry of Education, Science, Sports and Culture,, Grant-in-Aid for Young Scientists (B), 20700168, 2008. This research was also supported by "Establishment of Consolidated Research Institute for Advanced Science and Medical Care," Encouraging Development Strategic Research Centers Program, the Special Coordination Funds for Promoting Science and Technology, Ministry of Education, Culture, Sports, Science and Technology, Japan, by the CREST project "Foundation of technology supporting the creation of digital media contents" of JST, by the Grant-in-Aid for the WABOT-HOUSE Project by Gifu Prefecture, by Project for Strategic Development of Advanced Robotics Elemental Technologies (NEDO: 06002090), and the Global-COE Program, "Global Robot Academia," Waseda University.

[1] W. Li, C. Miyajima, T. Nishino, K. Itou, K. Takeda, and F. Itakura, "Adaptive nonlinear regression using multiple distributed microphones for in-car speech recognition," IEICE Trans. Fundamentals **E88-A**, 1716–1723 (2005).
[2] J. Cho and A. Krishnamurthy, "Speech enhancement using microphone array in moving vehicle environment," *Proceedings of IEEE Intelligent Vehicles Symposium*, 2003, pp. 366–371.
[3] J.-M. Valin, J. Rouat, and F. Michaud, "Enhanced robot audition based on microphone array source separation with post-filter," *Proceedings of IEEE/RSJ International Conference on Robots and Intelligent Systems*, 2004, pp. 2123–2128.
[4] A. Wang, K. Yao, R. E. Hudson, D. Korompis, and F. Lorenzelli, "Microphone array for hearing aid and speech enhancement applications," *International Conference on Application-Specific Systems, Architectures, and Processors*, 1996, pp. 231–239.
[5] H. Luts, J. Maj, W. Soede, and J. Wouters, "Better speech perception in noise with an assistive multi-microphone array for hearing aids," Ear Hear. **25**, 411–420 (2004).
[6] T. Nishiura, M. Nakamura, A. Lee, H. Saruwatari, and K. Shikano, "Talker tracking display on autonomous mobile robot with a moving microphone array," *Proceedings of International Conference on Auditory Display*, 2002, pp. 244–247.
[7] J.-M. Valin, F. Michaud, B. Hadjou, and J. Rouat, "Localization of simultaneous moving sound sources for mobile robot using a frequency-domain steered beamformer approach," *IEEE International Conference on Robotics and Automation*, 2004, pp. 1033–1038.
[8] K. Kiyohara, Y. Kaneda, S. Takahashi, H. Nomura, and J. Kojima, "A microphone array system for speech recognition," *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, 1997, pp. 215–218.
[9] Y. Kaneda and J. Ohga, "Adaptive microphone array system for noise reduction," IEEE Trans. Acoust., Speech, Signal Process. **34**, 1391–1400 (1986).
[10] K. Takao, M. Fujita, and T. Nishi, "An adaptive antenna array under directional constraint," IEEE Trans. Antennas Propag. **24**, 662–669 (1976).
[11] M. Matsumoto and S. Hashimoto, "A miniaturized adaptive microphone array under directional constraint utilizing aggregated microphones," J. Acoust. Soc. Am. **119**, 352 359 (2006).
[12] J. Capon, "High resolution frequency-wavenumber spectrum analysis," Proc. IEEE **57**, 2408–2418 (1969).
[13] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," IEEE Trans. Antennas Propag. **34**, 276–280 (1986).
[14] M. Matsumoto and S. Hashimoto, "Multiple signal classification by aggregated microphones," IEICE Trans. Fundamentals **E88-A**, 1701–1707 (2005).
[15] M. Matsumoto and S. Hashimoto, "Noise reduction combining microphones and piezoelectric Devices," *Proceedings of International Symposium on Communications, Control and Signal Process*, 2008, pp. 368–371.
[16] M. Matsumoto and S. Hashimoto, "An acoustical array combining microphones and piezoelectric Devices," J. Acoust. Soc. Am. **123**, 2117–2125 (2008).
[17] T. S. Ho, H. Matsuhisa, and Y. Honda, "Passive vibration suppression of beams by piezoelectric elements," Trans. Jpn. Soc. Mech. Eng., Ser. C **66**, 737–743 (2000).
[18] K. Yamada, H. Matsuhisa, and H. Utsuno, "Hybrid vibration suppression of multiple vibrational modes of flexible structures using piezoelectric elements," Trans. Jpn. Soc. Mech. Eng., Ser. C **73**, 461–469 (2007).
[19] K. Takagi, K. Nagase, K. Oshima, Y. Hayakawa, and H. Ichikawa, "Optimization of passive piezoelectric shunt damping based on the robust performance index," Trans. Jpn. Soc. Mech. Eng., Ser. C **69**, 3183–3190 (2003).
[20] A. Belouchrani, K. Abed Meraim, J. F. Cardoso, and E. Moulines, "A blind source separation technique based on second order statistics," IEEE Trans. Signal Process. **45**, 434–444 (1997).
[21] D. Nuzillard and J. M. Nuzillard, "Second order blind source separation on the Fourier space of data," Signal Process. **83**, 627–631 (2003).

[22]K. I. Diamantaras, A. P. Petropulu, and B. Chen, "Blind two-input-two-output FIR channel identification based on frequency domain second-order statistics," IEEE Trans. Signal Process. **48**, 534–542 (2000).
[23]B. Chen and A. P. Petropulu, "Frequency domain blind MIMO system identification based on second- and higher-order statistics," IEEE Trans. Signal Process. **49**, 1677–1688 (2001).
[24]T. Ihara, M. Handa, T. Nagai, and A. Kurematsu, "Multi-channel speech separation and localization by frequency assignment," IEICE Trans. Fundamentals **J86-A**, 998–1009 (2003).
[25]S. Rickard and O. Yilmaz, "On the approximate W-disjoint orthogonality of speech," *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing 2002*, pp. 529–532.
[26]O. Yilmaz and S. Rickard, "Blind Separation of Speech Mixtures via Time-frequency Masking," IEEE Trans. Signal Process. **52**, 1830–1847 (2004).
[27]P. D. O'Grady, B. A. Pearlmutter, and S. T. Rickard, "Survey of sparse and non-sparse methods in source separation," Int. J. Imaging Syst. Technol. **15**, 18–33 (2005).
[28]D. O'Shaughnessy, *Speech Communications: Human and Machine* (IEEE, New York, 1999).

On the applicability of the spherical wave expansion with a single origin for near-field acoustical holography

Jesper Gomes[a)]
Institute of Sensors, Signals and Electrotechnics, University of Southern Denmark, Campusvej 55, DK-5230 Odense M, Denmark

Jørgen Hald
Brüel & Kjær Sound and Vibration Measurement A/S, Skodsborgvej 307, DK-2850 Nærum, Denmark

Peter Juhl
Institute of Sensors, Signals and Electrotechnics, University of Southern Denmark, Campusvej 55, DK-5230 Odense M, Denmark

Finn Jacobsen
Acoustic Technology, Department of Electrical Engineering, Technical University of Denmark, Building 352, Ørsteds Plads, DK-2800 Kgs. Lyngby, Denmark

(Received 29 August 2008; revised 20 November 2008; accepted 12 December 2008)

The spherical wave expansion with a single origin is sometimes used in connection with near-field acoustical holography to determine the sound field on the surface of a source. The radiated field is approximated by a truncated expansion, and the expansion coefficients are determined by matching the sound field model to the measured pressure close to the source. This problem is ill posed, and therefore regularization is required. The present paper investigates the consequence of using only the expansion truncation as regularization approach and compares it with results obtained when additional regularization (the truncated singular value decomposition) is introduced. Important differences between applying the method when using a microphone array surrounding the source completely and an array covering only a part of the source are described. Another relevant issue is the scaling of the wave functions. It is shown that it is important for the additional regularization to work properly that the wave functions are scaled in such a way that their magnitude on the measurement surface decreases with the order. Finally, the method is applied on nonspherical sources using a vibrating plate in both simulations and an experiment, and the performance is compared with the equivalent source method.
 [DOI: 10.1121/1.3068451]

PACS number(s): 43.60.Sx, 43.60.Pt, 43.20.Rz [EGW] Pages: 1529–1537

I. INTRODUCTION

Near-field acoustical holography[1] (NAH) is a technique that reconstructs the sound pressure, particle velocity, and/or sound intensity on the surface of a sound source based on the acoustic field measured at a set of positions close to the source. Knowing precisely the areas with highest noise radiation is an important prerequisite for choosing an effective noise reduction strategy.

One approach within a variety of NAH methods makes use of the spherical wave expansion with a single origin. The idea is to express the sound field as the truncated wave expansion, and then estimate the coefficients of the included expansion functions. Assuming that the truncated series is an acceptable approximation to the true field, the calculated coefficients are reused to estimate the sound field at positions different from the measurement positions, e.g., on the source. This type of approach can be divided into two subcategories. The first subcategory exploits the fact that the expansion functions are orthogonal on a spherical surface, and by sampling the sound field with a spherical microphone array the coefficients can be estimated using numerical integration.[2] The second subcategory finds a set of coefficients to the truncated expansion by fitting the model to the measured data, e.g., by minimization of the difference between the measured sound pressure and the predicted sound pressure at the microphone positions in a least squares sense.[3–7] The latter approach does not rely on the orthogonality of the wave functions, and hence the array does not need to be of spherical shape. The present paper focuses only on this approach.

The truncation is required for practical reasons since the expansion has an infinite number of terms, but it also benefits the reconstruction accuracy because of its inherent regularizing effect. The strong radial decay of the high order terms makes the problem more ill posed, which means that the truncation of the expansion makes the method less sensitive to noise in the measured data. Other ways of regularizing the problem is by use of Tikhonov regularization or truncated singular value decomposition (TSVD).[8] Often, the truncation of the spherical waves is used as the only regularization

[a)]Author to whom correspondence should be addressed. Electronic mail: jgomes@bksv.com. Also at Brüel & Kjær Sound and Vibration, Skodsborgvej 307, DK-2850 Nærum, Denmark.

method. In that case, the SVD, which is required in the TSVD and normally also for Tikhonov regularization, is avoided, which saves computation time. This paper investigates how well the truncation of the spherical wave expansion works as a regularization method, and determines the circumstances where other regularization approaches such as Tikhonov or the TSVD become necessary.

Research within NAH based on the single origin spherical wave expansion (SOSWE) often uses the wave expansion directly as it appears from the mathematical derivation without scaling or normalizing the spherical Hankel functions.[3–7] In Ref. 9 the Hankel functions are normalized on the measurement surface, but without any arguments given for this choice. To the authors' knowledge there are no published studies on the consequence of not scaling the functions. The present paper demonstrates the importance of the scaling and suggests scaling the waves on the surface of the source.

In principle the SOSWE approach has problems with mapping on irregular source surfaces. For example, for exterior problems the spherical wave model is valid only outside the minimum sphere (the smallest sphere that encloses the source with its center at the origin of the spherical expansion). This means that all measurement and calculation points should be outside the minimum sphere, which is possible only if the source surface to be mapped has a spherical shape. However, results have shown that an acceptable accuracy can be achieved even when dealing with nonspherical source surfaces.[3,5,7]

The equivalent source method (ESM)[10–12] will be used as a benchmark method when investigating the performance of the SOSWE approach on nonspherical sources. ESM approximates the sound field by a superposition of the sound fields of simple sources distributed inside the source, and it does not require the source to be of separable geometry.

II. OUTLINE OF THEORY

A. Single origin spherical wave expansion

In this paper only exterior (free-field) problems are considered, and hence the expansion only includes outgoing spherical waves. At a point outside the minimum sphere, $\mathbf{r}=(r,\theta,\phi)$, where θ and ϕ are the zenith and azimuth angle, respectively, the sound pressure can then be expressed as

$$p(\mathbf{r})=\sum_{n=0}^{\infty}\sum_{m=-n}^{n}\tilde{a}_{nm}h_n(kr)Y_n^m(\theta,\phi), \tag{1}$$

where h_n is the nth order spherical Hankel function of the second kind (with the $e^{i\omega t}$ sign convention), k is the wavenumber, $\tilde{a}_{nm}$ are the unknown expansion coefficients, and Y_n^m is a "spherical harmonic."[1]

In the SOSWE method the sound field is approximated as a truncated version of Eq. (1),

$$p(\mathbf{r})\approx\sum_{n=0}^{N}\sum_{m=-n}^{n}\tilde{a}_{nm}h_n(kr)Y_n^m(\theta,\phi)=\sum_{j=1}^{J}\tilde{a}_j\tilde{\Psi}_j(\mathbf{r}), \tag{2}$$

where $\tilde{a}_j=\tilde{a}_{nm}$ and $\tilde{\Psi}_j(\mathbf{r})=h_n(kr)Y_n^m(\theta,\phi)$ are elementary wave functions with $j\in\{n^2+n+m+1\}=\{1,2,\ldots,(N+1)^2\}$. The method assumes that the consequence of truncating the expansion at some number, J, is acceptable. At some limit the high order elementary waves cannot be sampled adequately with a given microphone spacing, which makes it reasonable to truncate the series. Also, the high order Hankel functions have a strong decay in the region between the reconstruction surface and the measurement surface, which causes potentially large errors during the reconstruction process if the measurement noise contains high order components, i.e., the truncation has a regularizing effect.

The magnitude of the radial part of the elementary wave functions in Eq. (2) increases with the order n for a given distance from the expansion origin. As will be demonstrated in Sec. III A, the SOSWE method in general performs better if the elementary wave functions are scaled so that their magnitude decreases on the measurement surface with the order. This can be done by rewriting Eq. (2) as

$$p(\mathbf{r})\approx\sum_{n=0}^{N}\sum_{m=-n}^{n}a_{nm}\frac{h_n(kr)}{h_n(kr_s)}Y_n^m(\theta,\phi)=\sum_{j=1}^{J}a_j\Psi_j(\mathbf{r}), \tag{3}$$

where $a_j=a_{nm}=\tilde{a}_{nm}h_n(kr_s)$ and $\Psi_j(\mathbf{r})=[h_n(kr)/h_n(kr_s)]Y_n^m(\theta,\phi)$ defines a new set of elementary wave functions. The radial distance, r_s, defines the surface of a sphere at which the radial part of the functions has equal amplitude. There are in principle no restrictions on the position of the scaling surface. If it is placed between the expansion origin and the source surface, then the elementary wave functions corresponding to a high order, n, have the lowest magnitude on both the source and measurement surface, and if it coincides with the source surface the magnitude is roughly the same on this surface for all waves (assuming a spherical source shape), whereas it decreases with n on the measurement surface.

Equation (3) can be used to express the sound pressure at M field positions, and on matrix form this yields

$$\mathbf{p}=\mathbf{B}\mathbf{a}, \tag{4}$$

where $[\mathbf{a}]^{J\times 1}$ contains the coefficients, a_j, $[\mathbf{p}]^{M\times 1}$ contains the sound pressure at M field points, and the elements of $[\mathbf{B}]^{M\times J}$ are the values of the jth elementary wave at the mth field point.

If $\mathbf{p}$ contains measured pressure data, then the expansion coefficients are the only unknowns in Eq. (4). The number of waves and the number of measurement points are not necessarily equal, yielding an overdetermined or underdetermined system of equations. To get a unique solution the problem is solved using least squares and least norm for $M\geq J$ and $M<J$, respectively,

$$\mathbf{a}=\begin{cases}(\mathbf{B}^H\mathbf{B})^{-1}\mathbf{B}^H\mathbf{p} & \text{for } M\geq J\\ \mathbf{B}^H(\mathbf{B}\mathbf{B}^H)^{-1}\mathbf{p} & \text{for } M<J,\end{cases} \tag{5}$$

where ^{H} denotes the Hermitian transpose.

The computation of the SVD yields $\mathbf{B}=\Sigma_{i=1}^{I}\mathbf{u}_i\sigma_i\mathbf{v}_i^H$, where $\mathbf{u}_i$ and $\mathbf{v}_i$ are the left and right singular vectors, respectively, and σ_i are the singular values. If only the terms corresponding to nonzero singular values are included, i.e., $I=\text{rank}(\mathbf{B})$, Eq. (5) can then be written as

$$\mathbf{a} = \sum_{i=1}^{I} \frac{\mathbf{u}_i^H \mathbf{p}}{\sigma_i} \mathbf{v}_i \tag{6}$$

for both the overdetermined and underdetermined cases. If the condition number of $\mathbf{B}$ is high, additional regularization is needed in order to prevent strong amplifications of the measurement noise in $\mathbf{p}$. This can be done using the TSVD to get a regularized solution vector

$$\mathbf{a}_{\text{reg}} = \sum_{i=1}^{I_{\text{reg}}} \frac{\mathbf{u}_i^H \mathbf{p}}{\sigma_i} \mathbf{v}_i. \tag{7}$$

The truncation parameter, I_{reg}, can be determined using automated parameter choice methods such as the L-curve analysis or generalized cross-validation (GCV).[8]

As mentioned earlier, the truncation of the wave expansion is often used as the only type of regularization, which means that the expansion coefficients can be found directly from Eq. (5). The computation time increases if the TSVD is introduced in addition, but as demonstrated in Sec. III C, in many cases it will give a more accurate reconstruction result.

After the calculation of $\mathbf{a}_{\text{reg}}$, these coefficients can be reused in Eq. (3) to calculate the pressure on the source surface, and the particle velocity on the surface of the source can be obtained by use of Euler's equation.[1]

B. Equivalent source method

ESM will be used for benchmarking in this paper. The principle behind the method is quite similar to that of SOSWE, but it is based on a distributed set of equivalent sources, i.e., multiple origins, and the method normally only uses the monopoles from the wave expansion. The sound field in a point $\mathbf{r}$ is estimated as

$$p(\mathbf{r}) \approx \sum_{j=1}^{J} a_j G(\mathbf{r}, \mathbf{r}_j), \tag{8}$$

where $G(\mathbf{r}, \mathbf{r}_j)$ is the free-space Green's function with its origin at a point $\mathbf{r}_j$. The pressure at a set of field points can then be expressed in the same form as for SOSWE in Eq. (4), and the coefficients can be found by use of, for instance, Eq. (7). Finally, the resulting coefficients of the equivalent sources, $\mathbf{a}_{\text{reg}}$, can be used to estimate the pressure and/or particle velocity on the source surface.

III. SIMULATED MEASUREMENTS

In all the test cases that follow, only the prediction of the particle velocity is considered, and the resulting reconstruction error is calculated as

$$\frac{\|\tilde{\mathbf{v}} - \mathbf{v}\|_2}{\|\tilde{\mathbf{v}}\|_2} \cdot 100\,\%, \tag{9}$$

where $\tilde{\mathbf{v}}$ is a vector of the "true" velocities at a set of reconstruction points, and $\mathbf{v}$ contains the estimated velocities. Imperfections in the setup are simulated by introducing normal distributed random amplitude and phase mismatch between the microphones with standard deviations of 0.2 dB and 1°, respectively. The resulting signal-to-noise ratio (SNR) was approximately 30 dB in all the simulated results. Tests have also been made with random background noise with the same SNR, but this did not have any effect on the overall conclusions. Hence, only the results based on mismatch errors are shown in the present paper.

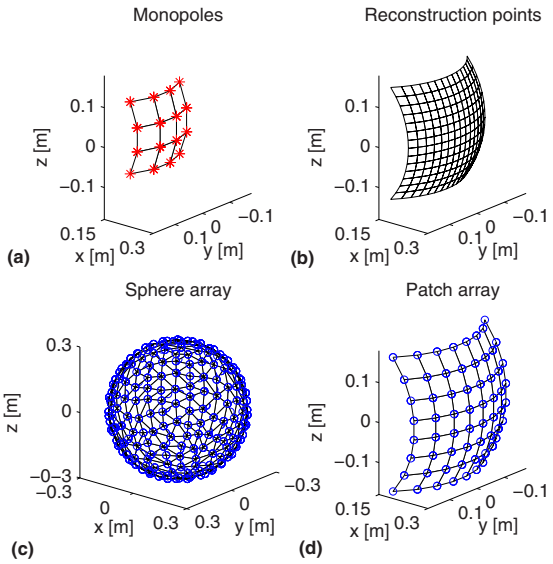


FIG. 1. (Color online) Simulated test case with (a) 4×4 monopoles located on a patch of a virtual spherical surface with a radius of 0.2 m, (b) 15×15 reconstruction points on a patch of a another spherical surface with a radius of 0.25 m, (c) an array of 303 measurement positions located on sphere with a radius of $r_m = 0.3$ m, and (d) an 8×8 array with all measurement points located at $r_m = 0.3$ m.

The first test case to be considered consists of monopoles that are distributed on a patch of a virtual spherical surface with a radius of 0.2 m, as shown in Fig. 1(a). Their positions correspond to a regular grid of 4×4 points in the yz-direction with a spacing of 6 cm, which is projected onto the virtual sphere. The phase difference between neighboring monopoles is set to 180° to create an evanescent sound field. The reconstruction points in Fig. 1(b) correspond to 15×15 points with a spacing of 2 cm in the yz-direction, which are projected onto a spherical surface of radius $r_s = 0.25$ m. Two different simulated arrays will be considered. In Fig. 1(c) there are 303 measurement positions placed on a spherical surface with a radius of $r_m = 0.3$ m with an average spacing of 5 cm between the measurement positions. The patch array in Fig. 1(d) is the projection of every other of the 15×15 reconstruction points [see Fig. 1(b)] onto the spherical surface with a radius of $r_m = 0.3$ m, yielding 8×8 points covering an area slightly larger than the reconstruction surface. The origin of the spherical wave expansion will be placed at $\mathbf{r} = (0, 0, 0)$ m.

A. Scaling the wave functions

The influence of the scaling on the reconstruction accuracy is investigated using the simulated setup just mentioned. However, only the sphere array in Fig. 1(c) will be considered in this section.

The SVD of the matrix, $\mathbf{B}$, gives valuable insight into the problem to be solved, and Fig. 2 shows the first and last three left singular vectors with and without the scaling of the

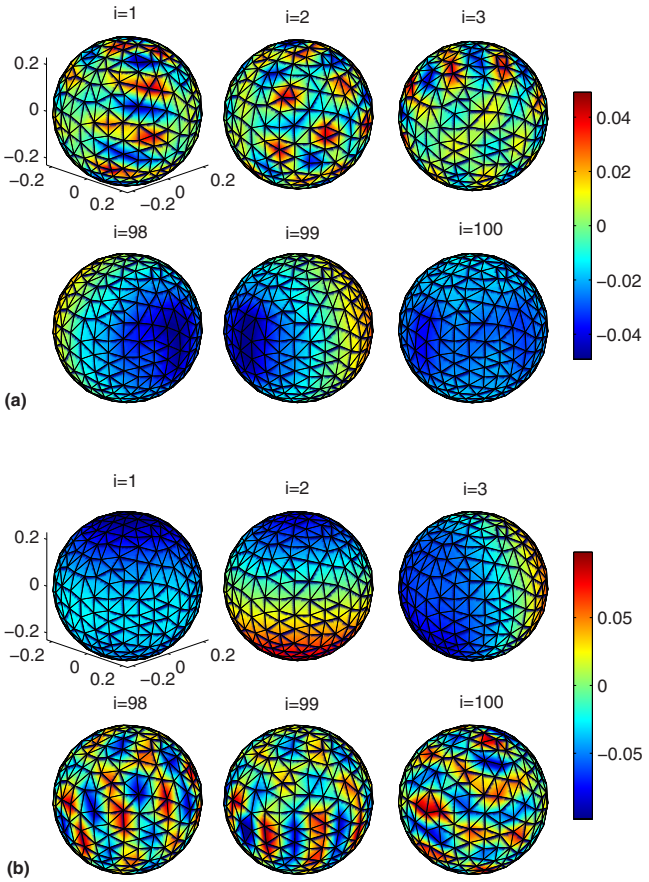


FIG. 2. (Color online) Examples of real part of the left singular vectors, $\mathbf{u}_i$, plotted on the spherical measurement surface with radius, $r_m=0.3$ m. The frequency is $f=500$ Hz, and the number of expansion terms is set to $J=100$. (a) Without scaling the elementary wave functions; (b) With the elementary waves scaled on a sphere with radius, $r_s=0.25$ m.

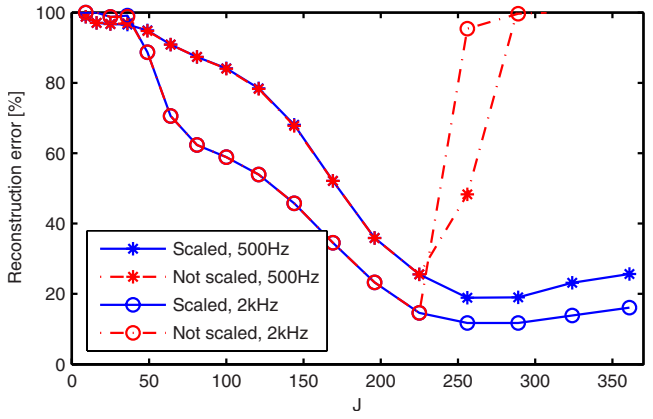


FIG. 3. (Color online) Error in reconstructed velocity in the 15×15 reconstruction points as a function of the truncation of the elementary waves, J. The TSVD is used for all results with the regularization parameter chosen optimally.

elementary wave functions, i.e., Fig. 2(a) is based on the elementary waves, Ψ_j, in Eq. (2), while Fig. 2(b) is based on $\tilde{\Psi}_j$ from Eq. (3). The singular vectors with high index i correspond to the small singular values, and normally in NAH these singular vectors mainly contain the high spatial frequencies. This is clearly not the case in Fig. 2(a); in fact, the spatial frequency seems to decrease with i. In Fig. 2(b) the waves are scaled on a sphere with radius, $r_s=0.25$ m, and as a result, the left singular vectors with low indices are the least fluctuating ones. If the elementary waves are not scaled, the amplitude of the radial part of the high order expansion terms will be significantly higher than those of low order. For the present setup the SVD decomposes $\mathbf{B}$ into basis vectors that behave much like the elementary waves, and the high amplitudes of the radial part will therefore be seen as large singular values associated with the left singular values with high spatial frequencies.

As a consequence of the opposite behavior of the singular vectors, the first basis vectors to be removed from the solution when applying the TSVD in Eq. (7) are those with the lowest spatial frequencies. This is undesirable since these wave components are very important in the solution. Hence, it is reasonable to expect that the TSVD (or Tikhonov) will not benefit the solution at all for this setup.

The particle velocity created by the monopoles will now be estimated in the normal direction of the reconstruction surface, and Fig. 3 shows the relative errors as a function of the truncation of the spherical wave expansion with and without the wave scaling. The upper limit, $J=(18+1)^2=361$, is set so that the elementary waves can be sampled on equator and in the θ-direction with approximately two measurement positions per half-period. The regularization parameter in the TSVD is chosen optimally, i.e., yielding the lowest possible reconstruction error. Up to $J=225$ the optimal results are obtained when no regularization is introduced, which is the reason why the curves coincide up to this truncation number. Above this value the error does not change much for the scaled version since the terms above $J=225$ are effectively removed by the TSVD. For the unscaled waves, however, the SVD cannot be truncated because the low order components are important in the solution.

B. Open measurement surface

In practice it is often not possible to measure the sound field on a surface that surrounds the source. The next simulation uses the open measurement surface from Fig. 1(d) that covers only a patch on the spherical surface from before. This type of array will be denoted as a patch array. The elementary waves are again scaled on the reconstruction surface ($r_s=0.25$ m).

The left singular vectors for the patch array are shown in Fig. 4. The vectors seem to be ordered in the traditional way with the spatial frequencies increasing with i. It should be kept in mind, though, that an elementary wave with high order might not vary much across the small patch, which means that high order elementary waves then behave as singular vectors with low index. As an example, the elementary wave function, Ψ_{100}, from Eq. (3) only has approximately one half-period across the patch in the θ-direction and two periods in the ϕ-direction (not shown). This means that some elementary waves with relatively high order mainly consist of combinations of singular vectors with relatively low index. The singular vectors with low index might even contribute significantly to both a high and a low order elementary wave, respectively. The remains, which cannot be repre-

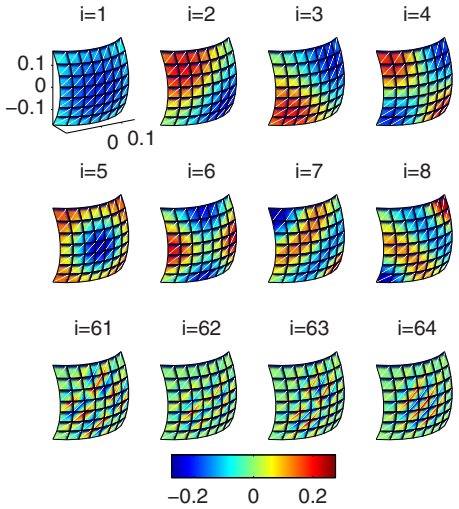


FIG. 4. (Color online) Real values of the left singular vectors, $\mathbf{u}_i$, plotted on the open measurement surface. The elementary waves are scaled on r_s =0.25 m; J=100 and f=500 Hz.

sented by the first singular vectors, will be seen as the noise-like singular vectors in the bottom row in Fig. 4.

To analyze the inverse problem further, Eq. (6) is multiplied by $\mathbf{v}_i^H$ yielding

$$\mathbf{v}_i^H \mathbf{a} = \frac{1}{\sigma_i} \mathbf{u}_i^H \mathbf{p}, \quad i = 1,2,\ldots,I, \tag{10}$$

where the orthogonality of the right singular vectors, $\mathbf{v}_i$, has been used. The inner product, $\mathbf{u}_i^H \mathbf{p}$, describes the ith left singular vector's contribution to the measured pressure, and the jth element in $\mathbf{v}_i$ is multiplied on the amplitude of the jth elementary wave, which means that the magnitude of the jth element in $\mathbf{v}_i$ is proportional to the participation factors of the jth elementary wave. In the simple case where all the coefficients are unity, the left-hand side becomes a sum of the elements in $\mathbf{v}_i^H$, and the element with highest amplitude participates the most to the inner product, $\mathbf{u}_i^H \mathbf{p}$. Figure 5 shows the modulus of the elements of the matrix, $\mathbf{V}^H = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_I\}^H$, where I=rank($\mathbf{B}$), when using the sphere array and the patch array, respectively. For the sphere array the high-magnitude elements are mainly located around the diagonal, which means that the elementary waves with low j (the first columns in $\mathbf{V}^H$) are closely related to the singular vectors with low i (the first rows in $\mathbf{V}^H$) and vice versa. Notice that the matrix is very well divided in clusters with $2n+1$ rows and columns, where n is the order of the wave functions belonging to the cluster. For the patch array the concentration of high-magnitude elements is more blurred and the matrix is not dominated around the diagonal. The singular vectors with relatively low index, i, involve elementary waves with relatively high index, j, and already at $i \approx 15$ all the included terms have come into play. This is in agreement with the arguments given in connection with Fig. 4 regarding the fact that a high order elementary wave does not necessarily vary much across the patch.

Knowing that the singular values decrease when i is increased, it can be concluded from Fig. 5 that for the sphere array, the level of the singular values is directly related to the

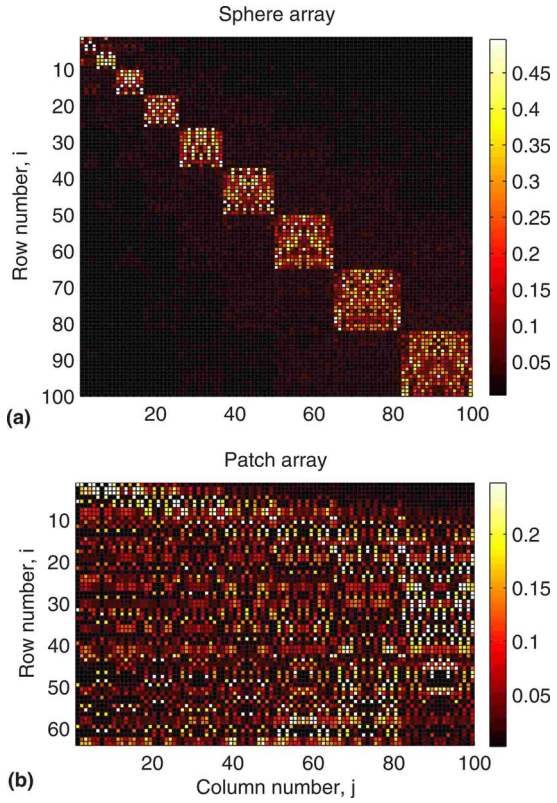


FIG. 5. (Color online) Modulus of the matrix elements in $\mathbf{V}^H$ for f =500 Hz and J=100, (a) with the sphere array, and (b) with the 8×8 array covering only a part of the spherical measurement surface.

order of the elementary waves. This is not the case for the patch array, which is recognized by the fact that $\mathbf{V}^H$ is not dominated around its diagonal.

Figure 6 shows the singular values for both the patch array and the sphere array. The plot shows that σ_i drops faster for the patch array, which can again be explained by the fact that for the patch array the elementary waves' variation across the array area consists of combinations of the singular vectors with low index i, and those with high index in Fig. 4 make up a small contribution to the elementary waves, resulting in small singular values. That is, there exists some level of linear dependence between the elementary

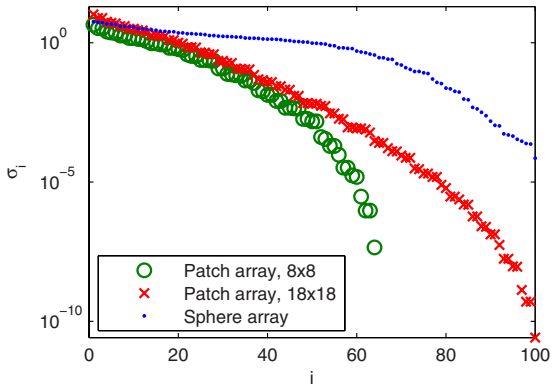


FIG. 6. (Color online) Singular values for an open (patch) array and a closed (sphere) array. Two patch arrays are used: One with 8×8 and one with 18×18 measurement positions; the two patch arrays cover the same area on the measurement sphere. The elementary waves are scaled on r_s=0.25 m; J=100 and f=500 Hz.

waves seen across the small patch area and this linear dependence translates into a faster decay of σ_i. For the sphere array the elementary waves behave very differently seen over the closed measurement surface. In fact, the elementary waves are even orthogonal when considering the entire sphere, and therefore the wave functions are "more linearly independent" than for the patch array, which results in a slower decay of the singular values. Figure 6 also shows the singular values corresponding to a patch array with 18×18 measurement positions, covering the same area as the 8×8 array. This array is included in the plot to show that the conclusion also holds for a patch array with roughly the same number of microphones as the sphere array. It should be emphasized that the measurement surface need not be of spherical shape to have a low condition number, i.e., a small ratio between the largest and the smallest singular value. For instance, spot checks with a box-shaped measurement surface also resulted in matrix with a high concentration around the diagonal as the one in Fig. 5 and small condition number (not shown).

C. Comparison of the regularization approaches

In this section, the SOSWE results are compared with and without the use of the TSVD, respectively. The test case from Fig. 1 is used again, and the resulting reconstruction errors are shown in Fig. 7 for the 8×8 patch array and for the sphere array. From Fig. 7(a) it is seen that up to $J=225$ the TSVD is unnecessary when using the sphere array. The condition number increases with J due to the strong decay of the elementary waves with highest order, and at $J>225$ the condition number becomes so large that the error of the unregularized solution starts to increase. Hence, for this array configuration the TSVD can be omitted if the truncation number, J, is chosen optimally. However, this requires an automated method for the selection of J, and a wrong selection will result in large errors. When introducing the TSVD, the error does not increase much at high values of J because the part of the solution stemming from the small singular values is removed. Therefore, when the TSVD is used, the method is not very sensitive to the choice of J, as long as it is chosen large enough.

As mentioned, the TSVD is superfluous at low values of J for the sphere array, but as seen in Fig. 7(b), this is not the case for the patch array. The noise in $\mathbf{p}$ will in general have components as those in the bottom row of Fig. 4 and the associated small singular values result in an unwanted amplification of these components during the reconstruction. Therefore, the additional regularization is required for most of the values of J in Fig. 7(b). These plots clearly show that there is a risk of getting high errors if additional regularization is not implemented in the SOSWE; at least if the measurement surface covers only a patch on an otherwise closed surface.

D. Nonspherical source shape

Very often in applied acoustics the vibrator of interest has a nonspherical shape. The theory behind SOSWE dictates that the source must be of spherical geometry in order to reconstruct the sound field on its surface, but as mentioned, earlier results have shown reasonable accuracy for the method when applied on nonspherical source shapes as well. This section investigates the reconstruction errors using a planar vibrator.

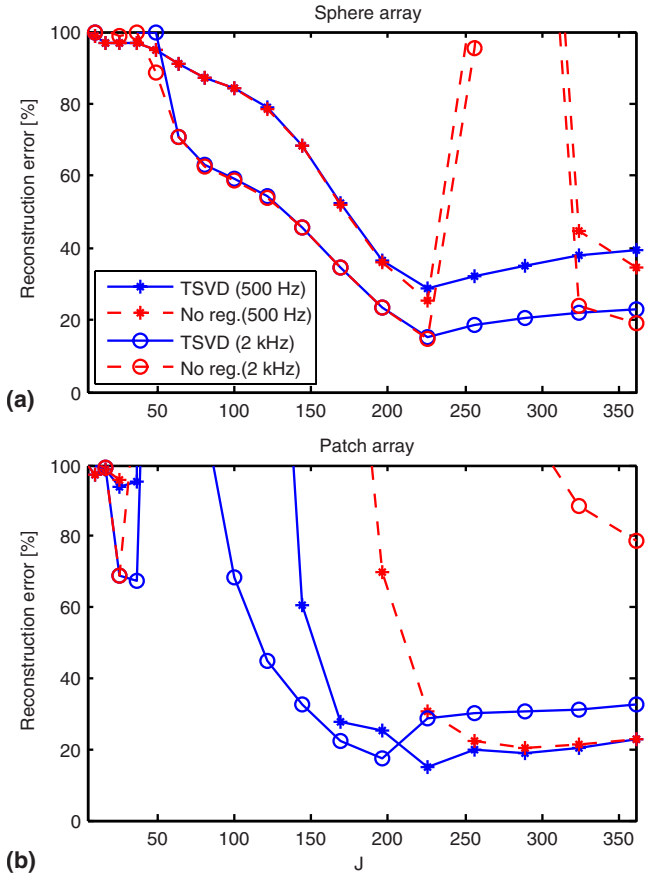


FIG. 7. (Color online) Error in the velocity as a function of the number of elementary waves included with (a) patch array, and (b) sphere array. The dashed lines are based on Eq. (5), i.e., without the TSVD as additional regularization, and the solid lines are based the TSVD solution in Eq. (7) with GCV as parameter choice method. The waves are scaled at $r_s=0.25$ m.

The simulated source object is a $0.08\times0.36\times0.21$ m^3 rectangular box with five rigid sides and the sixth side being a 0.36×0.21 m^2 simply supported steel plate. The plate is point excited near a corner, and the radiated sound field is calculated using the boundary element method. The sound pressure is computed at 8×8 field positions situated 3 cm from the plate (see Fig. 8). Microphone mismatch is again added to the sound pressure, and the result is used as input to SOSWE to find a set of expansion coefficients. Finally, the particle velocity is estimated at 15×15 points on the plate directly under the array. The waves are scaled on a sphere with a radius equal to the average distance between the expansion origin and the reconstruction points.

Two thicknesses have been used for the plate, 5 and 1.5 mm. In general, the dominating modes in the steel plate at a given excitation frequency will have a shorter spatial wavelength for the thin plate. This means that the radiated sound field is more evanescent for the 1.5 mm plate than for the 5 mm plate. The resulting reconstruction errors are shown in Fig. 9 for the 5 mm plate at two different frequen-

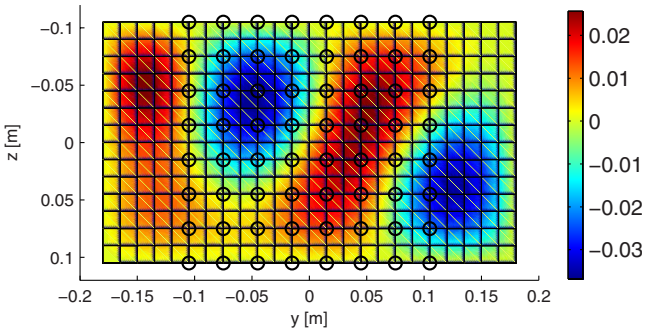


FIG. 8. (Color online) Imaginary part (the real part i zero) of the normal velocity in m/s of a simply supported steel plate (in a rigid rectangular box). The circles indicate 8×8 microphone positions placed 3 cm from the plate. The displayed vibration pattern is for a 5 mm plate point excited at 500 Hz.

cies. Each curve on the plots corresponds to a new distance between the expansion origin and the measurement plane, a. The results show that there is an interrelation between the optimal number of elementary waves and the distance to the expansion origin, and the optimal combination of a and J clearly depends on the frequency. In general, the further away the origin is, the higher is the optimal number of functions. This can be explained by the fact that the elementary waves propagate for $ka > n$, which means that the order, n, must be high to represent the evanescent behavior of the field when a is large.

From Fig. 9 it seems that good results can be achieved if

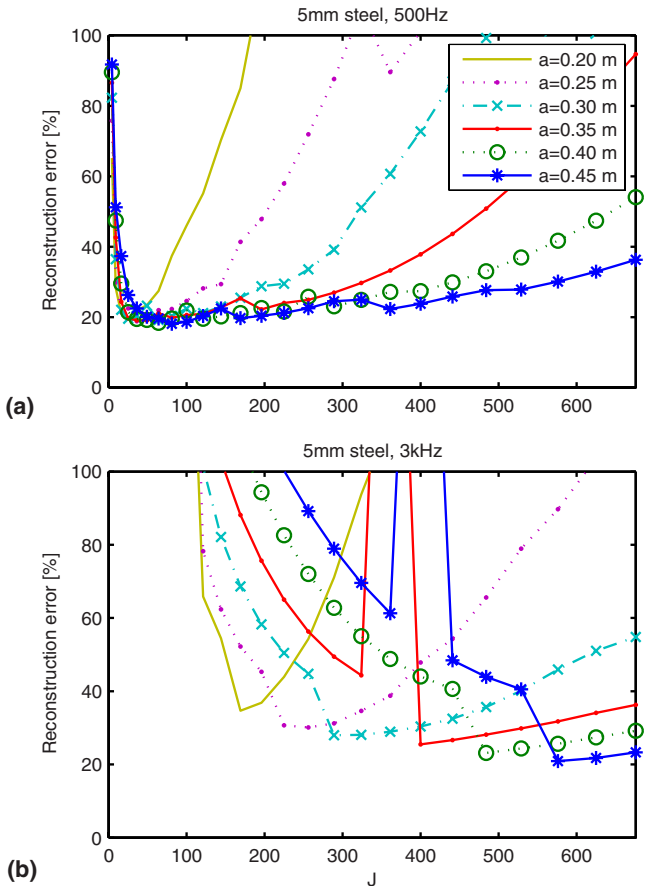


FIG. 9. (Color online) Relative error in the reconstucted particle velocity on the vibrating 5 mm steel plate for different distances, a, to the expansion origin. GCV is used to find a regularization parameter for the TSVD.

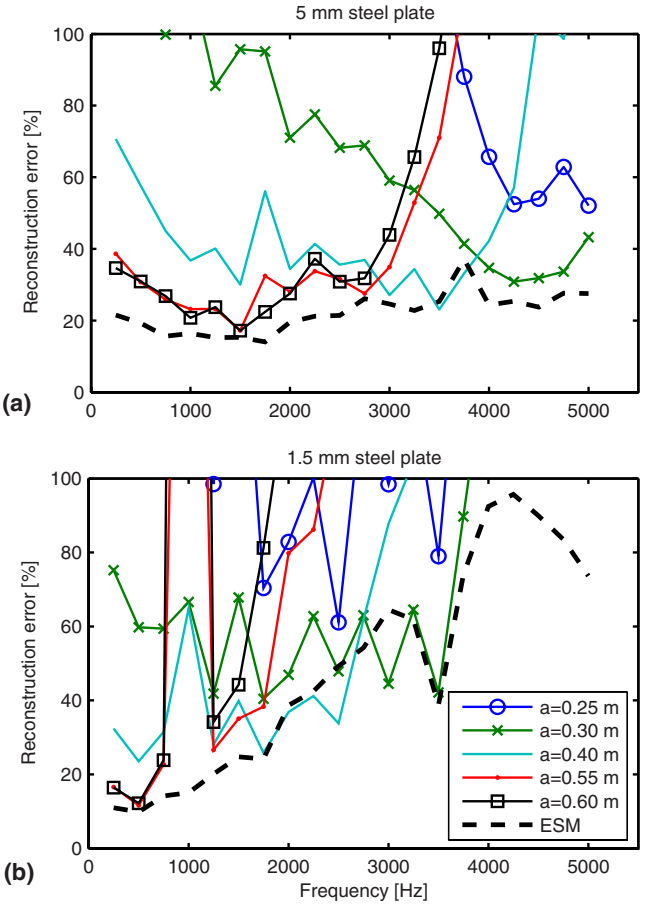


FIG. 10. (Color online) Reconstruction error as a function of frequency; (a) 5 mm steel plate, and (b) 1.5 mm steel plate. Each curve corresponds to a new distance to the expansion origin, a; $J=676$; GCV is used with the TSVD. The ESM results are also shown for the sake of benchmarking.

J is set high and if the origin is adjusted accordingly. Figure 10 shows the error as a function of frequency for the two plates using different positions of the expansion origin with a fixed number of terms, $J=676$. No single value for a benefits the whole frequency range. For the low frequencies the best results are obtained with the origin far from the plate, and for the high frequencies the preferred origin is closer to the plate. For the furthest origins, e.g., $a=0.6$ m, better results are achieved if a higher value of J is used (not shown), but the choice for J is a compromise between computational cost and the (assumed) reconstruction accuracy.

The errors are also shown for ESM in Fig. 10 for benchmarking. The equivalent sources (monopoles) are positioned below the plate at a distance equal to the spacing of the reconstruction points (1.5 cm). There is one monopole below each reconstruction point, and then two additional rims of monopoles at the edges, i.e., 19×19 in total with a 1.5 cm spacing. ESM is clearly more robust and accurate than SOSWE, and the position of the monopoles does not have to be varied with the frequency in order to get satisfactory results.

IV. EXPERIMENTAL RESULTS

Some experiments have been carried out in the anechoic room at the University of Southern Denmark. The source

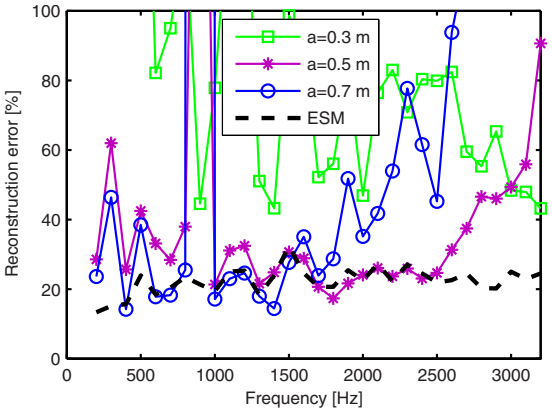


FIG. 11. (Color online) Experiment with a vibrating steel plate. Error in reconstructed velocity using SOSWE with different expansion origins and ESM.

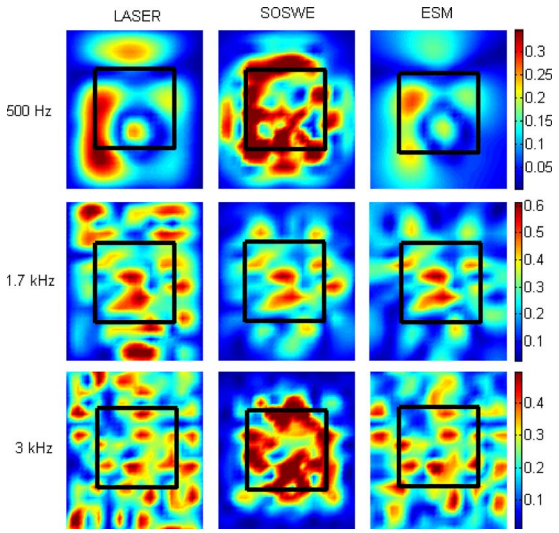


FIG. 12. (Color online) Measured (first column) and estimated (second and third column) normal velocity in (m/s)/N on the steel plate. The black squares indicate the position of the array. The origins in SOSWE are found from the optimization procedure: $a=0.4$ m at 500 Hz, $a=0.5$ m at 1.7 kHz, and $a=0.3$ m at 3 kHz.

object is a $0.4\times0.4\times0.5$ m^3 rectangular box of 19 mm fiberboard with one side replaced by a 3 mm steel plate.

The steel plate was excited at a point near its center with broad band noise using a Brüel & Kjær (B&K) exciter (type 4809). A stinger was mounted on the exciter, and a force transducer connected the stinger to the plate (using beeswax on the transducer). The pressure was measured with a B&K two-layer microphone array with 8×8 1/4 in. microphones (type 4959) in each layer. The microphones constitute a regular grid with 3 cm spacing, and only the front layer of microphones was used. The array was positioned symmetrically with respect to the steel plate at a standoff distance of 3 cm. A B&K "PULSE" analyzer with 65 channels was used in all measurements.

The normal velocity of the steel plate was measured with a laser vibrometer (Ometron VH-1000-D) in 16×14 points covering the entire plate with a spacing of 3 cm between the points. The force transducer was used as a reference for both pressure and velocity measurements, and to circumvent problems such as drift in the setup, only transfer functions (between force transducer and laser/microphone) were used. Hence all values in the following are given in (m/s)/N instead of m/s.

The results from the laser are considered as the true velocity, and Eq. (9) is used as an error measure taking into account only the 8×8 points in front of the array. Since the difference in phase response between the laser and the microphones is unknown, only absolute values are used in the error calculation.

In the results that follow, the truncation of the wave expansion is set to $J=676$, and GCV is used as parameter choice method in the TSVD. The reconstruction errors using SOSWE with different origins are displayed in Fig. 11. As for the simulated steel plate in Sec. III D, the optimal position of the origin is further away from the plate at the low frequencies and closer to the plate for the higher frequencies. The ESM errors are also shown in the plot, and again, this method is as good, or better, than SOSWE. The equivalent sources in ESM are distributed in the same way as in the simulations.

The optimal location of the origin in SOSWE depends strongly on the number of terms included, the frequency, and on the source object. In Ref. 6 an optimization procedure is suggested to find an estimate to the optimal value for J with a fixed expansion origin. In the following, a similar procedure will be applied, but with J being fixed and the expansion origin being the optimization parameter. The measurement points are divided into two groups (M_1 and M_2), and then the pressure data from M_1 are used as input to the SOSWE algorithm to estimate the pressure in M_2. This is repeated for different origins, and the origin that yields the lowest error in M_2 is assumed to be the optimal choice. It should be emphasized, though, that the lowest error in M_2 does not necessarily correspond to the lowest error on the source surface. Every other measurement position is used for M_1, and the reconstruction error at the remaining points, M_2, is found for $a=\{0.25,0.3,\ldots,0.7\}$ m. That is, there are 32 measurement positions in each group, and no neighboring positions belong to the same group. The resulting reconstruction results are shown in Fig. 12 together with the laser results and the ESM-based reconstruction. The plots show the reconstruction results on the entire steel plate (the error calculations in Fig. 11 are based only on the area below the array). The origins used in SOSWE are selected by the optimization procedure, and the erroneous reconstruction pattern at $f=500$ Hz and $f=3$ kHz is due to an inappropriate choice of the origin. That is, the origin that yields the lowest error in M_2 does not correspond to a well reconstructed pattern on the source surface. ESM predicts the velocity pattern quite well in the region below the array and, in fact, also in the region just outside the measurement area. In general, however, the ESM prediction should only be trusted in the area covered by the array.

V. CONCLUSION

The SOSWE approach is sometimes used for NAH purposes, and the truncation of the wave expansion is often used as the only regularization method in order to prevent the

unwanted amplifications of noise during the reconstruction process. The results in the present paper showed that when the measurement points do not constitute a closed surface surrounding the expansion origin, but only a small patch, the truncation of the wave expansion is in general not sufficient to get accurate results since the condition number can be high even for a relatively low number of expansion terms when using the open array. Therefore, additional regularization such as the TSVD or Tikhonov regularization should be applied in connection with SOSWE methods when using open measurement surfaces. The TSVD can be omitted when using a spherical array surrounding the source, but since the truncation of the expansion then works as the only regularization, the number of terms should be selected carefully.

It has been shown that in order for the additional regularization to work properly the spherical wave functions should be scaled in such a way that the waves with highest order have the lowest amplitude on the measurement surface. This can be done by scaling the functions on the surface of the source. By doing that, the regularization will mainly remove information from the high order terms, leaving the information from the low order terms in the regularized solution.

Acceptable accuracy can be achieved with the method on sources of nonspherical geometry, but the errors depend strongly on the choice of the expansion origin and the number of terms included. However, as shown with simulations and practical measurements, if one of these parameters is fixed the other most be changed with the frequency to get useful results. The optimal value of this parameter is unknown in practice, which leaves a risk of high reconstruction errors. This is a main drawback of the method, and as shown in this paper the ESM, which uses distributed monopoles to represent the sound field, is a more robust and accurate method.

[1]E. G. Williams, *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography* (Academic, London, 1999).

[2]E. G. Williams, N. Valdivia, P. C. Herdic, and J. Klos, "Volumetric acoustic vector intensity imager," J. Acoust. Soc. Am. **120**, 1887–1897 (2006).

[3]Z. Wang and S. F. Wu, "Helmholtz equation-least-squares method for reconstructing the acoustic pressure field," J. Acoust. Soc. Am. **102**, 2020–2032 (1997).

[4]S. F. Wu, "On reconstruction of acoustic pressure fields using the Helmholtz equation least squares method," J. Acoust. Soc. Am. **107**, 2511–2522 (2000).

[5]N. Rayess and S. F. Wu, "Experimental validation of the HELS method for reconstructing acoustic radiation from a complex vibrating structure," J. Acoust. Soc. Am. **107**, 2955–2964 (2000).

[6]M. S. Moondra and S. F. Wu, "Visualization of vehicle interior sound fields using nearfield acoustical holography based on Helmholtz-equation least-squares (HELS)," Noise Control Eng. J. **53**, 145–154 (2005).

[7]E. G. Williams, "Approaches to patch NAH," Proceedings of Inter-Noise 2003, 2003, pp. 2187–2194.

[8]P. C. Hansen, *Rank-Deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion* (Society for Industrial and Applied Mathematics, Philadelphia, PA, 1997).

[9]J. Prager, "Approximate reconstruction of sound fields close to the source surface using spherical nearfield acoustical holography," J. Acoust. Soc. Am. **122**, 2067–2073 (2007).

[10]A. Sarkissian, "Method of superposition applied to patch near-field acoustical holography," J. Acoust. Soc. Am. **118**, 671–678 (2005).

[11]N. Valdivia and E. G. Williams, "Study on the comparison of the methods of equivalent sources and boundary element methods for near-field acoustic holography," J. Acoust. Soc. Am. **120**, 3694–3705 (2006).

[12]J. Gomes, "Patch holography using a double layer microphone array," Proceedings of Inter-Noise 2007, in07-122.

Reconstruction of vibroacoustic responses of a highly nonspherical structure using Helmholtz equation least-squares method

Huancai Lu and Sean F. Wu
Department of Mechanical Engineering, Wayne State University, Detroit, Michigan 48202

(Received 26 February 2008; revised 11 November 2008; accepted 12 December 2008)

The vibroacoustic responses of a highly nonspherical vibrating object are reconstructed using Helmholtz equation least-squares (HELS) method. The objectives of this study are to examine the accuracy of reconstruction and the impacts of various parameters involved in reconstruction using HELS. The test object is a simply supported and baffled thin plate. The reason for selecting this object is that it represents a class of structures that cannot be exactly described by the spherical Hankel functions and spherical harmonics, which are taken as the basis functions in the HELS formulation, yet the analytic solutions to vibroacoustic responses of a baffled plate are readily available so the accuracy of reconstruction can be checked accurately. The input field acoustic pressures for reconstruction are generated by the Rayleigh integral. The reconstructed normal surface velocities are validated against the benchmark values, and the out-of-plane vibration patterns at several natural frequencies are compared with the natural modes of a simply supported plate. The impacts of various parameters such as number of measurement points, measurement distance, location of the origin of the coordinate system, microphone spacing, and ratio of measurement aperture size to the area of source surface of reconstruction on the resultant accuracy of reconstruction are examined. [DOI: 10.1121/1.3068449]

PACS number(s): 43.60.Sx, 43.60.Pt, 43.60.Jn [EGW] Pages: 1538–1548

I. INTRODUCTION

Near-field acoustical holography (NAH) is a methodology that allows for reconstruction of vibroacoustic responses on the surface of an elastic vibrating structure based on the acoustic pressures measured on a hologram surface at very close distances to a structure. Fourier acoustics[1–5] is the first generation in NAH development, which decomposes the measured acoustic pressure into the wave number domain by using two-dimensional spatial Fourier transform, projects it to a parallel plane by multiplying the angular spectrum of the measured acoustic pressure by a propagator, and converts the result to the temporal-frequency domain by taking an inverse spatial Fourier transform. This Fourier transform is applicable for an object that contains a level of constant coordinate, for example, an infinite plane and a cylinder and a sphere, and is valid in a source free region.

To extend NAH to arbitrary surfaces, an inverse boundary element method (IBEM)[6–11] based Helmholtz integral formulation is developed. This second generation of NAH technology offers users much greater flexibilities than Fourier acoustics does in the exterior and interior regions, and in selecting measurement and reconstruction locations for an arbitrary object.

An alternative to Fourier acoustics and IBEM is the Helmholtz equation least-squares (HELS) method.[12–14] Unlike its predecessors, HELS attempts to find an approximate solution to an acoustic field produced by an arbitrary object using superposition of spherical wave functions. Note that one can use different coordinate systems and different types of wave functions in HELS expansions. When the spherical coordinates are selected, one obtains an approximate solution using the spherical wave functions, and when the cylindrical coordinates are used, the approximate solution is expressed in terms of cylindrical wave functions. The coefficients in the expansion are determined by matching an assumed-form solution to the measured acoustic pressures and the errors are minimized by the least-squares method.

The main advantage of HELS is that it is immune to the nonuniqueness difficulty inherent in IBEM since HELS solves the Helmholtz equation directly. The main limitation of HELS is that there is no single set of coordinate system that can provide good approximations for all surface geometry. For example, the spherical wave functions are good for blunt and convex objects, while the cylindrical wave functions are ideal for slender bodies. If the spherical wave functions are used to approximate the acoustic quantities on a nonspherical surface, say, a thin plate discussed in this paper, the level of difficulty is significantly increased. Naturally, one may ask the question, "How accurate can HELS results be when the spherical wave functions are used to approximate an acoustic field produced by a highly nonspherical surface?" More importantly, one wants to know how various parameters such as number of measurement points, measurement distance, location of the origin of the coordinate system, microphone spacing and ratio of measurement aperture size to the reconstruction size may affect the resultant accuracy of reconstruction. These questions are addressed in this paper.

Note that the number of measurement points, measurement distance, and microphone spacing are common reconstruction parameters that need to be determined for all NAH methods. The location of the origin of the coordinate system

is such a parameter that requires special consideration when the spherical coordinates are used for nonspherical objects in HELS and IBEM based NAH. Similarly, the ratio of measurement aperture size to the reconstruction size is critical not only for HELS but also for Fourier acoustics based NAH and patch hologram approaches.[15–17]

Section II summarizes mathematical formulations of HELS method. Section III illustrates an example of reconstructing vibroacoustic responses of a simply supported baffled plate. The reconstructed normal surface velocities are validated against benchmark values, and out-of-plane vibration patterns at various natural frequencies are compared with the corresponding natural modes of this plate. The impacts of various reconstruction parameters on the resultant accuracy in reconstruction are discussed in Sec. IV. Conclusions are drawn in Sec. V.

II. MATHEMATICAL FORMULATIONS

In HELS method, the acoustic pressure is expressed as an expansion of a set of mutually orthogonal and uniformly convergent basis functions. When the spherical coordinate system is selected, the basis functions consist of the spherical Hankel functions and spherical harmonics.[12] The expansion coefficients are determined by matching the assumed-form solution to the acoustic pressures measured on a conformal surface around the source at very close distances. The errors incurred in this process are minimized by the least-squares method. In a matrix form, the HELS formulations can be written as[12]

$$\mathbf{p}(\mathbf{x};\omega) = \mathbf{G}_p(\mathbf{x}|\mathbf{x}_m;\omega)\mathbf{p}(\mathbf{x}_m;\omega) \quad \text{and}$$

$$\mathbf{v}_n(\mathbf{x};\omega) = \mathbf{G}_v(\mathbf{x}|\mathbf{x}_m;\omega)\mathbf{p}(\mathbf{x}_m;\omega), \tag{1}$$

where $\mathbf{p}(\mathbf{x};\omega)$ and $\mathbf{v}_n(\mathbf{x};\omega)$ represent, respectively, the column vectors of the acoustic pressure and normal component of particle velocity at any desired location $\mathbf{x}$, which can be in the field or on a source surface, $\mathbf{p}(\mathbf{x}_m;\omega)$ are the column vector that contains the acoustic pressure measured on a hologram surface, and $\mathbf{G}_p(\mathbf{x}|\mathbf{x}_m;\omega)$ and $\mathbf{G}_v(\mathbf{x}|\mathbf{x}_m;\omega)$ stand for the transfer matrices that correlate $\mathbf{p}(\mathbf{x}_m;\omega)$ to $\mathbf{p}(\mathbf{x};\omega)$ or to $\mathbf{v}_n(\mathbf{x};\omega)$, respectively,

$$\mathbf{G}_p(\mathbf{x}|\mathbf{x}_m;\omega) = \boldsymbol{\Psi}(\mathbf{x};\omega)\boldsymbol{\Psi}(\mathbf{x}_m;\omega)^{\dagger} \quad \text{and}$$

$$\mathbf{G}_v(\mathbf{x}|\mathbf{x}_m;\omega) = \frac{1}{i\omega\rho_0}\frac{\partial\boldsymbol{\Psi}(\mathbf{x};\omega)}{\partial n}\boldsymbol{\Psi}(\mathbf{x}_m;\omega)^{\dagger}, \tag{2}$$

where $\boldsymbol{\Psi}(\mathbf{x}_m;\omega)^{\dagger}$ is known as pseudoinverse,

$$\boldsymbol{\Psi}(\mathbf{x}_m;\omega)^{\dagger} = [\boldsymbol{\Psi}(\mathbf{x}_m;\omega)^H\boldsymbol{\Psi}(\mathbf{x}_m;\omega)]^{-1}\boldsymbol{\Psi}(\mathbf{x}_m;\omega)^H, \tag{3}$$

where the superscript H indicates a conjugate transpose and the elements of matrix $\boldsymbol{\Psi}$ are the particular solutions to the Helmholtz equation, which are expressible in the spherical coordinates as

$$\Psi_j(r,\theta,\phi;\omega) \equiv \Psi_{nl}(r,\theta,\phi;\omega) = h_n^{(1)}(kr)Y_n^l(\theta,\phi), \tag{4}$$

where $h_n^{(1)}(kr)$ and $Y_n^l(\theta,\phi)$ are the spherical Hankel functions of the first kind and the spherical harmonics, respectively, and the indices j, n, and l in Eq. (4) are related via $j=n^2+n+l+1$ with n starting from 0 to N and l from $-n$ to $+n$. For each n and l, we have $j=1-J$, where J indicates the total number of expansion functions.

For reconstructing the surface acoustic pressure using Eq. (1), J can be optimized by using a subset of measured acoustic pressures to reconstruct the full set of pressures on the measurement locations.[12,13] The minimum least-squares error between reconstructed pressures and measured pressures on the measurement locations is utilized as a criterion in iterations to search for an optimal J value in reconstruction process. Mathematically, this criterion of optimization is expressed as $\min_{J_{\text{opt}}}\|\hat{\mathbf{p}}(\mathbf{x}_m;\omega)-\mathbf{p}(\mathbf{x}_m;\omega)\|_2^2$, where $\|\cdot\|_2$ represents the L^2 norm, J_{opt} is the value of the optimized number of expansion functions, $\hat{\mathbf{p}}(\mathbf{x}_m;\omega)$ is the column vector of reconstructed pressures on the measurement locations, and $\mathbf{p}(\mathbf{x}_m;\omega)$ is the column vector of measured pressures in the hologram, respectively.[18] Detailed procedures for determining J_{opt} are described in Ref. 13 and are omitted for brevity here. For reconstructing the normal surface velocity, a different method must be used since the transfer matrix $\mathbf{G}_v(\mathbf{x}|\mathbf{x}_m;\omega)$ contains much more near-field effects than $\mathbf{G}_p(\mathbf{x}|\mathbf{x}_m;\omega)$ does. In other words, $\mathbf{G}_v(\mathbf{x}|\mathbf{x}_m;\omega)$ is much more ill-conditioned than $\mathbf{G}_p(\mathbf{x}|\mathbf{x}_m;\omega)$ is. This is precisely why reconstruction of the normal surface velocities is much more challenging than that of the surface acoustic pressures.

In this study, we use a modified Tikhonov regularization[19,20] implemented through singular value decomposition (SVD) with regularization parameters determined by the generalized cross validation (GCV).[21] The goal of Tikhonov regularization is to strike a balance between a good match to the data measured on a hologram surface and a good match to some prior knowledge of the source field. This balance is achieved by minimizing the following functional:

$$J(\mathbf{v}_n) = \|\mathbf{G}_v(\mathbf{x}|\mathbf{x}_m;\omega)^{\dagger}\mathbf{v}_n(\mathbf{x};\omega) - \mathbf{p}(\mathbf{x}_m;\omega)\|_2^2 + \alpha\|L\mathbf{v}_n\|_2^2, \tag{5}$$

where $\mathbf{v}_n(\mathbf{x};\omega)$ represents the reconstructed normal surface velocity that depends on a regularization parameter α. The first term on the right side represents a measure of closeness of the reconstructed and benchmark values, while the second term is a penalty term for the "roughness" of the reconstructed results. L is the matrix that depends on the type of regularization. Here we choose L to be an identity matrix, and Eq. (5) becomes the standard Tikhonov regularization.[22–24] The choice of a regularization parameter α allows users to decide how far to go along this path to strike the best balance. For example, a good match to measured data is achieved when $\alpha=0$, whereas a good match to some prior knowledge of a source field is achieved when $\alpha=\infty$.

Using SVD and applying regularization to Eq. (1) yield the following factorization:

$$\mathbf{p}(\mathbf{x};\omega) = V_p F_p^{\alpha}\Sigma_p^{-1}U_p^H\mathbf{p}(\mathbf{x}_m;\omega) \quad \text{and}$$

$$\mathbf{v}_n(\mathbf{x};\omega) = V_v F_v^{\alpha} \Sigma_v^{-1} U_v^H \mathbf{p}(\mathbf{x}_m;\omega), \tag{6}$$

where the reconstructed acoustic pressure $\mathbf{p}(\mathbf{x};\omega)$ and normal velocity $\mathbf{v}_n(\mathbf{x};\omega)$ on a source surface are α dependent, the subscripts p and v indicate reconstructions for $\mathbf{p}(\mathbf{x};\omega)$ and $\mathbf{v}_n(\mathbf{x};\omega)$, respectively; Σ_p^{-1} and Σ_v^{-1} represent inversions of the diagonal matrices containing singular values of the transfer matrices $\mathbf{G}_p(\mathbf{x}|\mathbf{x}_m;\omega)$ and $\mathbf{G}_v(\mathbf{x}|\mathbf{x}_m;\omega)$ defined in Eq. (2), respectively; V_p and V_v denote $N \times N$ unitary matrices, where N is the total number of reconstruction points; U_p^H and U_v^H denote $M \times M$ unitary matrices that are complex conjugate of U_p and U_v, respectively, where M is the total number of measurement points; and F_p^{α} and F_v^{α} signify the low-pass filters for eliminating the high wave numbers (evanescent waves) in reconstructing $\mathbf{p}(\mathbf{x};\omega)$ and $\mathbf{v}_n(\mathbf{x};\omega)$, respectively. Note that because the singular values in Σ_p^{-1} and Σ_v^{-1} are different, the regularization parameters will be different. As a result, the low-pass filters F_p^{α} and F_v^{α} will be different, which are given by[22]

$$F_p^{\alpha} = \mathrm{diag}\left(\ldots, \frac{\varepsilon_{pj}^2}{\alpha^3/(\alpha+\varepsilon_{pj}^2)^2 + \varepsilon_{pj}^2}, \ldots\right) \quad \text{and}$$

$$F_v^{\alpha} = \mathrm{diag}\left(\cdots, \frac{\varepsilon_{vj}^2}{\alpha^3/(\alpha+\varepsilon_{vj}^2)^2 + \varepsilon_{vj}^2}, \cdots\right), \tag{7}$$

where ε_{pj} and ε_{vj} are the jth singular values of the transfer matrices $\mathbf{G}_p(\mathbf{x}|\mathbf{x}_m;\omega)$ and $\mathbf{G}_v(\mathbf{x}|\mathbf{x}_m;\omega)$, respectively. Once the surface acoustic pressure and normal surface velocity are determined, the normal acoustic intensity on the source surface is given by $\mathbf{I}_n(\mathbf{x};\omega) = 1/2\ \mathrm{Re}[\mathbf{p}(\mathbf{x};\omega)\mathbf{v}_n^*(\mathbf{x};\omega)]$.

III. VALIDATIONS OF HELS FORMULATIONS

A. Test object

In this study, a simply supported plate in an infinite baffle is selected as a test object. The dimensions of this plate are 0.5 m wide, 0.5 m long, and 0.001 m thick. The Cartesian coordinates are oriented in such a way that the y-axis is in the normal direction of plate, and x and z axes are in transverse and longitudinal directions, respectively (see Fig. 1). Once again, the main reasons for selecting this plate are that it is highly nonspherical, yet the analytic solutions to structural vibrations and sound radiation are readily available.

The out-of-plane vibration responses of a simply supported thin plate are given by[25,26]

$$w(x,y_0,z) = \sum_{\mu=1}^{\infty}\sum_{\nu=1}^{\infty} A_{\mu\nu} \sin\frac{\mu\pi(x+L_x/2)}{L_x} \sin\frac{\nu\pi(z+L_z/2)}{L_z},$$

$$-\frac{L_x}{2} \le x \le \frac{L_x}{2}, \quad -\frac{L_z}{2} \le z \le \frac{L_z}{2}, \tag{8}$$

$$v_n(x,y_0,z) = -i\omega w(x,y_0,z), \tag{9}$$

where $w(x,y_0,z)$ and $v_n(x,y_0,z)$ represent the complex amplitudes of the flexural vibration displacement and normal surface velocity of plate, respectively; $A_{\mu\nu}$ is the amplitude coefficient of the μth and νth normal modes determined from the initial conditions of the problem,[26] μ and υ are the indices indicating the order of the natural mode in the x and z axes, respectively; L_x and L_z are the length and width of the plate, respectively; and y_0 is the position of the plate in the coordinate system.

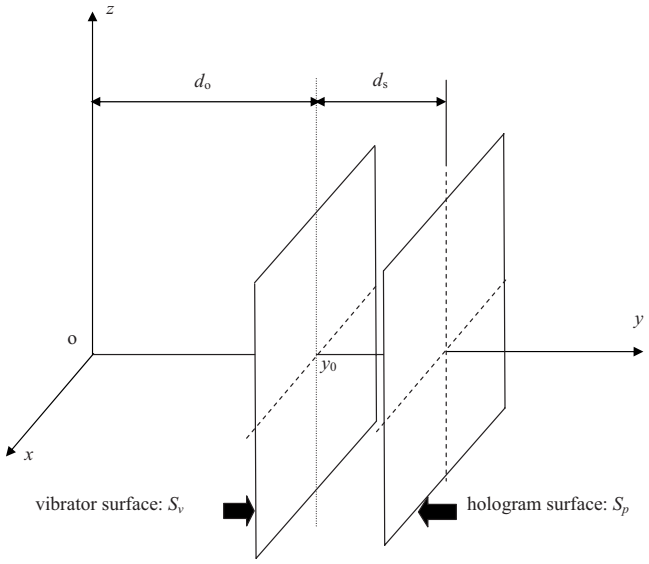


FIG. 1. Schematic of the test setup, where d_s is the standoff distance from hologram surface S_p to plate surface S_v, and d_0 is the distance from origin of coordinate system to plate surface.

Given the normal surface velocity of this plate, $v_n(x,y_0,z)$, we can then generate the field acoustic pressure by using the Rayleigh integral:[27]

$$p(x',y',z';\omega) = \frac{-i\omega\rho_0}{2\pi} \int_{-L_x/2}^{L_x/2} \int_{-L_z/2}^{L_z/2} v_n(x,y_0,z) \frac{e^{ikR}}{R} dx\, dz, \tag{10}$$

where ω is the angular frequency, ρ_0 is the ambient density of the surrounding fluid medium, and $R = \sqrt{(x-x')^2+(y-y_0)^2+(z-z')^2}$ is the distance between field and surface points.

Note that in analyzing structural vibrations attention is typically paid to certain dominant resonant (normal) modes of a structure. It is particularly important when the plate is driven near a resonance frequency corresponding to one of the modes, since the vibration is then dominated by a single normal mode. Therefore, we choose to validate the reconstructed vibroacoustic quantities at the normal modes of this baffled plate. Also in practice the input data will inevitably contain errors due to the presence of background noise. So in this case, the input data generated by using Eq. (10) are added with additional random noise to represent a 30 dB signal to noise ratio, as suggested by a reviewer, where the signal level is the rms pressure of the acoustic pressure on a hologram surface. These field acoustic pressures with random noise are then taken as input to the HELS formulations (6) to reconstruct surface acoustic pressure and normal surface velocity. The reconstructed normal surface velocities are compared to the benchmark values given by Eq. (9), and the errors in the reconstruction are obtained using the normalized least-squares error $\mathfrak{I}_{\mathrm{NLSE}}$:[28]

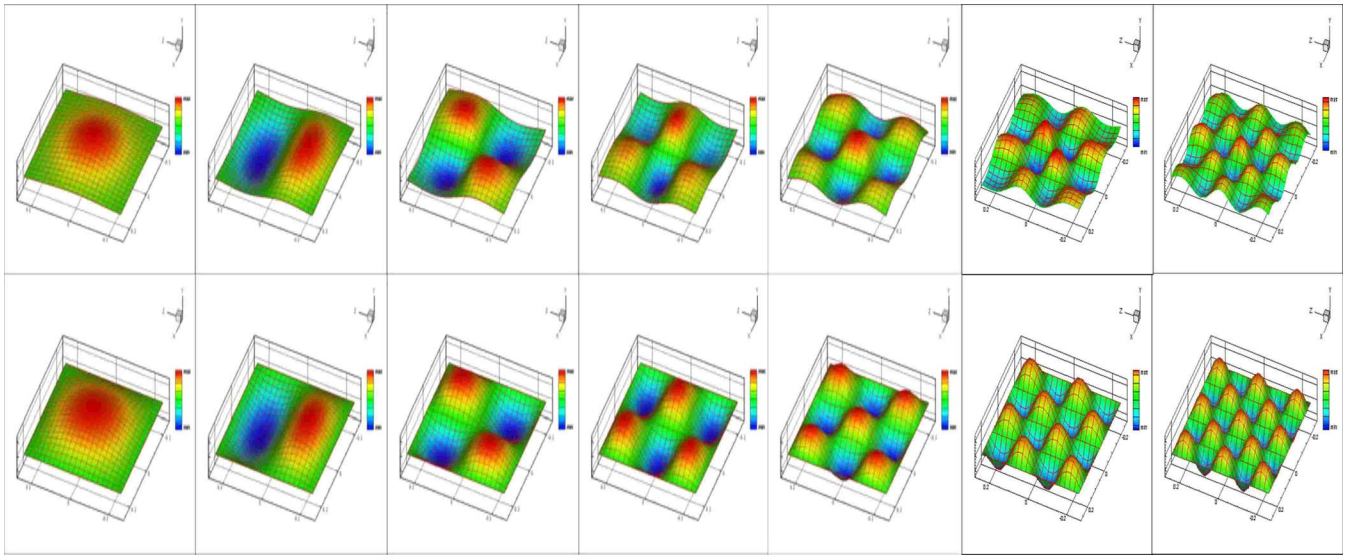


FIG. 2. (Color online) Comparison of the reconstructed normal surface velocities (top row) and theoretical ones (bottom row) for the (1, 1), (1, 2), (2, 2), (2, 3), (3, 3), (4, 4), and (5, 5) natural modes with a standoff distance $d_s=0.015$ m, normalized origin location $d_0/D=0.919$, ratio of the measurement aperture to plate surface $\gamma=1.25$, number of measurement points $M=441$ with uniform microphone spacing $\Delta d_m=0.028$ m, and reconstruction points $N=441$ with a uniform spatial interval $\Delta d_m=0.025$ m.

$$\mathfrak{I}_{\mathrm{NLSE}}=\frac{\sum_i^N \|\hat{v}_{n,i}(x,y_0,z;\omega)-v_{n,i}(x,y_0,z;\omega)\|_2^2}{\sum_i^N \|v_{n,i}(x,y_0,z;\omega)\|_2^2}\times 100\%. \tag{11}$$

Note that in all cases, the number of reconstruction points is fixed at $N=441$ and the spacing among individual reconstruction points is held constant at 0.025 m over the plate surface.

B. Discussions of reconstructed vibroacoustic responses

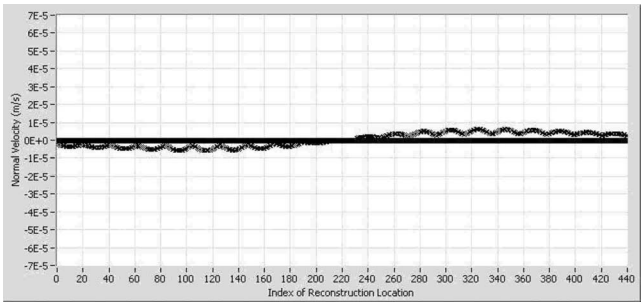
To obtain an optimal reconstruction of the vibroacoustic responses on the surface of plate, we set the origin of the Cartesian coordinate system outside the plate while reconstructing the vibroacoustic quantities on the opposite side of the plate. Section IV discusses the impacts of selecting the origin of the Cartesian coordinate system and other parameters on the accuracy in reconstruction. Here, we just designate the origin at the optimal distance, $d_0=0.325$ m, behind the plate front surface, and take the same number of measurement points as that of reconstruction points $M=N=441$. The spacing among individual measurement points ($\Delta d_m=0.028$ m), however, is different from that of reconstruction points ($\Delta d_r=0.025$ m). Thus the measurement points are not in line with the reconstruction points, and the measurement aperture, $S_p=(0.028\times 20)(0.028\times 20)=0.3136$ m^2, is slightly larger than plate surface, $S_v=0.5\times 0.5=0.25$ m^2. The ratio of the measurement aperture to the source surface, $\gamma=S_p/S_v\approx 1.25$, is much less than the 4:1 ratio required by Fourier acoustics based NAH.[2,6] This should not be a surprise because by virtue of formulations (1)–(6), HELS method has the inherent capability of performing patch NAH. As such, the measurement area only needs to be slightly larger than the area of interest. This makes it very convenient to reconstruct acoustic quantities using HELS in practice.[15]

Figure 2 shows comparisons of three-dimensional graphs of the reconstructed normal surface velocity patterns and analytic results for seven natural modes at (1, 1), (2, 1), (2, 2), (3, 2), (3, 3), (4, 4), and (5, 5), respectively. Good agreement is obtained, specifically for the lower-order modes. Table I summarizes the natural frequencies and structural wavelengths associated with the natural modes.

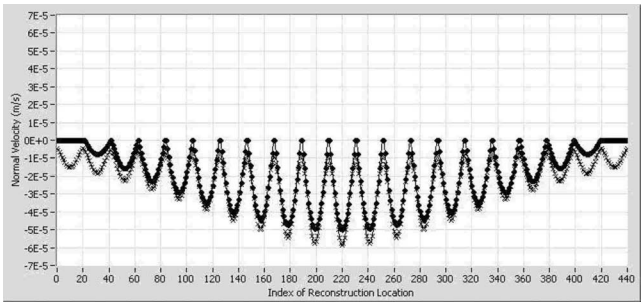
To examine the accuracy in reconstruction, we plot the real and imaginary components of reconstructed and analytic normal surface velocities vs the reconstruction locations in Figs. 3–9. Results show that the reconstructed velocities (lines with symbol ×) for all seven natural modes agree very well with the benchmark values (lines with symbol ●) in general. The largest errors occur near the edges of plate for all cases because (1) the filtering process in regularization removes the high wave number information that accounts for the slope discontinuity at the edges, and (2) HELS formulations use the spherical wave functions to approximate the

TABLE I. Mode order, natural frequency, and structural wavelength of a simply supported thin plate with a dimension of 0.5×0.5 m^2.

Mode (μ,υ)	Frequency f (Hz)	Structural wavelengths λ_s (m)	
		$\lambda_{s,x}$	$\lambda_{s,z}$
(1, 1)	19.1	1	1
(2, 1)	47.8	0.5	1
(2, 2)	76.4	0.5	0.5
(3, 2)	124.2	0.33	0.5
(3, 3)	172	0.33	0.33
(4, 4)	305.7	0.25	0.25
(5, 5)	477.6	0.2	0.2



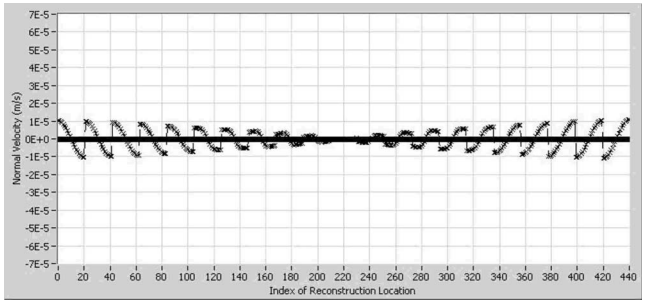
(a) REAL [V_n]



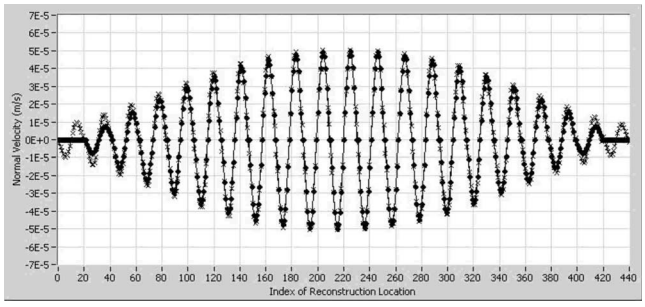
(b) IMAG [V_n]

FIG. 3. Real (a) and imaginary (b) components of reconstructed normal surface velocities (×) vs theoretical ones (●) for the (1, 1) natural mode.

acoustic radiation from a plate, which are sufficient around the center but insufficient along the edges because they are furthest from a hypothetical spherical surface. Results also show that reconstruction errors increase with the frequency at a fixed point, which is common for all expansion theories.

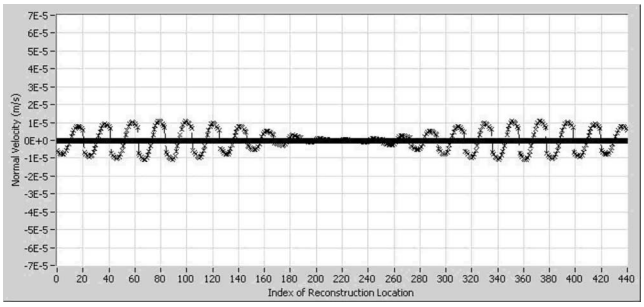


(a) REAL [V_n]

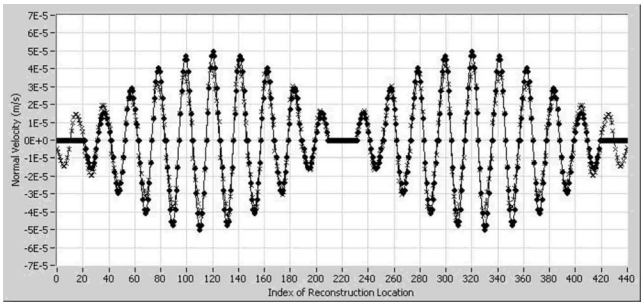


(b) IMAG [V_n]

FIG. 4. Real (a) and imaginary (b) components of reconstructed normal surface velocities (×) vs theoretical ones (●) for the (1, 2) natural mode.



(a) REAL [V_n]

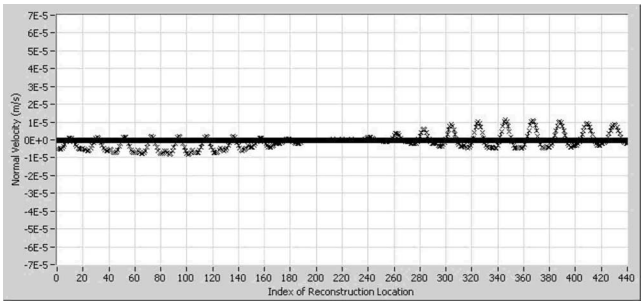


(b) IMAG [V_n]

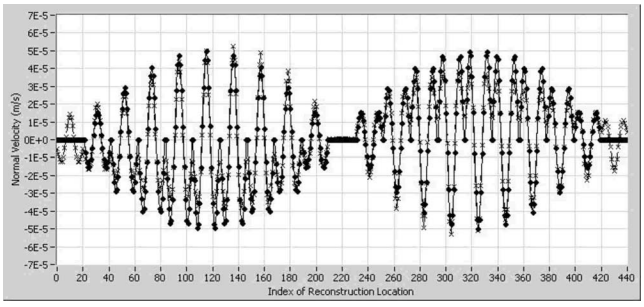
FIG. 5. Real (a) and imaginary (b) components of reconstructed normal surface velocities (×) vs theoretical ones (●) for the (2, 2) natural mode.

IV. EFFECTS OF VARIOUS RECONSTRUCTION PARAMETERS

To address the effects of parameters, for example, the location of the origin of the coordinate system, measurement distance, number of measurement points, microphone spacing, and ratio of measurement aperture to reconstruction surface area on the accuracy in reconstruction, we conduct sys-

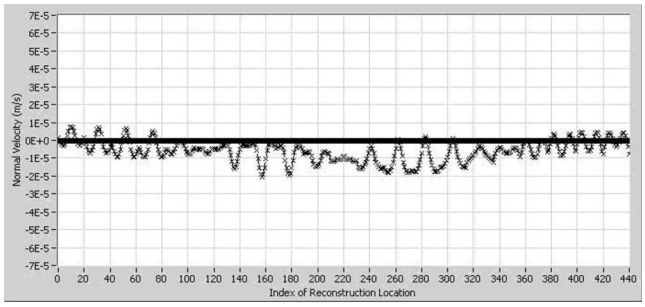


(a) REAL [V_n]

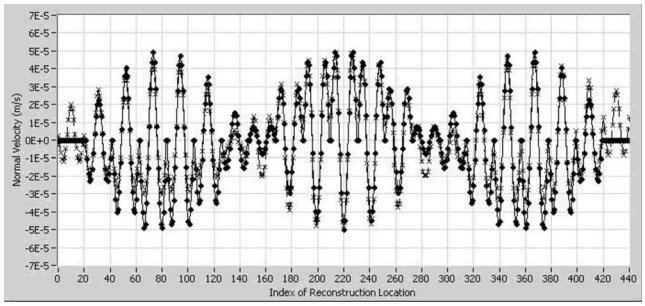


(b) IMAG [V_n]

FIG. 6. Real (a) and imaginary (b) components of reconstructed normal surface velocities (×) vs theoretical ones (●) for the (2, 3) natural mode.



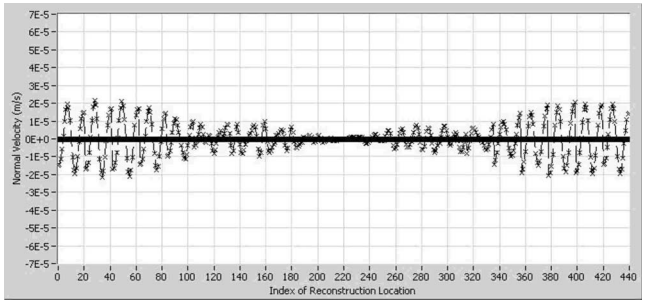
(a) REAL [V_n]



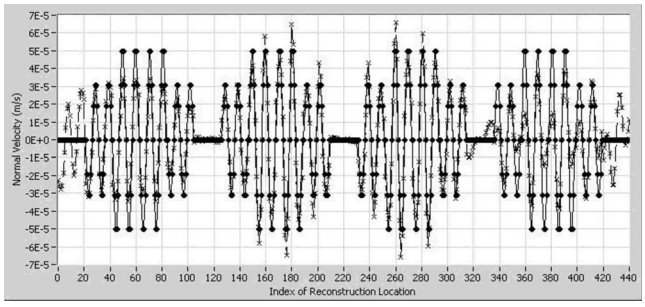
(b) IMAG [V_n]

FIG. 7. Real (a) and imaginary (b) components of reconstructed normal surface velocities (×) vs theoretical ones (●) for the (3, 3) natural mode.

tematic numerical experiments in which normal surface velocities are reconstructed and the normalized least-squares errors $\Im_{\mathrm{NLSE}}$ are calculated with only one parameter varying while the rest being held constant. The results of these experiments are expected to shed light on how to select these parameters to produce best results and provide guidelines for reconstructing vibroacoustic responses of a vibrating structure.

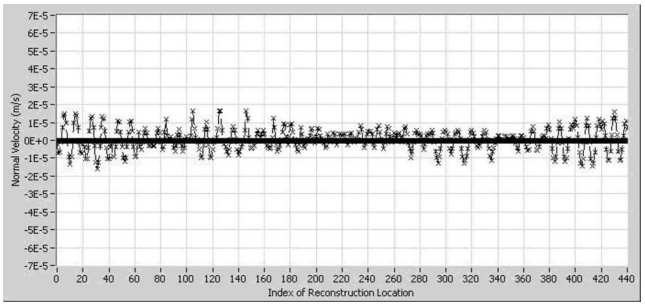


(a) REAL [V_n]

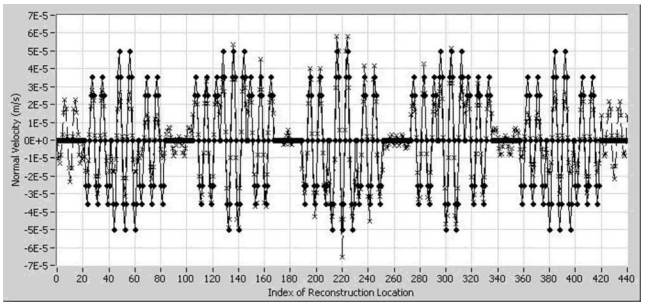


(b) IMAG [V_n]

FIG. 8. Real (a) and imaginary (b) components of reconstructed normal surface velocities (×) vs theoretical ones (●) for the (4, 4) natural mode.



(a) REAL [V_n]



(b) IMAG [V_n]

FIG. 9. Real (a) and imaginary (b) components of reconstructed normal surface velocities (×) vs theoretical ones (●) for the (5, 5) natural mode.

A. Location of the origin of the Cartesian coordinate system

In dealing with a thin plate, it is not feasible to place the origin of a coordinate system at the geometric center because the thickness of the plate is too small. Under this circumstance, it is better to move the origin outside the plate and place it in the opposite side of the surface where the vibroacoustic quantities are to be reconstructed. For example, to reconstruct normal velocity on the front surface of the plate, $y=0.5h^+$, where h is the thickness of a plate, the origin should be moved to $y<0$ region. The location of the origin directly affects the accuracy in reconstruction. Summarized here is the effect of the location of the origin on reconstruction by varying d_0 with other parameters held constant.

Figure 10 depicts the impact of varying the origin location of Cartesian coordinate system on the accuracy in reconstruction. Results show that when the normalized distance of the origin is around $d_0/D \approx 0.9$, where $D=0.5\sqrt{L_x^2+L_z^2}$ is the characteristic dimension of a plate, the errors in the reconstructed normal surface velocities are minimum, $\Im_{\mathrm{NLSE}} \leq 30\%$ for all natural modes tested. Note that the range where errors are minimal becomes narrower for a higher-order mode than for a lower-order one. This is consistent with the observation that the higher the frequency is, the more sensitive it becomes in selecting the origin location.

Results show that there is an optimal distance for the origin around $d_0/D \approx 0.9$. Setting the origin of the Cartesian coordinate system too close to or too far from a thin plate surface tends to increase the reconstruction error at any frequency. In particular, the errors in the reconstructed normal velocities deteriorate more rapidly at higher frequencies than those in lower frequencies.

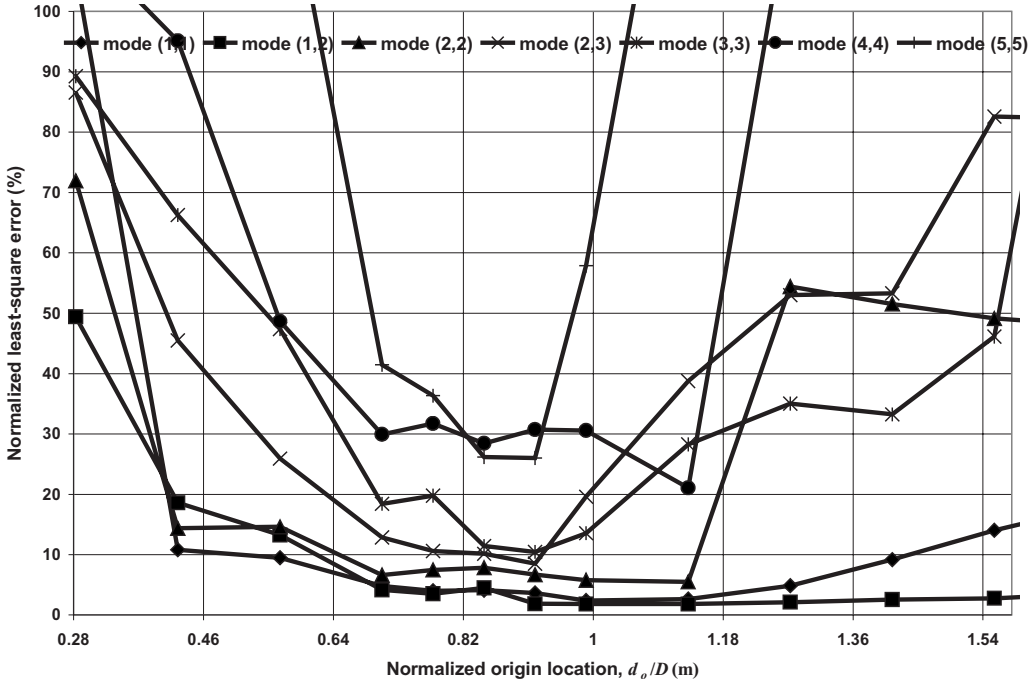


FIG. 10. Normalized least-squares errors in reconstructing the normal surface velocities with respect to normalized origin location of the Cartesian coordinate system. In this case, number of measurement points $M=441$; number of reconstruction points $N=441$; spacing among individual reconstruction points $\Delta d_r=0.025$ m; spacing among microphones $\Delta d_m=0.028$ m; standoff distance between hologram surface and plate surface is $d_s=0.015$ m; the ratio of measurement aperture to plate surface $\gamma=1.25$.

B. Measurement distance

A basic requirement in NAH is to capture as much near-field information as possible. So we should always place a hologram surface as close to a source surface as possible. In practice, however, this may not be possible because of the test environment and test condition. Therefore, it is important to know how close is close enough. This issue is addressed in this section.

Figure 11 illustrates how the measurement distance affects the reconstruction errors when all other parameters are

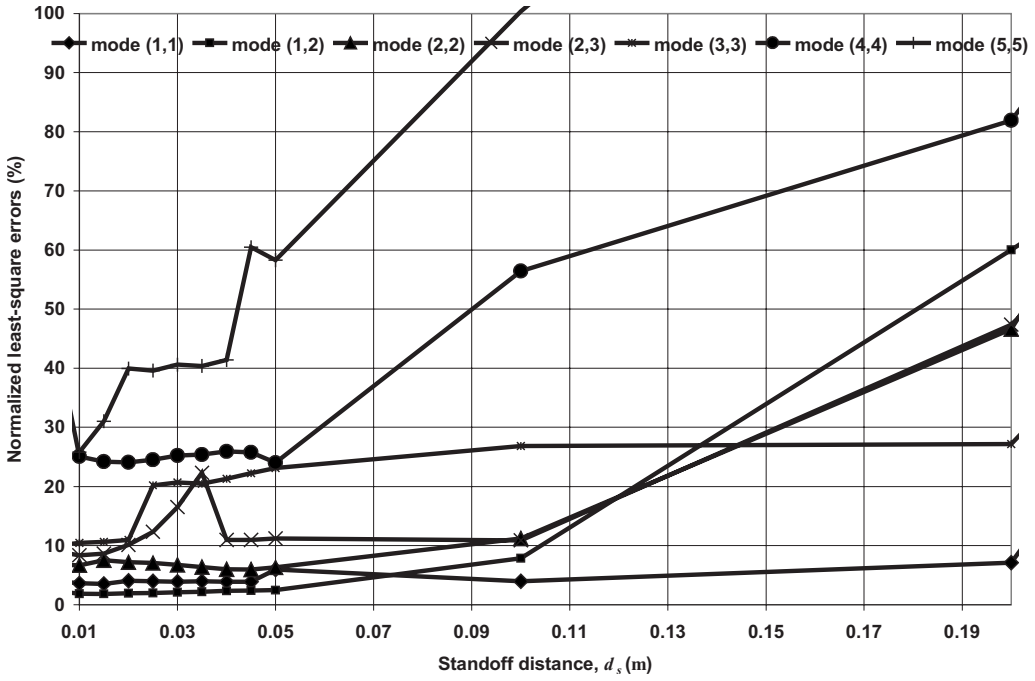


FIG. 11. Normalized least-squares errors in reconstructing the normal surface velocities with respect to standoff distance. In this case, number of measurement points $M=441$; number of reconstruction points $N=441$; spacing among individual reconstruction points $\Delta d_r=0.025$ m; spacing among microphones $\Delta d_m=0.028$ m; normalized origin locations $d_0/D=0.919$; ratio of measurement aperture to plate surface $\gamma=1.25$.

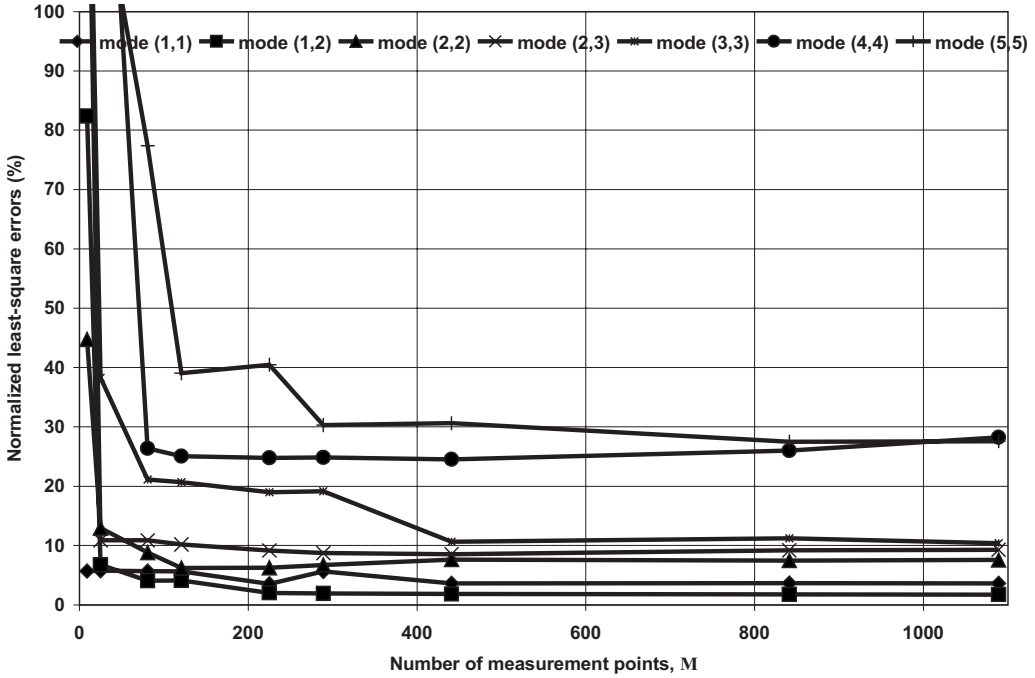


FIG. 12. Normalized least-squares errors in reconstructing the normal surface velocities with respect to number of measurement points. In this case, number of reconstruction points $N=441$; spacing among individual reconstruction points $\Delta d_r=0.025$ m; normalized origin locations $d_0/D=0.919$; standoff distance between hologram surface and plate surface is $d_s=0.015$ m; ratio of measurement aperture to plate surface $\gamma=1.25$. The spacing between measurement points varies when the number of measurement points changes because the measurement aperture holds constant.

held constant. As expected, the closer the hologram surface to the source surface is, the more evanescent waves are captured, and the smaller the reconstruction errors are. This is true for all frequencies. However, the impact of measurement distances on reconstruction errors at lower-order modes is different from those at higher-order modes.

For example, the (4, 4) mode reconstruction error can be held under $\mathfrak{I}_{\mathrm{NLSE}} \leq 30\%$ when standoff distance is less than one-quarter of the structural wavelength of the (4, 4) mode $\lambda_{4,4}$; here $\lambda_{4,4}=0.25$ m (see Table I). On the other hand, the (5, 5) or higher-order modes, the measurement distance and microphone spacing should be reduced simultaneously in order to obtain the same level of accuracy as that of lower-order modes. Note that the standoff distance should be aimed at the wavelength of the highest mode of interest not the wavelength of an acoustic wave.

The results displayed in Fig. 11 show how difficult it is to reconstruct accurately the normal surface velocity distributions of high-order modes. This is because the high-order modes contain much more near-field effects than the low-order modes do. Those near-field effects decay rapidly with distances. As a result, computational errors or measurement and background noise could be taken as input data in reconstruction, which will significantly distort the reconstructed image as acoustic quantities are projected back toward a source surface. Therefore, the measurement surface must be placed at very close range to a target surface in order to capture the near-field effects. Sometimes, this may not be possible in practice either because of hostile measurement environment very close to a vibrating structure or because of the presence of many accessories around a target structure. Therefore, there will always be a limit on the accuracy and spatial resolution in the reconstructed normal surface velocity in practice.

C. Microphone spacing

When there is no *a priori* information of the locations and characteristics of sound sources, a uniform grid of microphone array is safe to use because it captures near-field effects uniformly. The spacing among individual microphones is very important. Typically, the smaller the microphone spacing is, the higher the natural modes of a target structure can be resolved and therefore, the higher the spatial resolution in reconstruction may be. However, smaller microphone spacing means a larger number of measurement points when the measurement aperture remains constant. More numbers of measurement points are required in reconstruction of the natural modes with higher orders because the more basis functions are needed in the assumed solutions of reconstruction in Eqs. (1)–(4) to describe the targeted source. However, in many practical situations it is not likely that very high-order modes of a structure could be excited. Hence there is no need to reduce the microphone spacing very small and to increase the number of measurement points too large for it only increases overall costs of the reconstruction process.

Figure 12 demonstrates that reconstruction errors reduce as microphone spacing becomes smaller. Note that in this case the decreasing in spacing is obtained by increasing the number of microphones so the measurement aperture S_p and all other reconstruction parameters are constant. Results show that when $\Delta d_m \geq 0.05$ m or $\Delta d_m \geq (1/4)\lambda_s$, where λ_s is

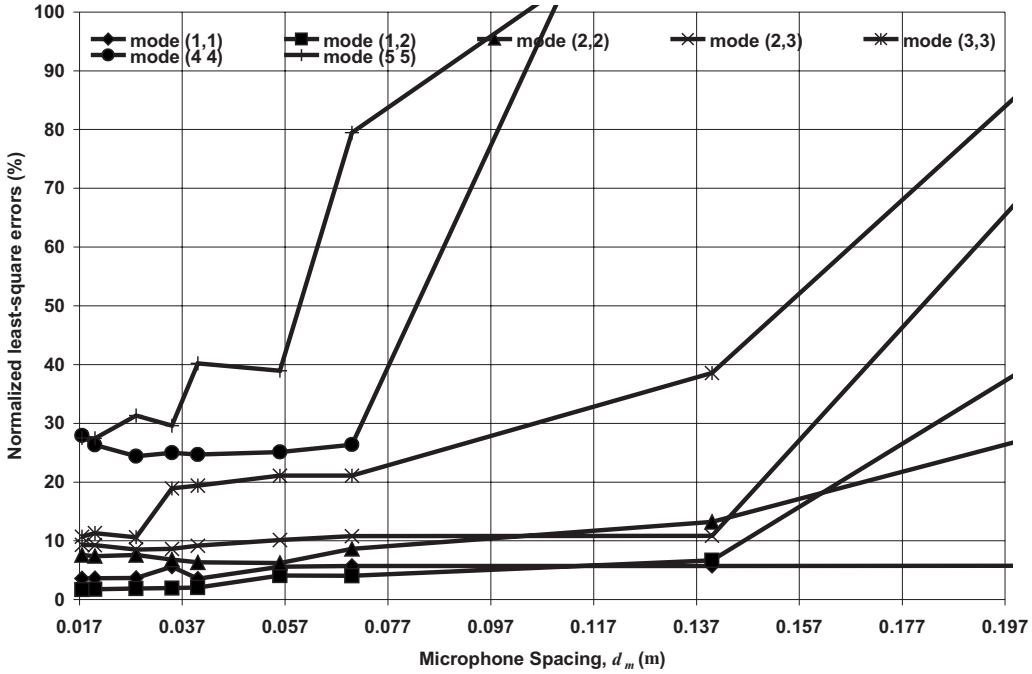


FIG. 13. Normalized least-squares errors in reconstructing the normal surface velocities with respect to microphone spacing. In this case, number of reconstruction points $N=441$; spacing among individual reconstruction points $\Delta d_r=0.025$ m; normalized origin locations $d_0/D=0.919$; standoff distance between hologram surface and plate surface is $d_s=0.015$ m; ratio of the measurement aperture to plate surface $\gamma=1.25$. The number of measurement points varies when the spacing changes because the measurement aperture holds constant.

the structural wavelength of the highest mode of interest, which corresponds to a number of measurement points $M<121$ in Fig. 12, the reconstruction errors exceed $\mathfrak{I}_{\mathrm{NLSE}} \geq 40\%$. When microphone spacing is decreased to $\Delta d_m \leq 0.03$ m or $\Delta d_m \leq (1/7)\lambda_s$, which corresponds to more measurement points $M>289$, the reconstruction errors becomes $\mathfrak{I}_{\mathrm{NLSE}} \leq 30\%$. Note that the reconstruction errors tend to approach a constant level as Δd_m further decreases and number of measurements further increases. In other words, improvement in spatial resolution and accuracy in reconstruction is negligible, even the spacing among measurement microphones is further reduced and number of measurements further increased.

For instance, the error levels off around 25% when the number of microphones increases from 400 to 800. Figure 13 depicts the relationship between the accuracy in reconstruction and microphone spacing. For example, $\mathfrak{I}_{\mathrm{NLSE}} \leq 40\%$ if spacing $\Delta d_m \leq 0.05$ m$=(1/4)\lambda_s$ for the cases considered.

In practice, very small microphone spacing may not be realistic because of test condition and environment, as well as costs considerations. As a compromise, we recommend that the microphone spacing be no less than $\Delta d_m \leq (1/4)\lambda_s$, where λ_s is the structural wavelength of the highest mode of interest.

D. Ratio of measurement aperture to reconstruction surface area

In general, the larger the ratio of measurement aperture to reconstruction surface area γ is, the more information of the spherical spreading waves can be collected, and the higher the accuracy in reconstruction may be. However, this also means higher costs for reconstruction due to a larger number of measurement points to be taken, more basis functions to be included in reconstruction when microphone spacing of is fixed. In that case, mathematically speaking, the high-order functions which generally decay in an exponential-like manner greatly amplify the measurement and background noise on the reconstruction results on source surface.[13,29] Thus, one is faced with the situation where too small γ that implies too few basis functions is not enough to describe the sound field and too large γ that implies too many basis functions is even more problematic. This phenomenon is typically shown when the frequency is high like mode (5, 5) in Fig. 14. In this study, we want to find an optimal ratio γ that will allow one to get satisfactory reconstruction results in the most cost-effective manner.

To study the impact of γ on the reconstruction accuracy, we fix microphone spacing and standoff distance. So an increase in γ indicates an increase in the number of measurement points and measurement aperture.

Figure 14 depicts how the impacts of changing γ on the reconstruction accuracy. Here, we set constant microphone spacing at $\Delta d_m=0.02$ m, and then change the number of measurement points that in true change the measurement aperture size. Results indicate that for lower-order modes the reconstruction errors decrease with an increase in γ. However, for the (5, 5) mode the reconstruction errors decrease at first and then increase with the value of γ. Moreover, the range of this minimal reconstruction error seems to be getting narrower as the frequency becomes higher. This is because there is a limit on the degree of accuracy in using the spherical wave functions to approximate an acoustic field generated by a planar surface. It will be difficult for the spherical wave functions to cover a large plane. That is why

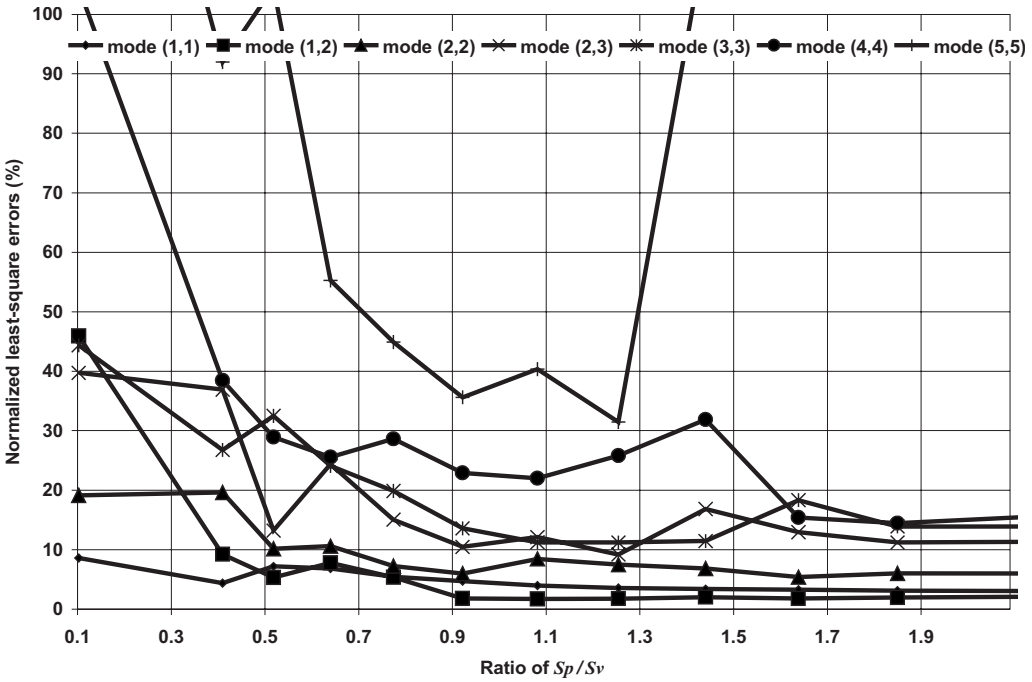


FIG. 14. Normalized least-squares errors in reconstructing the normal surface velocities with respect to ratio of measurement aperture to area of reconstruction surface. In this case, number of reconstruction points $N=441$; spacing among individual reconstruction points $\Delta d_r=0.025$ m; normalized origin locations $d_0/D=0.919$; standoff distance between hologram surface and plate surface is $d_s=0.015$ m; spacing among microphones $\Delta d_m=0.02$ m.

when the measurement aperture exceeds a certain limit, the accuracy in reconstruction actually goes down because there is no way for HELS to get a good fit for the entire area. On the other hand, if the measurement aperture is too small, not enough information is captured, so the errors in reconstruction are large. Therefore, there is an optimal ratio of aperture size to reconstruction size. This explains why the errors for the (5, 5) mode are large when the value of γ is either small or large, and there is an optimal range of γ that the accuracy in reconstruction for the (5, 5) mode is the highest.

Test results in Fig. 14 show that for frequencies up to the natural frequency of the (5, 5) mode, the maximum errors in reconstruction are $\Im_{\mathrm{NLSE}} \leq 40\%$, when $0.9 < \gamma < 1.3$. When the measurement aperture is less than 90% of the source surface, the reconstruction errors increase because some vital information is lost in data collection. On the other hand, if the ratio is greater than 130%, reconstruction errors may increase, especially for the higher-order modes. A good compromise for engineering applications is to set one row and one column larger in a measurement aperture than a target source surface.

V. CONCLUSIONS

Vibroacoustic responses of a highly nonspherical object such as a simply supported baffled plate are reconstructed using HELS method based on the expansion of the spherical wave functions. The input data are generated by the Rayleigh integral and benchmark values of normal surface velocity distributions at seven natural modes of this simply supported plate are obtained by analytic formulations. This enables us to examine the reconstructed normal surface velocity distributions rigorously. Most importantly, the impacts of various reconstruction parameters on reconstruction accuracy are investigated systematically. Results demonstrate that HELS method can be used to reconstruct the vibroacoustic responses of a highly nonspherical object provided that certain parameters are properly selected. For a plate type structure the optimal location of the Cartesian coordinate system can be set approximately at $d_0/D \approx 0.9$, where $D=0.5\sqrt{L_x^2+L_z^2}$ is the characteristic dimension of a plate; the measurement distance can be set at $d_s \leq (1/8)\lambda_s$, where λ_s is the structural wavelength of the highest mode of interest; the microphone spacing can be set at $\Delta d_m \leq (1/4)\lambda_s$, and the measurement aperture can be set at one row and one column larger than the source surface. Under this setting, the reconstruction errors in the normal surface velocity can be held under 30%.

[1]E. G. Williams and J. D. Maynard, "Holographic imaging without the wavelength resolution limit," Phys. Rev. Lett. **45**, 554–557 (1980).
[2]J. D. Maynard, E. G. Williams, and Y. Lee, "Nearfield acoustic holography: I. Theory of generalized holography and the development of NAH," J. Acoust. Soc. Am. **78**, 1395–1413 (1985).
[3]E. G. Williams, J. D. Maynard, and E. Skudrzyk, "Sound source reconstructions using a microphone array," J. Acoust. Soc. Am. **68**, 340–344 (1980).
[4]E. G. Williams and J. D. Maynard, "Numerical evaluation of the Rayleigh integral for planar radiators using the FFT," J. Acoust. Soc. Am. **72**, 2020–2030 (1982).
[5]E. G. Williams, H. D. Dardy, and R. G. Fink, "Nearfield acoustical holography using an underwater automated scanner," J. Acoust. Soc. Am. **78**, 789–798 (1985).
[6]E. G. Williams, *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography* (Academic, San Diego, 1999), Chap. 3.
[7]B. K. Gardner and R. J. Bernhard, "A noise source identification technique using an inverse Helmholtz integral equation method," ASME J. Vib., Acoust., Stress, Reliab. Des. **110**, 84–90 (1988).
[8]W. A. Veronesi and J. D. Maynard, "Digital holographic reconstruction of sources with arbitrarily shaped surfaces," J. Acoust. Soc. Am. **85**, 588–598 (1989).
[9]B.-K. Kim and J.-G. Ih, "On the reconstruction of vibro-acoustic field over the surface enclosing an interior space using the boundary element

method," J. Acoust. Soc. Am. **100**, 3003–3016 (1996).
[10]B.-K. Kim and J.-G. Ih, "Design of an optimal wave-vector filter for enhancing the resolution of reconstructed source field by near-field acoustical holography (NAH)," J. Acoust. Soc. Am. **107**, 3289–3297 (2000).
[11]N. Valdivia and E. G. Williams, "Implicit methods of solution to integral formulations in boundary element method based nearfield acoustic holography," J. Acoust. Soc. Am. **116**, 1559–1573 (2004).
[12]Z. Wang and S. F. Wu, "Helmholtz equation-least-squares method for reconstructing the acoustic pressure field," J. Acoust. Soc. Am. **102**, 2020–2032 (1997).
[13]S. F. Wu, "On reconstruction of acoustic pressure fields using the Helmholtz equation least squares method," J. Acoust. Soc. Am. **107**, 2511–2522 (2000).
[14]N. E. Rayess and S. F. Wu, "Experimental validations of the HELS method for reconstructing acoustic radiation from a complex vibrating structure," J. Acoust. Soc. Am. **107**, 2955–2964 (2000).
[15]E. G. Williams, "Fourier acoustics: Uncovering the origins of sound," 11th International Congress on Sound and Vibration, St. Petersburg, Russia, 5–8 July 2004.
[16]E. G. Williams, "Continuation of acoustic near-fields," J. Acoust. Soc. Am. **113**, 1273–1281 (2003).
[17]A. Sarkissian, C. F. Gaumond, E. G. Williams, and B. H. Houston, "Reconstruction of the acoustic field over a limited surface area on a vibrating cylinder," J. Acoust. Soc. Am. **93**, 48–54 (1993).
[18]H. Lu, "Reconstruction of vibroacoustic responses using Helmholtz equation least squares method," Ph.D. thesis, Wayne State University, Detroit, MI (2007).
[19]A. N. Tikhonov, "Solution of incorrectly formulated problems and the regularization method," Sov. Math. Dokl. **4**, 1035–1038 (1963).
[20]D. M. Titterington, "General structure of regularization procedures in image restoration," Astron. Astrophys. **144**, 381–387 (1985).
[21]G. H. Golub, M. Heath, and G. Wahba, "Generalized cross-validation as a method for choosing a good ridge parameter," Technometrics **21**, 215–223 (1979).
[22]E. G. Williams, "Regularization methods for nearfield acoustical holography," J. Acoust. Soc. Am. **110**, 1976–1988 (2001).
[23]S. F. Wu, "Hybrid near-field acoustic holography," J. Acoust. Soc. Am. **115**, 207–217 (2004).
[24]P. C. Hansen, "Rank-deficient and discrete ill-posed problems," (SIAM, Philadelphia, PA, 1998).
[25]A. W. Leissa, *Vibration of Plates* (Scientific Information Division, NASA, Washington, DC, 1969), Chap. 10, pp. 267–284.
[26]G. B. Warburton, "The vibration of rectangular plates," Proc. Inst. Mech. Eng. **168**, 371–384 (1954).
[27]A. D. Pierce, *Acoustics: An Introduction to Its Physical Principles and Applications* (McGraw-Hill, New York, 1981), Chap. 4, pp. 165–194.
[28]K. W. Pratt, *Digital Image Processing*, 2nd ed. (Wiley, New York, 1991), pp. 369–370 and 685–686.
[29]N. E. Rayess, "An investigation in acoustic holography using the Helmholtz equation least-squares method, Ph.D. thesis, Wayne State University, Detroit, MI (2001).

Using time-reversal to generate generalized transversely localized transient waves (X-waves)

S. C. Walker[a)]
Marine Physical Laboratory, Scripps Institution of Oceanography, University of California San Diego, La Jolla, California 92093-0238

(Received 14 March 2008; revised 6 January 2009; accepted 7 January 2009)

In the traditional approach to X-waves, the X-wave field is synthesized from a superposition of solutions to the homogenous wave equation (in three-dimensions) without regard to boundary conditions. As a consequence the synthesized solution is acausal. Here, it is shown that the solution to the inhomogenous scalar wave equation for the acoustic field from a supersonic source distribution consistent with the radiation condition, i.e., a Mach front, defines a causal X-wave. Using the connection between X-waves and a physical source, it is shown that an X-wave can be generated from a planar aperture using time-reversal. By appealing to the demonstrated self-adaptivity of time-reversal processes, the method should allow for the generation of X-waves in arbitrary (inhomogenous) media. Typically, the generation of approximate acoustic X-waves from a planar aperture is achieved using a complicated annular transducer arrangement. Here, the time-reversal method for the generation of approximate acoustic X-waves is experimentally proven using a line transducer array in two-dimensional geometry in free space.
 [DOI: 10.1121/1.3075769]

PACS number(s): 43.60.Tj, 43.60.Gk, 43.20.Bi [DRD] Pages: 1549–1557

I. INTRODUCTION

Because of their unique properties, i.e., that they are nondiffracting, supersonic (superluminal), and intensely focused along a single axis, X-waves have generated a great deal of interest in both acoustics[1–3] and optics.[4–6] While it has been demonstrated that X-waves cannot transmit information at a speed that exceeds the speed at which waves move through the propagation medium,[7–9] due to the large depth of field, they still hold the potential for use in a variety of applications, particularly in the areas of imaging[2,10,11] and confinement.[12–14]

In order to experimentally study the properties and possible applications of X-waves, much work has been done pertaining to the generation of X-waves from planar apertures in both acoustics[15,16] and optics.[17–19] One standard strategy in the literature has been to calculate the desired X-wave solution projected over the plane of the aperture, and then to apply it as a signal to broadcast from the aperture.[15,20] The X-wave solution, as formulated by Lu and Greenleaf,[21] is an idealized solution to the homogenous wave equation that is infinite in both time and space. As a result, it is acausal and unphysical. One limitation of the standard approach to X-waves is that the solution can only be calculated analytically in homogenous media in situations involving a high degree of symmetry. While the infinite X-wave is diffractionless, broadcasting the X-wave signal from a finite aperture results in an approximate X-wave over a finite region of space during a finite interval of time. With the understanding that the underlying principles pertain to the optical case as well, this paper focuses on the X-waves and their generation in acoustic media.

Here, the field from a supersonic source in free space is derived and shown to be a causal form of the X-wave solution of Lu and Greenleaf that obeys the radiation condition. In essence, the Mach cone defines the X-wave. In this physical context, the Mach angle replaces the axicon parameter of the traditional X-wave derivation. Or, more correctly, the axicon parameter is shown to represent the axial phase speed of the source. It is then shown that the Mach cone interpretation of the X-wave can be exploited to generate an approximate X-wave from a finite planar aperture using time-reversal (TR). By using a line phased array to propagate a supersonic source signal and measuring the resulting field with a planar aperture, the field needed to produce an X-wave can be directly measured rather than calculated. Time-reversing the signal and broadcasting from the receive aperture then results in the aforementioned approximate X-wave. In this sense, the receive aperture functions as a time-reversal mirror (TRM).[22] Using TR, then, the generation of X-waves from planar apertures is no longer restricted to free space homogenous media. The self-adaptivity of TR[23,24] allows for the generation of X-waves along chosen trajectories (for example, curved trajectories) in any static medium (where TR will work), limited only by the feasibility of measuring the aperture field. As with any field generated by a finite aperture, X-wave or otherwise, the ability to focus energy is limited by diffraction according to Rayleigh resolution considerations. In addition to self-adaptivity, TR methods offer the advantage of implementation from simple phased arrays of Cartesian geometries.[24,25]

While the generation of approximate X-waves is typically implemented using annular arrays,[20,26] here experimental approximate acoustic X-waves are generated using a simple line acoustic transducer array in a two-dimensional (2D) geometry. Combining the focusing properties of

[a)]Electronic mail: shane@mpl.ucsd.edu; wak@mpl.ucsd.edu

X-waves with the self-adaptivity and ease of deployment offered by a TR approach to acoustic X-wave generation potentially broadens the scope of opportunities to which the unique focusing properties of X-waves can be applied. In addition to offering potential new methods for deployment in medical ultrasonic applications, TR X-waves might find application in nondestructive testing,[27,28] ocean acoustic communications[29] and focusing,[30] target scattering characterization,[2] and complex media propagation[23] (including waveguides [31]). While broadband TR techniques are not yet directly applicable to electromagnetic applications (due to the high frequencies involved), applications combining narrow band X-waves with multi-wave phase conjugation techniques are a possibility as well.

Section II provides a brief overview of the derivation and properties of the acausal X-wave solution followed by the derivation of the causal X-wave resulting from a supersonic source. The derivation expands the analysis presented by the authors in Ref. 9, in which it is shown that the field from a Cerenkov–Vavilov particle defines a causal X-wave, to include arbitrary signals. In Sec. III the connection between TR and the generation of X-waves from planar apertures is developed, and the implications for the physics of X-waves are discussed. The results of an experiment in which X-waves are generated using TR are presented in Sec. IV. The final section closes with concluding remarks.

II. BACKGROUND THEORY OF X-WAVES

In this section it is demonstrated that the Mach front field generated in a homogenous medium by a uniformly propagating supersonic source signal (for example, along a linear geometry phased array) defines a causally consistent version of the X-wave field proposed by Lu and Greenleaf. It will ultimately be argued that the fields resulting from signals propagating at supersonic speeds along selected trajectories (i.e., trajectories other than a straight line along the axial direction) are, in fact, also X-wave variants. While these X-wave variants are beyond the reach of traditional X-wave analytical methods, they are easily obtained experimentally using curvilinear phased arrays.

A. Radially standing X-waves: Brief summary of the traditional approach to the formulation of X-waves

The ideal, diffractionless X-wave is a superposition of conical wave solutions of the homogenous wave equation in cylindrical geometry in a homogenous medium. Denoting the medium sound speed by c, the standard X-wave formulation takes the form[1]

$$\chi(r,z,t) = \int_{-\infty}^{\infty} d\omega A_{\omega}(\omega) \left\{ J_0\left(\frac{\sin\zeta}{c}\omega r\right) \times \exp\left[i\omega\left(t - \left(\frac{\cos\zeta}{c}\right) z\right)\right]\right\}, \quad (1)$$

where J_0 is the zeroth order Bessel function of the first kind, and r, z, and t denote the radial, axial, and time coordinates, respectively. Here the quantity in the brackets is the conical wave solution of the homogenous wave equation with ω representing a dummy variable of integration and $A_{\omega}(\omega)$ denoting the coefficient (at the value ω) of the ω component of the conical wave solution. The conical waves, parametrized by the axicon angle ζ (see Fig. 1), define a cone that propagates supersonically along the z-axis, with the speed of propagation, v, related to the axicon angle of the cone, $v=c/\cos\zeta$. Making the substitution $\omega=-ck_z/\cos\zeta=-k_z v$ and rescaling the coefficient, $A_{k_z}(k_z)=|v|A_{\omega}(-vk_z)$, the solution can be recast,

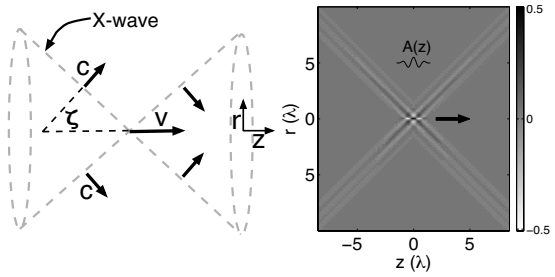


FIG. 1. Conical geometry of the acausal, infinite X-wave. As shown schematically (left panel), the X-wave, propagating at speed $|v|>c$, comprises both expanding and collapsing cylindrical waves. The axial speed of the X-wave is related to the axicon angle, $v=c/\cos\zeta$. A simulated X-wave solution (right panel) corresponding to a signal $A(z)$ exhibits the distinctive cross-cross pattern. The simulation is shown in units of the wavelength, λ, of the signal.

$$\chi(r,z,t) = \int_{-\infty}^{\infty} dk_z A_{k_z}(k_z) J_0(\tan\zeta k_z r) e^{ik_z(z-vt)}. \quad (2)$$

The above formulations take the form of Fourier transforms, and the integration variables ω and k_z can be interpreted to represent angular frequency and axial wavenumber, respectively. In accordance with the physical interpretations of the integration variables, the coefficients A_{ω} and A_{k_z} are the Fourier transforms (over time and axial coordinates, respectively) of a real signal function, $A=A(z-vt)$, provided that

$$A_{k_z}(k_z) = \begin{cases} |v|A^*_{\omega}(vk_z), & -\pi/2 < \zeta < \pi/2 \ (v>0) \\ |v|A_{\omega}(vk_z), & \pi/2 < \zeta < 3\pi/2 \ (v<0), \end{cases} \quad (3)$$

where (*) denotes complex conjugation. The X-wave solutions of Eqs. (1) and (2) describe an azimuthally symmetric superposition of conical waves that constructively interfere along the z-axis and that propagate without diffraction along the axial direction at a constant supersonic speed. This formulation of X-waves is termed radially standing as it comprises (see Fig. 1) a superposition of radially inbound and outbound components. As a consequence of the fact that the solution has been constructed in an infinite source-free medium without the imposition of boundary conditions, the field is infinitely supported, both spatially and temporally. As a result, the ideal, diffractionless X-wave solution is acausal and unphysical.

B. Causal X-waves: A formulation of X-waves for their production from phased arrays

This section demonstrates the connection between the X-wave solution described in the previous section and the

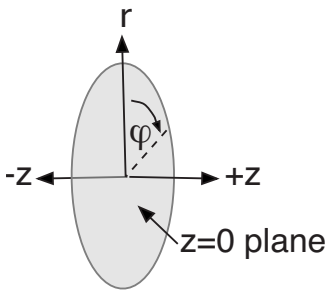


FIG. 2. Three-dimensional cylindrical geometry.

field generated by a supersonic source. As the aim is to ultimately produce X-waves experimentally, the analysis is carried out in the context of an acoustic signal propagated supersonically along an infinite line phased array of acoustic transducers.

Consider the inhomogenous wave equation for the velocity potential,

$$\left(\nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right)\psi(\boldsymbol{r},t) = S(\boldsymbol{r},t), \tag{4}$$

where $\boldsymbol{r}$ represents the position vector, t represents time, c is the speed of waves in the medium (assuming that medium is isotropic and homogenous), and the source, $S(\boldsymbol{r},t)$, represents a volume injection rate (per unit volume). In the three-dimensional cylindrical coordinate system shown in Fig. 2, the velocity potential obeys the following equation:

$$\left(\frac{1}{r}\frac{\partial}{\partial r}r\frac{\partial}{\partial r} + \frac{1}{r^2}\frac{\partial^2}{\partial \phi^2} + \frac{\partial^2}{\partial z^2} - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right)\psi(r,\phi,z,t) = S(r,\phi,z,t). \tag{5}$$

Performing the appropriate Hankel and Fourier transforms, the single valued solution for the velocity potential becomes

$$\psi_n(k_r,k_z,\omega) = -\frac{S_n(k_r,k_z,\omega)}{k_r^2 + k_z^2 - (\omega/c)^2}, \quad n = \{0, \pm 1, \ldots\}. \tag{6}$$

where k_r, k_z, and ω denote the radial wavenumber, the axial wavenumber, and the angular frequency, respectively, and the nth transform coefficient, S_n, is defined as

$$S_n(k_r,k_z,\omega) \equiv \int_{-\infty}^{\infty} dt \int_{-\infty}^{\infty} dz \frac{1}{2\pi}\int_0^{2\pi} d\phi \int_0^{\infty} dr r J_n(k_r r) \times S(r,\phi,z,t)e^{-i(2n\phi + k_z z + \omega t)}. \tag{7}$$

Here J_n denotes the nth order Bessel function of the first kind. The final solution for the velocity potential is given by a superposition of the solutions above,

$$\psi(r,\phi,z,t) = -\frac{1}{4\pi^2}\sum_{n=-\infty}^{\infty} e^{i2n\phi}\int_{-\infty}^{\infty} d\omega \int_{-\infty}^{\infty} dk_z \times \int_0^{\infty} dk_r k_r J_n(k_r r)\psi_n(k_r,k_{k_z},\omega)e^{i(k_z z + \omega t)}. \tag{8}$$

Introducing a source corresponding to the case of a real valued signal over a linear array contiguous with the z-axis,

$$S(r,z,t) = -A(z,t)\frac{\delta(r)}{r}, \tag{9}$$

where $A(z,t)$ represents the axio-temporal distribution of the source and $\delta(r)$ denotes the Dirac-delta function, results in a great simplification of the expression for the line source (LS) velocity potential, ψ_{LS},

$$\begin{aligned}\psi_{\mathrm{LS}}(r,z,t) &= \frac{1}{4\pi^2}\int_{-\infty}^{\infty} d\omega \int_{-\infty}^{\infty} dk_z \times \int_0^{\infty} dk_r k_r \frac{A(k_z,\omega)J_0(k_r r)}{k_r^2 + k_z^2 - (\omega/c)^2}e^{i(k_z z + \omega t)} \\ &= \frac{1}{8\pi^2}\int_{-\infty}^{\infty} d\omega \int_{-\infty}^{\infty} dk_z \times \int_{-\infty}^{\infty} dk_r k_r \frac{A(k_z,\omega)H_0^{(1)}(k_r r)}{k_r^2 + k_z^2 - (\omega/c)^2}e^{i(k_z z + \omega t)},\end{aligned} \tag{10}$$

where $A(k_z,\omega)$ is the Fourier transform of $A(z,t)$ and $H_0^{(1)}$ represents the zeroth order Hankel function of the first kind. Restricting the analysis to sources in uniform axial motion at velocity v, such that $A(z,t) \rightarrow A(z-vt)$, results in a further simplification,

$$\psi_{\mathrm{LS}}(r,z,t) = \frac{1}{4\pi}\int_{-\infty}^{\infty} dk_z \int_{-\infty}^{\infty} dk_r k_r \frac{A_{k_z}(k_z)H_0^{(1)}(k_r r)}{k_r^2 + (\gamma k_z)^2}e^{ik_z(z-vt)}, \tag{11}$$

where $A_{k_z}(k_z)$ represents the Fourier transform over the z coordinate and $\gamma \equiv \sqrt{1-(v/c)^2}$ is the Lorentz factor. Invoking the radiation condition, integration over the radial wavenumber results in the final form of the LS velocity potential,

$$\psi_{\mathrm{LS}}(r,z,t) = \frac{i}{2}\int_{-\infty}^{\infty} dk_z A_{k_z}(k_z)H_0^{(1)}(i\gamma k_z r)e^{ik_z(z-vt)}. \tag{12}$$

In the case that $|v| < c$, the Lorentz factor, γ, is real so that the Hankel function becomes the modified Bessel function, $H_0^{(1)}(i\gamma k_z r) \rightarrow (-2i/\pi)K_0(\gamma k_z r)$, which decays exponentially with r. The result is that the field from the line array source does not radiate. In the case that $|v| > c$, γ is imaginary. Using the definition, $\bar{\gamma} \equiv \sqrt{(v/c)^2 - 1} = -i\gamma$, the solution becomes

$$\psi_{\mathrm{LS}}(r,z,t) = \frac{i}{2}\int_{-\infty}^{\infty} dk_z A_{k_z}(k_z)H_0^{(1)}(\bar{\gamma}k_z r)e^{ik_z(z-vt)}, \tag{13}$$

which describes a radiating field. In fact, as shown in Fig. 3, it describes the Mach cone, or bow wave generated by a supersonic source, where $\zeta \equiv \arctan\bar{\gamma}$ is the angle at which the Mach front propagates with respect to the direction of source motion (in this case the positive z-axis). Note that Eq. (13) describes a convolution of the axial distribution of the source signal, $A_{k_z}(k_z)$, with the Green's function specific to a point source in uniform supersonic motion along an infinite line array along the z-axis, $H_0^{(1)}(\bar{\gamma}k_z r)e^{-ik_z vt}$.

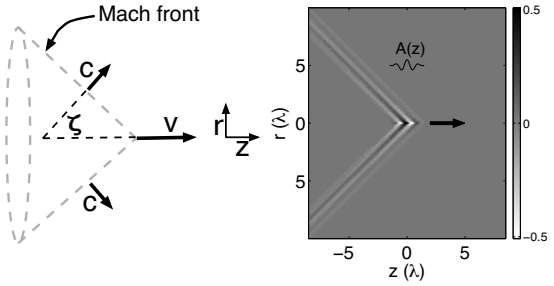


FIG. 3. The causal X-wave. The Mach cone (schematic of left panel) generated by a supersonic source, traveling at speed $v>c$, defines a causal X-wave comprising only expanding cylindrical waves. The angle at which the Mach cone propagates, $\zeta \equiv \arctan \bar{\gamma}$, is identically the axicon angle that characterizes the X-wave. The right panel shows a simulated causal X-wave (right panel) corresponding to a signal $A(z)$ The simulation is shown in units of the wavelength, λ, of the signal.

As shown in Fig. 4(a), a signal propagated along a phased array is effectively a wave packet (in the continuous limit), comprising both spatial and temporal components, propagating through space along the path defined by the array. In one interpretation, the signal represents a spatially distributed wave packet propagating along the array, $A(z-vt) \rightarrow A_{k_z}(k_z)e^{-ik_z vt}$. While an axial point source $[A_{k_z}(k_z)=1]$, being of infinite bandwidth, produces an infinite bandwidth front [Fig. 4(b)], an extended axial source distribution, being of finite bandwidth, results in a finite bandwidth front [Fig. 4(c)]. Equivalently, the signal may be interpreted as a phase delayed temporal signal, $A(t-z/v) \rightarrow A_\omega(\omega)e^{-i\omega z/v}$, where the relation between the spatial and temporal interpretations given in Eq. (3) applies. This relation can be used to recast Eq. (13) in terms of its angular frequency Fourier representation,

$$\psi_{\mathrm{LS}}(r,z,t) = \frac{i}{2}\int_{-\infty}^{\infty} d\omega A_\omega(\omega) H_0^{(1)*}\left(\frac{\bar{\gamma}}{v}\omega r\right) e^{i\omega(t-z/v)}. \qquad (14)$$

On comparison, the LS velocity potential solutions of Eqs. (13) and (14) bear striking similarity to the ideal, diffractionless X-wave solutions of Eqs. (1) and (2). As demonstrated in Ref. 9, the LS velocity potential solution describes a causal X-wave, with the Mach front defining the characteristic lobes of the X-wave and the angle of front

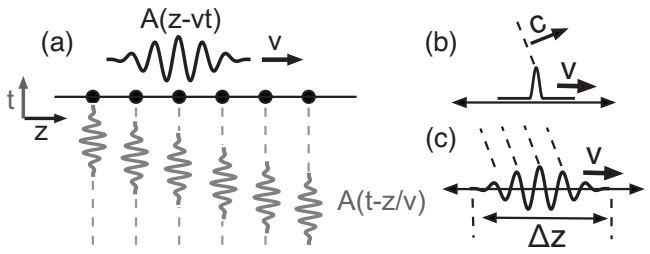


FIG. 4. Phased array signal schematic. Synchronizing the elements of a phased array produces a signal that propagates along the array. The signal is effectively a wave packet in the continuous limit. The gray vertical signal (along the dashed line left panel) represents the time-domain signal emitted by each element (dark circles). The black solid curve represents the spatial distribution of the synchronized signals moving at speed v along the array. A monopole pulse (center panel), with $\Delta z \rightarrow 0$, results in an infinite bandwidth spatio-temporal front, while a source distribution (right) results in a finite bandwidth front.

propagation, ζ, identified as the axicon angle, which obeys the relation $\bar{\gamma}=\tan\zeta$. The fundamental difference between the solutions is related to the method of derivation. In particular, the derivation of the X-wave solution, vis-à-vis an *ad hoc* superposition of solutions to the homogenous (source-free) wave equation, results in a radially standing solution that is acausal and unphysical. The LS solution, on the other hand, because it is derived from a physical source, obeys the radiation condition so that it is both physical and causal. Thus, the line array formulation offers a straightforward method for generating a causal X-wave experimentally.

III. X-WAVE GENERATION THROUGH TIME-REVERSAL

In this section, two methods for generating an approximate X-wave field from a planar aperture are discussed. The first employs the standard strategy popular in the literature,[15,16] in which the aperture signal for producing an approximate X-wave is calculated using the techniques of Sec. II A. Because an analytic X-wave solution is only possible for a homogenous medium and relies on radial symmetry, it is only possible to create approximate X-waves that propagate along a single axis and only in a homogenous medium. In the alternate method, the aperture signal for producing an approximate X-wave is determined experimentally using a supersonic source signal. TR of the resulting aperture signal then produces an approximate X-wave that propagates in the opposite direction to the original source signal. Consequently, using this method, it is possible to generate an X-wave that propagates along a selected trajectory in any static medium.

A. The standard approach to generating approximate X-waves

Consider the creation of an X-wave from a circular plane aperture perpendicular to the axial direction, as shown in Fig. 2. From Eq. (1), the aperture source signal, χ_{ap}, for the case of an aperture in the $z=0$ plane, is given by

$$\chi_{\mathrm{ap}}(r,t) = \int_{-\infty}^{\infty} d\omega A_\omega(\omega) J_0\left(\frac{\bar{\gamma}}{v}\omega r\right) e^{i\omega t}.$$

Figure 5 shows series of snapshots of χ_{ap} on the aperture. Notice that the aperture signal closely resembles a beam-steer signal in circular geometry.[32] Invoking Green's theorem (assuming an aperture of radius, σ), the field generated by the aperture signal is

$$\begin{aligned}\bar{\chi}(r,z,t) &= \int_0^{\sigma} r'dr' \int_0^{2\pi} d\phi' \int_{-\infty}^{\infty} d\omega A_\omega(\omega) J_0\left(\frac{\bar{\gamma}}{v}\omega r'\right) \\ &\quad \times \left\{\frac{\exp\left(-i\omega\frac{R}{c}\right)}{2\pi R}\right\} e^{i\omega t} \\ &= \int_0^{\sigma} r'dr' \int_0^{2\pi} d\phi' \frac{\chi_{\mathrm{ap}}(r', t-R/c)}{2\pi R}, \qquad (15)\end{aligned}$$

where the term in curly brackets is recognized as the three-dimensional spectral Green's function and R

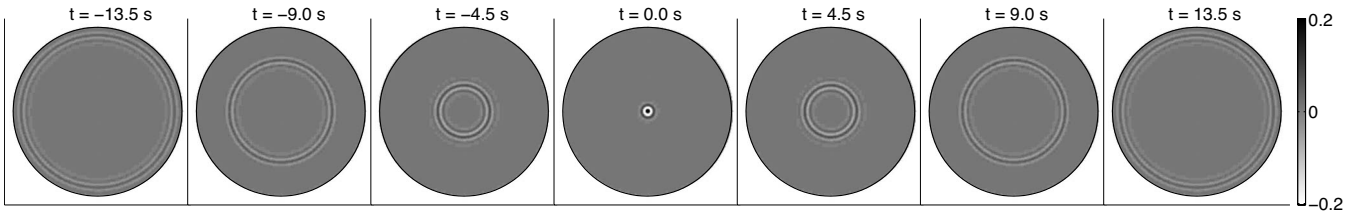


FIG. 5. Simulated X-wave aperture signal. Snapshots of the X-wave field projected over a finite planar aperture in the $z=0$ plane are visualized. In the case of the acausal X-wave, the field is described by a signal ring that collapses toward (for $t<0$), then expands from (for $t>0$), the center of the aperture. The acausal X-wave aperture signal is TR invariant. The causal X-wave aperture signal, on the other hand, exists only for times $t\geqslant 0$ and is not TR invariant.

$=\sqrt{r^2+r'^2-2rr'\cos(\phi-\phi')+z^2}$ is the distance between the source and field points.

The snapshots of the solution visualized in Fig. 6 demonstrate that the aperture signal, χ_{ap}, generates a field that approximates an X-wave over a finite region of space and a finite interval of time. As demonstrated schematically in Fig. 6, the axial region over which the approximate X-wave exists (in the high frequency limit) is given by[33] $\Delta z_\chi \equiv (\sigma_{\max}-\sigma_{\min})/\bar{\gamma}$, where $\sigma_{\max}$ and $\sigma_{\min}$ are the inner and outer radii of the signal aperture, respectively. In this case, the inner radius is $\sigma_{\min}=0$, so that the X-wave exists between $z=0$ and $z_{\max}=\sigma/\bar{\gamma}$. Thus, any experimentally realizable aperture generated approximate X-wave must be both finite in extent and approximately diffractionless over only a limited region in space. To produce the full X-wave that is infinitely supported and diffractionless over all space and time would require an infinite aperture and an infinite amount of time. It is in this sense that the X-wave solution of Sec. II A is unphysical. Also evident is the edge wave that results from the fact that the source aperture is of finite radius.

B. Using time-reversal to produce an approximate X-wave

Although it is not apparent in the case of the standing X-wave, the generation of X-waves from an aperture can be interpreted as a TR operation. To see this, consider the field, given by Eq. (14), produced by an axial source signal distribution, $A(t-z/v)$, propagated at uniform speed $v>c$ along a line array coincident with the z-axis such that at $t=0$ the source signal is at $z=0$. Recall that this field describes a causal X-wave. Referring to the geometry of Fig. 7, the field measured over a 2D transducer aperture at $z=0$ oriented perpendicularly to the line array is given by

$$\psi_{\mathrm{ap}}(r,t)=\frac{i}{2}\int_{-\infty}^{\infty}d\omega A_\omega(\omega)H_0^{(1)*}\left(\frac{\bar{\gamma}}{v}\omega r\right)e^{i\omega t}. \tag{16}$$

In contrast to χ_{ap}, which is acausal and can be TR invariant, the aperture signal, ψ_{ap}, derived from the measurement of the field generated by the LS, is never time-reversal invariant. For positive times the measured aperture field [Eq. (16)] is identical to the standing X-wave aperture field (see Fig. 5 for $t\geqslant 0$), while for negative times, the measured aperture field is zero,

$$\psi_{\mathrm{ap}}(r,t)=\chi_{\mathrm{ap}}(r,t),\quad t\geqslant 0\quad (\text{see } t\geqslant 0 \text{ snapshots in Fig. 5}),$$

$$\psi_{\mathrm{ap}}(r,t)=0,\quad t<0. \tag{17}$$

As shown in Fig. 8, TR and transmission of the measured signal, ψ_{ap}, generates an approximate X-wave supported over a finite region of space (out to a distance $z_{\max}=\sigma/\bar{\gamma}$) and time. For times $t\leqslant 0$ the field generated by the time-reversed aperture signal is identical to $\bar{\chi}$ [see Eq. (15)]. Consequently, these snapshots are not included in Fig. 8. However, because the time-reversed aperture signal is zero for $t>0$, the lagging extremities of the field (dashed boxes) are not supported (compared to the acausal field, $\bar{\chi}$). These extremities do not contribute to the axial behavior of the approximate X-wave and are not needed for its support.

For applications in homogenous media requiring only cylindrical symmetry, the TR method for producing approximate X-waves provides no advantage over the standard method as the required aperture field can be calculated directly using the ideal, diffractionless X-wave solution. Suppose, on the other hand, that the aim is to experimentally generate an X-wave in more complicated media for which it is not possible to generate an analytic or a numerical solution

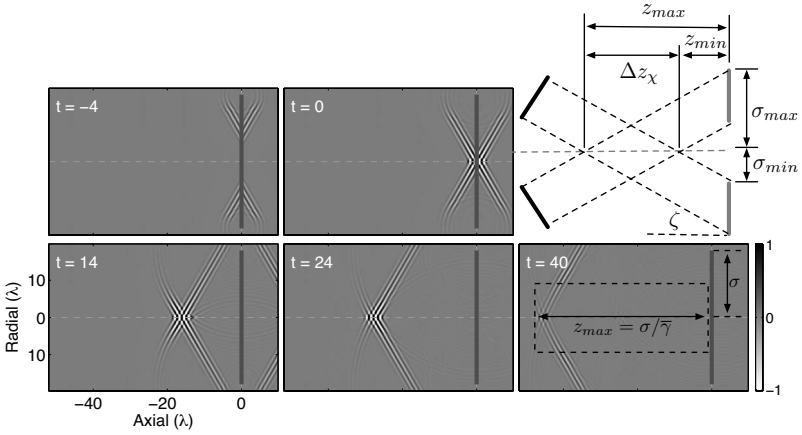


FIG. 6. Simulated field generated by an acausal X-wave aperture signal. The acausal X-wave aperture signal generates an approximate X-wave. The X-wave is supported out to a distance $z_{\max}=\sigma/\bar{\gamma}$, where σ is the radius of the aperture (thick gray vertical line) and $\bar{\gamma}=\tan\zeta$ quantifies the axial speed of the X-wave. The visualization is shown in units of wavelength, λ, of the signal field. Also depicted schematically is the geometric relation between $z_{\max}$ ($z_{\min}$), $\sigma_{\max}$ ($\sigma_{\min}$), and the range over which an approximate X-wave exists, Δz_χ. In the simulation, $\sigma_{\min}=0$.

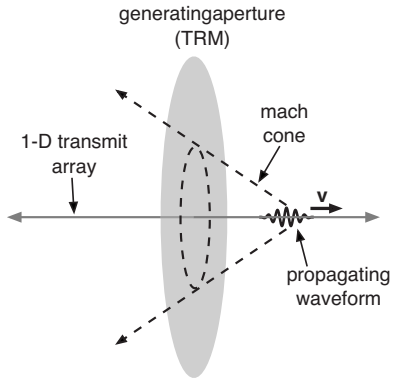


FIG. 7. Schematic for experimentally measuring a causal X-wave over a planar aperture. A planar aperture TRM (gray region) in the $z=0$ plane measures the field from a supersonic signal (black curve) propagated along a linear phased array (horizontal gray line) along the z-axis. The field generated by the supersonic signal is a causal X-wave.

for the required aperture signal. By invoking the implicit TR properties of X-waves, it is possible to self-adaptively produce X-waves along selected trajectories from aperture source arrays of arbitrary geometry and orientation in an arbitrary static medium. The possibilities are limited only by diffraction and experimental difficulties associated with deploying the arrays and measuring the fields.

C. A note on the supersonic properties of acausal and causal X-waves

An added consequence of the causal X-wave derivation is that it demonstrates that the acoustic intensity points in the direction propagation of the Mach cone, i.e., at angle ζ with respect to the z-axis. Thus, while the axial phase speed of the X-wave is equal to the source speed, the group velocity, $\boldsymbol{u}$, is defined such that $\boldsymbol{u}=c(\cos\zeta\hat{z}+\sin\zeta\hat{x})$; i.e., the group travels at speed c. The same applies to the acausal X-wave. The ideal, diffractionless, acausal X-wave can be thought of as the Mach front generated by an annular ring signal distribution collapsing radially (for $-\infty<t<t_0$) then expanding radially (for $t_0<t<\infty$) at supersonic speed $|v_r|=v/\bar{\gamma}$ on an infinite planar array (refer to Fig. 5, for a visualization) located at $z=-\infty$ (or $z=\infty$ for $v<-c$), where t_0 is the time at which the radius of the annular signal is zero. For times $t>t_0$, the cylindrical waves of the Mach front, propagating at group speed c in the direction $\zeta=\arctan(\bar{\gamma})$ with respect to the z-axis, constructively interfere along the z-axis, resulting in a localized high intensity component that propagates axially at phase speed $v>c$. In the infinite time/space limit, this field defines the ideal, diffractionless X-wave. Thus, in accordance with discussion presented in Ref. 7, the ideal, diffractionless, acausal X-wave is a manifestation of the scissor effect.

IV. USING TIME-REVERSAL TO GENERATE X-WAVES: EXPERIMENTAL RESULTS

While the conventional approach to X-wave theory focuses, out of necessity, on cylindrical geometry, in light of the discussion of Sec. III, using the TR approach, X-waves can be generated in many geometries. This section presents the results from an experiment in which X-waves are produced through TR in a 2D geometry rather than in the conventional three-dimensional cylindrical geometry. As a result, a brief derivation of the 2D X-wave is presented along with the experimental setup and results.

A. The two-dimensional X-wave

A similar analysis to that in Sec. II in 2D Cartesian coordinates (rather than three-dimensional cylindrical coordinates) results in the 2D X-wave formulation. Consider the wave equation [Eq. (4)] in a 2D Cartesian coordinate system (x,z), where, in analogy with the cylindrical X-wave result, z is still referred to as the axial coordinate and x represents the perpendicular coordinate. Assuming a source of the form,

$$S(x,z,t)=-A(z-vt)\delta(x),\quad |v|>c, \tag{18}$$

representing a supersonic source in uniform motion at velocity v along the axial direction leads to the 2D causal X-wave solution,

$$\psi^{(2D)}(x,z,t)=\frac{i}{2}\int_{-\infty}^{\infty}dk_z A_{k_z}(k_z)\frac{\exp(i\bar{\gamma}k_z|x|)}{\bar{\gamma}k_z}e^{ik_z(z-vt)}. \tag{19}$$

The ideal, diffractionless, 2D version of the Lu and Greenleaf X-wave is given by

$$\chi^{(2D)}(x,z,t)=\int_{-\infty}^{\infty}dk_z A_{k_z}(k_z)\frac{\sin(\bar{\gamma}k_z|x|)}{\bar{\gamma}k_z}e^{ik_z(z-vt)}. \tag{20}$$

Thus, in two-dimensions, the X-wave comprises a superposition of plane waves, rather than cylindrical waves, traveling at an angle ζ (defined as $\tan\zeta=\bar{\gamma}$) with respect to the axial direction. In terms of the angular frequency, the aperture signal for producing causal X-waves from a planar aperture in the $z=0$ plane becomes

$$\psi_{ap}^{2D}(x,t)=\frac{i}{2}\int_{-\infty}^{\infty}d\omega A_{\omega}(\omega)\frac{\exp\left(-i\frac{\bar{\gamma}}{v}\omega|x|\right)}{\frac{\bar{\gamma}}{v}\omega}e^{i\omega t}. \tag{21}$$

As with the aperture field from a LS in three-dimensions [see discussion following Eq. (16)], the 2D aperture field, being zero for $t<0$, is never TR invariant. Time-reversal and broadcast of ψ_{ap}^{2D} from the TRM aperture is then given by

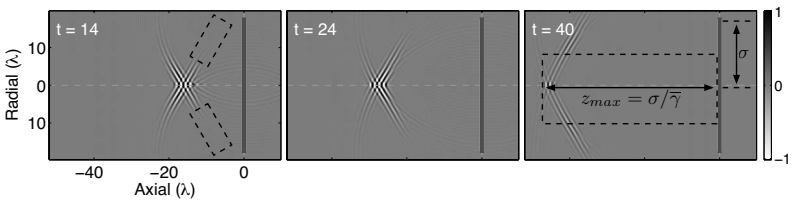


FIG. 8. Simulated field generated from causal X-wave aperture signal using TR. TR of the causal X-wave aperture signal generates an approximate X-wave. The X-wave is supported out to a distance $z_{max}=\sigma/\bar{\gamma}$. The field for $t<0$ is identical to the acausal case shown in Fig. 6. Because the TR aperture is zero for $t>0$, the lagging extremities of the field (dashed box) are not supported. The visualization is shown in units of wavelength, λ, of the signal field.

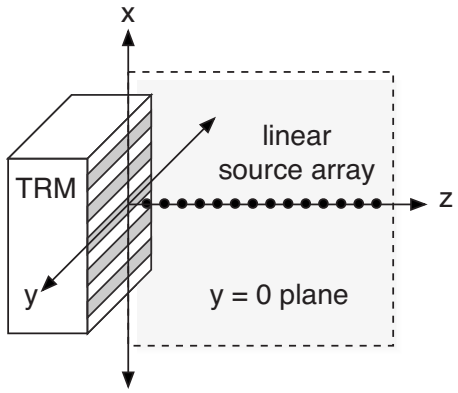


FIG. 9. Experimental setup schematic. An omnidirectional hydrophone was used to sample the field from the transducer elements of a TRM. The TRM comprised a planar aperture of acoustic transducers. The hydrophone sampled the field at evenly spaced points along the z-axis, thus synthesizing a HLA. Using reciprocity, the fields were then synchronized and summed to mimic a supersonic source. The summed field was then time-reversed and transmitted from the TRM and the resulting field densely sampled in the $y=0$ plane.

$$\psi_{\mathrm{TR}}^{\mathrm{2D}}(r,z,t)=\int_{-\sigma}^{\sigma}x'dx'\int_{-\infty}^{\infty}d\omega\frac{A_{\omega}^{*}(\omega)}{\frac{\bar{\gamma}}{v}\omega}\left\{H_0^{(1)}\left(\frac{\omega}{c}R\right)\right\}$$
$$\times\exp\left[i\omega\left(t+\bar{\gamma}\frac{|x'|}{v}\right)\right], \tag{22}$$

where the term in curly brackets is recognized as the 2D spectral Green's function and $R=\sqrt{(x-x')^2+z^2}$ is the distance between the source and field points.

B. Experimental setup

A series of experiments were conducted (the results from three of which are presented in Sec. IV C) in which approximate X-waves were generated from a planar aperture TRM using TR. The experiments were carried out in a homogenous, unbounded, static, fresh water medium at ultrasonic frequencies. As shown schematically in Fig. 9, an ultrabroadband Hann modulated pulse of bandwidth $\Delta f=1.5$ MHz at a carrier frequency of $f_c=1.5$ MHz was broadcast separately from each element of an ultrasonic TRM device and sampled by a piezo-electric ultrasonic hydrophone situated on the z-axis. Repeating the same measurement operation to other hydrophone locations resulted in a synthetic receive aperture along the z-axis. The synthetic aperture constituted the horizontal line array (HLA) along which the source signal was propagated. Using a value of $c=1487$ m/s for the medium, the wavelength at the carrier frequency was $\lambda\approx 1$ mm. In two of the experiments (cases A and B below), the HLA comprised 24 elements spaced at 1.5 mm ($3\lambda/2$) intervals along the z-axis. In a third experiment (case C below), the HLA comprised 75 elements spaced at approximately 0.5 mm intervals along a curved trajectory. The TRM, in the $z=0$ plane, comprised an array of 60 piezo-electric ultrasonic transducer elements evenly spaced along the x-axis at 0.5 mm ($\approx\lambda/2$) increments. Using reciprocity, the individual TRM-to-HLA fields were then synchronized and summed to mimic a supersonic source traveling along the HLA toward the TRM. The synthesized field was then time-reversed and broadcasted from the TRM. The resulting field was densely sampled (at 0.5 mm$\approx\lambda/2$ intervals in both z and x) in the $y=0$ plane. Because the TRM elements span a width of $\sigma=15$ mm ($\approx 15\lambda$) in the y-direction, the TRM is directional, producing a beam that does not vary appreciably in the y-direction. Due to the geometry of the TRM (i.e., because it is extended in the y-direction), the aperture signal and resulting time-reversed field (in the $y=0$ plane) are well modeled by the 2D case of Eqs. (21) and (22), respectively.

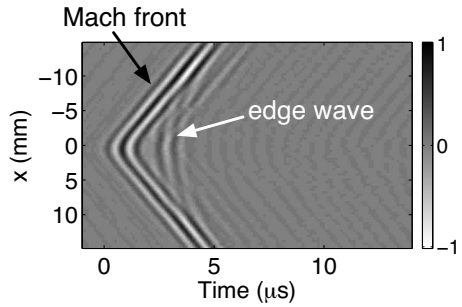


FIG. 10. Experimental TRM aperture signal. Visualized is the time-domain field measured over the TRM from a supersonic source traveling along the HLA at velocity $v=-1604$ m/s (toward the TRM). Both the Mach front and the edge wave generated by the finite length of the HLA are visible.

C. Experimental results

The results from three experimental cases are discussed. The first two cases (cases A and B) consider a signal in uniform supersonic motion along a linear geometry HLA at different speeds. In the third case (case C), the source signal was propagated along a curved geometry HLA. In all three cases the field generated by the HLA signal was sampled by a TRM, time-reversed, and rebroadcast.

In case A the source signal was propagated at speed $v_A=1604$ m/s toward the TRM, generating a Mach front traveling at an angle $\zeta=\arccos c/v_A=22°$ ($\bar{\gamma}=0.41$). The corresponding field measured on the TRM, visualized in Fig. 10, demonstrates the arrival of the Mach front. Note that time $t=0$ is defined to be the moment at which the source signal (traveling along the HLA) reaches the plane of the TRM aperture ($z=0$). Also visible is a trailing edge wave that results from the fact that the HLA is finite in length, extending from $z=0$ to a distance of 36 mm along the z-axis. The measured TRM signal was then time-reversed and broadcast. As expected, the time-reversed field, shown in Fig. 11, can be characterized as an approximate X-wave over a finite region of space and time. In this case the speed of the source was chosen so that the X-wave would be supported out to the distance of the furthest HLA element, $z_{\max}=\sigma/\bar{\gamma}=15$ mm$/0.41=36$ mm.

Recall in the simulation results of Sec. III that the finite edges of the TRM aperture produce an edge wave. In the current experimental case, both the TRM and the HLA are finite, so that one expects two sets of edge waves: one from the HLA that will be measured by the TRM aperture and one from the TRM when the signal is time-reversed. However, while the HLA edge wave is visible in both the measured

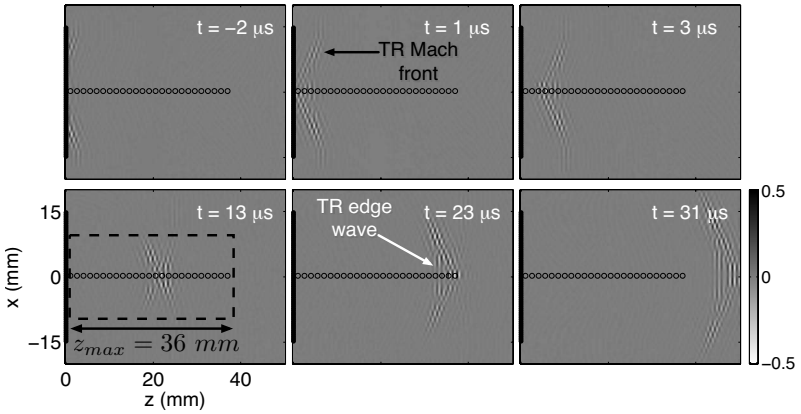


FIG. 11. Experimental TR approximate X-wave field. TR of the measured TRM signal from Fig. 10 results in an approximate X-wave over the length of the HLA. Also visible is received edge wave component that is now time-reversed. As mentioned in Fig. 9, the visualization is in the $y=0$ plane. The thick black line and open circles represent the TRM and the HLA, respectively.

TRM aperture field and the resulting time-reversed field (Figs. 10 and 11), the TRM generated edge wave seems to be missing entirely. This is easily explained when one considers the directionality of the individual TRM transducer elements. As shown in Fig. 12, the transducer directionality in the x-z ($y=0$) plane suppresses the high angle components of the received and transmitted fields.

In case B, the source signal was propagated along the HLA toward the TRM at a speed of $v_B=-2200$ m/s, producing a Mach front at an angle $\zeta=47°$ ($\bar{\gamma}=1.07$). As seen in the visualization of Fig. 13, the Mach front is highly suppressed due to the directionality of the TRM transducer elements. Only contributions from directly along the axis of the HLA and the edge wave are visible. As a result, TR of the measured TRM signal in this case does not result in an approximate X-wave.

In case C the source signal was propagated at a constant speed, $v_C=1604$ m/s, toward the TRM along the curved trajectory traced out by the HLA shown in Fig. 14. As shown, time-reversing the sampled TRM field results in an approximate X-wave whose vertex retraces the HLA trajectory as it propagates away from the TRM. In this case, the side-lobes of the generated approximate X-wave are slightly curved so that the field no longer assumes an exact X-shape. Using TR, then, this experimental result demonstrates that it is possible to generate exotic approximate X-waves that trace out curved trajectories along a large depth of field. The choice of trajectory is arbitrary with the following caveats. The transducer directionality of the TRM must be considered as field components that arrive on the TRM will be modulated by the transducer directionality. Care must be taken to define a source trajectory that does not generate field components that travel at steeper angles than the transducers can measure. In addition, the effects of the interference from the HLA edge wave must be considered. Finally, one must also realize that as the TRM aperture is finite, the approximate X-wave solution is diffractive so that the X-wave vertex resolution obeys the Rayleigh criterion. Consequently, the resolution of any approximate X-wave is expected to decrease with distance from the generating aperture.

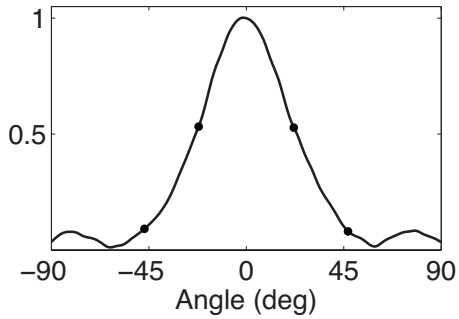


FIG. 12. Transducer/hydrophone pair directionality. The directionality of the measurement between a single TRM transducer element and the omnidirectional hydrophone was experimentally determined. Field components traveling in a relative direction of greater than 45° are highly suppressed.

V. CONCLUSIONS

In the Lu and Greenleaf formulation,[21] the X-wave solution is derived as a superposition of conical waves without connection to boundary conditions or a source. As a result, the solution, though it is diffractionless, is acausal. Also, the solution requires radial symmetry and a homogenous medium. Here, an alternate derivation of the X-wave field has been introduced whereby the field generated by a supersonic source distribution in free space is derived and shown to be a causal X-wave (in three-dimensions) that obeys the radiation condition with the Mach cone defining the X-wave. One implication of this result is that it demonstrates that all supersonic properties of X-waves, both acausal and causal, are phase rather than group phenomena. A method for generating X-waves from planar apertures using time-reversal that takes advantage of the connection with a physical source (and thus causality) has also been introduced. In contrast to existing

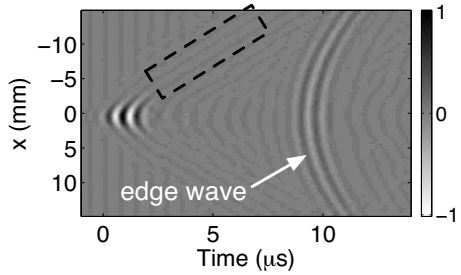


FIG. 13. Experimental TRM aperture signal: example of the effects of directional suppression on X-wave generation. Visualized is the field measured over the TRM from a supersonic source traveling along the HLA at velocity $v=-2200$ m/s (toward the TRM). The directionality of the transducer elements has suppressed the appearance of the Mach front (comparing to Fig. 10, the front should appear in the dashed box region). As a result it is not possible to generate an X-wave for steep axicon angles, ζ.

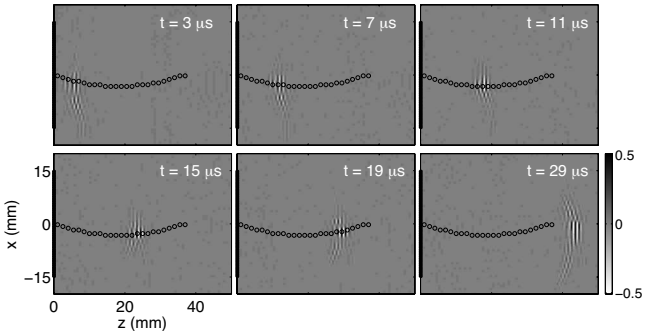


FIG. 14. Experimental curved trajectory approximate X-wave field. TR of the field generated by a signal propagated along a curved trajectory HLA results in an approximate X-wave that retraces the curved trajectory of the HLA. The thick black line and open circles represent the TRM and the HLA, respectively. The visualization is in the $y=0$ plane (see Fig. 9).

methods, where the generating aperture field is calculated using the Lu and Greenleaf X-wave solution (which limits applications to cylindrical geometry in homogenous media), the TR method is self-adaptive and should allow for the generation of X-waves in selected geometries in any static inhomogenous media. The method has been experimentally proven in a 2D experimental geometry in free space. To accompany the experimental results, the 2D form of the X-wave has been derived.

In addition to offering potential new methods for deployment in medical ultrasonic applications, acoustic TR X-waves might find application in nondestructive testing, ocean acoustics, complex media characterization, and waveguide physics. While broadband TR techniques are not yet directly applicable to electromagnetic applications (due to the high frequencies involved), applications combining narrow band X-waves with multi-wave phase conjugation techniques are a possibility. Consequently, the presented acoustic results have potential implications for electromagnetic field analogies in fiber laser and femtosecond laser applications, among others.

ACKNOWLEDGMENT

This work was supported by ONR.

[1]P. Stepanishen, "Acoustic bullets/transient Bessel beams: Near to far field transition via an impulse response approach," J. Acoust. Soc. Am. **103**, 1742–1751 (1998).
[2]P. Marston, "Scattering of a Bessel beam by a sphere," J. Acoust. Soc. Am. **121**, 753–758 (2007).
[3]K. Cunningham and M. Hamilton, "Bessel beams of finite amplitude in absorbing fluids," J. Acoust. Soc. Am. **108**, 519–525 (2000).
[4]D. McGloin and K. Dholakia, "Bessel beams: Diffraction in a new light," Contemp. Phys. **46**, 15–28 (2005).
[5]D. Faccio, A. Averchi, A. Dubietis, P. Polesana, A. Piskarskas, P. Trapani, and A. Couairon, "Stimulated Raman x waves in ultrashort optical pulse filamentation," Opt. Lett. **32**, 184–186 (2007).
[6]F. Xu, J. Liu, R. Li, and Z. Xu, "Nonlinear x-wave formation and conical emission at different powers of a femtosecond laser pulse in water," Opt. Commun. **275**, 433–439 (2007).
[7]W. Rodrigues, D. Thober, and A. Xavier, "Causal explanation for observed superluminal behavior of microwave propagation in free space," Phys. Lett. A **284**, 217–224 (2001).
[8]A. Shaarawi and I. Besieris, "Relativistic causality and superluminal signalling using x-shaped localized waves," J. Phys. A **33**, 7255–7263 (2000).
[9]S. Walker and W. Kuperman, "Cherenkov-Vavilov formulation of x waves," Phys. Rev. Lett. **99**, 244802 (2007).
[10]M. Moawad, A. Shaarawi, and I. Besieris, "Characterization of a nonrigid sphere using the backscattered fields of acoustic x waves," J. Acoust. Soc. Am. **115**, 2937–2946 (2004).
[11]J. Lu, T. Song, R. Kinnick, and J. Greenleaf, "In vitro and in vivo real-time imaging with ultrasonic limited diffraction beams," IEEE Trans. Med. Imaging **12**, 819–829 (1993).
[12]P. Marston, "Negative axial radiation forces on solid spheres and shells in a Bessel beam," J. Acoust. Soc. Am. **122**, 3162–3165 (2007).
[13]T. Grosjean, D. Courjon, and D. V. Labeke, "Bessel beams as virtual tips for near-field optics," J. Microsc. **210**, 319–323 (2003).
[14]J. Arlt, K. Dholakia, J. Soneson, and E. Wright, "Optical dipole traps and atomic waveguides based on Bessel light beams," Phys. Rev. A **63**, 063602 (2001).
[15]P. Stepanishen and J. Sun, "Acoustic bullets: Transient Bessel beams generated by planar apertures," J. Acoust. Soc. Am. **102**, 3308–3318 (1997).
[16]H. Masuyama, T. Yokoyama, K. Nagai, and K. Mizutani, "Generation of Bessel beam from equiamplitude-driven annular transducer array consisting of a few elements," Jpn. J. Appl. Phys., Part 1 **38**, 3080–3084 (1999).
[17]A. Chatzipetros, A. Shaarawi, I. Besieris, and M. Abdel-Rahman, "Aperture synthesis of time-limited x waves and analysis of their propagation characteristics," J. Acoust. Soc. Am. **103**, 2287–2295 (1998).
[18]J. Lu, B. Unit, M. Clinic, and M. Rochester, "Designing limited diffraction beams," Ultrasonics **44**, 181–193 (1997).
[19]P. Fox, J. Cheng, and J. Lu, "Theory and experiment of Fourier–Bessel field calculation and tuning of a pulsed wave annular array," J. Acoust. Soc. Am. **113**, 2412–2423 (2003).
[20]M. Fatemi and K. Ghaseminejad, "Generation of nondiffracting x waves using annular array transducers," J. Acoust. Soc. Am. **92**, 2457 (1992).
[21]J. Lu and J. Greenleaf, "Nondiffracting x waves-exact solutions to free-space scalar wave equation and their finite aperture realizations," Ultrasonics **39**, 19–31 (1992).
[22]M. Fink, "Time-reversal mirrors," J. Phys. D: Appl. Phys. **26**, 1333–1350 (1993).
[23]M. Fink, C. Prada, F. Wu, and D. Cassereau, "Self focusing in inhomogeneous media with time reversal acoustic mirrors," Proc.-IEEE Ultrason. Symp. 681–686 (1989).
[24]W. Kuperman, W. Hodgkiss, H. Song, T. Akal, C. Ferla, and D. R. Jackson, "Phase conjugation in the ocean: Experimental demonstration of an acoustic time-reversal mirror," J. Acoust. Soc. Am. **103**, 25–40 (1998).
[25]F. Wu, J. Thomas, and M. Fink, "Time reversal of ultrasonic fields. IL. Experimental results," Ultrasonics **39**, 567–578 (1992).
[26]S. Holm, "Bessel and conical beams and approximation with annular arrays," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **45**, 712–718 (1998).
[27]C. Prada, E. Kerbrat, D. Cassereau, and M. Fink, "Time reversal techniques in ultrasonic nondestructive testing of scattering media," Inverse Probl. **18**, 1761–1773 (2002).
[28]A. McNab and M. Campbell, "Ultrasonic phased arrays for nondestructive testing," NDT Int. **20**, 333–337 (1987).
[29]W. S. Hodgkiss, P. Roux, W. Kuperman, and T. Akal, "Coherent MIMO time reversal communications in shallow water," Oceans'04, MTS/IEEE Techno-Ocean'04, Vol. **4**, pp. 2225–2229 (2004).
[30]S. C. Walker, W. A. Kuperman, and P. Roux, "Synchronized time-reversal focusing with application to remote imaging from a distant virtual source array,"J. Acoust. Soc. Am. (2009), In press.
[31]P. Roux and M. Fink, "Time reversal in a waveguide: Study of the temporal and spatial focusing," J. Acoust. Soc. Am. **107**, 2418–2429 (2000).
[32]J. Lu and J. Greenleaf, "Steering of limited diffraction beams with a two-dimensional array transducer," Proc.-IEEE Ultrason. Symp. 603–607 (1992).
[33]J. Lu and J. Greenleaf, "Theory and acoustic experiments of nondiffracting x-waves [medicalus imaging application]," Proc.-IEEE Ultrason. Symp. **2**, 1155–1159 (1991).

Comparing two proposed measures of cochlear mechanical filter bandwidth based on stimulus frequency otoacoustic emissions[a)]

Ben Lineton[b)] and Catriona M. B. Wildgoose
Institute of Sound and Vibration Research, University of Southampton, Highfield, Southampton SO17 1BJ, United Kingdom

(Received 26 September 2008; revised 11 December 2008; accepted 12 December 2008)

It has been hypothesized that the sharpness of the cochlear mechanical filter is related to two measures based on stimulus frequency otoacoustic emissions (SFOAEs). The first is the group delay of the SFOAE; the second is the bandwidth of the SFOAE two-tone suppression isoinput tuning characteristic. A corollary of this is that natural variability in cochlear mechanical bandwidth within a population would lead to a positive correlation between these two SFOAE-based measures of tuning within that population. To test this prediction, SFOAE group delay and SFOAE two-tone suppression isoinput tuning characteristics were measured in a sample of 16 audiometrically normal subjects. Contrary to the prediction, no statistically significant correlation was found between the two bandwidth measures. Cochlear model simulations were used to aid the interpretation of this result. These suggested that a positive correlation between the two measures is expected, but that it may well be too weak to detect with the given sample size, due to the influence on the SFOAE measures of random inhomogeneities in basilar membrane impedance.
 [DOI: 10.1121/1.3068452]

PACS number(s): 43.64.Bt, 43.64.Jb, 43.64.Kc [BLM] Pages: 1558–1566

I. INTRODUCTION

The mechanical response of a given point on the basilar membrane to a sound presented in the ear canal can be represented as the output of a nonlinear bandpass filter. The filter bandwidth, denoted here $BW_{BM\text{-}Mech}$, depends both on stimulus frequency (SF) and stimulus level (e.g., Robles and Ruggero, 2001), and thus in the present article, $BW_{BM\text{-}Mech}$ is assumed to refer to an average value over a limited range of SF and level. The present article is concerned with the question of whether measurements of otoacoustic emissions (OAEs) can reveal differences in $BW_{BM\text{-}Mech}$ between subjects.

The currently prevailing theories of cochlear mechanics predict that the sharpness of a cochlear mechanical filter should be related to the latency of evoked OAEs, provided they are dominated by coherent reflection of the traveling wave (Shera and Zweig, 1993a; Zweig and Shera, 1995). Thus, for example, the group delay of SFOAEs has been taken as a measure of the cochlear filter bandwidth (Shera *et al.*, 2002; Shera and Guinan, 2003). In the current article, such bandwidth estimates based on SFOAE group delay will be denoted $BW_{SF\text{-}GD}$. Shera *et al.* (2002) argued that the magnitude of $BW_{SF\text{-}GD}$ in humans is roughly in agreement with psychoacoustic measures across a wide range of SFs, supporting the link between $BW_{SF\text{-}GD}$ and $BW_{BM\text{-}Mech}$.

Further indirect support for this theory comes from the nonlinear behavior of the cochlea. *In vivo* measurements in mammals show a broadening of cochlear tuning (i.e., an increase in $BW_{BM\text{-}Mech}$) with stimulus level (e.g., Robles and Ruggero, 2001). Thus a similar increase in $BW_{SF\text{-}GD}$ with stimulus level is also expected. This prediction has been borne out by measurements of SFOAE group delay, and related quantities (Kemp and Brown, 1983; Lineton and Lutman, 2003b; Konrad-Martin and Keefe, 2005; Schairer *et al.*, 2006), as well as by latency-versus-level measurements in transient-evoked OAEs, thought to be closely related to SFOAEs (Norton and Neely, 1987; Neely *et al.*, 1988; Tognola *et al.*, 1997). Thus there is a strong argument, based on both theory and experiment, for a link between $BW_{SF\text{-}GD}$ and $BW_{BM\text{-}Mech}$.

A second possible OAE-based measure of cochlear bandwidth is suggested by two-tone suppression (2TS) tuning curves of SFOAEs (e.g., Kemp and Chum, 1980; Brass and Kemp, 1993; Keefe *et al.*, 2008), which show a bandpass characteristic. The bandwidth estimate obtained from a 2TS paradigm is denoted here $BW_{SF\text{-}2TS}$. Whether current cochlear mechanical theories predict that $BW_{SF\text{-}2TS}$ should be related to $BW_{BM\text{-}Mech}$ is not as clear-cut as it is for $BW_{SF\text{-}GD}$. This is because the 2TS tuning characteristic is the outcome of complicated nonlinear responses to both a probe tone (i.e., the evoking tone) and a suppressor tone for which no generally agreed theory exists. Furthermore, whether or not measurements of $BW_{SF\text{-}2TS}$ show the expected increase with increasing stimulus level is also unclear, and may depend on the stimulus paradigm and method of analysis. For example, Keefe *et al.* (2008) found no clear level dependence in a

[a)]Portions of this work were presented in Lineton, B., and Wildgoose, C. "Attempts to estimate cochlear filter bandwidth using otoacoustic emissions: Experiments and models," *British Society of Audiology Short Papers Meeting on Experimental Studies of Hearing and Deafness, University College London, UK, 20–21 September 2007.*

[b)]Author to whom correspondence should be addressed. Electronic mail: bl@isvr.soton.ac.uk

bandwidth measure defined from isosuppression tuning curves of SFOAEs, in contrast to the effects of stimulus level seen in $BW_{SF\text{-}GD}$, mentioned above.

Nevertheless, the following heuristic argument, similar to that put forward by Brass and Kemp (1993), does predict that $BW_{SF\text{-}2TS}$ will be correlated with cochlear bandwidth. Consider a simple stimulus paradigm for obtaining $BW_{SF\text{-}2TS}$ in which a probe tone is presented at a fixed SF and level, while a suppressor tone is swept in frequency, holding its stimulus level equal to that of the probe tone. The 2TS-isoinput tuning characteristic is then obtained by measuring the amplitude of the SFOAE at the probe frequency as a function of suppressor frequency. If we assume that the degree of SFOAE suppression depends on the degree of overlap between the traveling wave envelopes of the probe tone and suppressor tone (when presented in isolation) then the bandwidth of the 2TS-isoinput tuning characteristic will depend on the widths of the traveling wave envelopes of the probe and suppressor tones, which in turn depend on $BW_{BM\text{-}Mech}$. A more explicit theoretical demonstration of how this could arise is presented as part of the present study, using a nonlinear model of cochlear mechanics due to Kanis and de Boer (1994), which has been adapted to allow SFOAE 2TS-isoinput tuning characteristics to be predicted, as in Lineton and Lutman (2003a). The results of this particular model confirm the heuristic argument: That the simulated $BW_{SF\text{-}2TS}$ is correlated with $BW_{BM\text{-}Mech}$. Note, however, that this result may depend on the particular implementation of nonlinear cochlear mechanics chosen by Kanis and de Boer (1994).

The aim of the present study is to estimate the degree of correlation between $BW_{SF\text{-}GD}$ and $BW_{SF\text{-}2TS}$ in otologically normal adults. If both these measures are correlated with $BW_{BM\text{-}Mech}$, and if there is any variation in $BW_{BM\text{-}Mech}$ within the population, then $BW_{SF\text{-}GD}$ and $BW_{SF\text{-}2TS}$ will show some degree of positive correlation with each other within the population.

II. METHODS

A. Subjects

Initially 20 volunteers (10 females and 10 males) aged 22–27 years participated in the study. One ear per subject was chosen at random for OAE measurements (ten left and ten right ears). All ears were normal on otoscopy, tympanometry (middle ear pressure range −100 to +50 daPa and middle ear compliance range 0.3–1.5 ml) and pure-tone audiometry (hearing threshold level ≤15 dB at 0.25, 0.5, 1, 2, 4, and 8 kHz). Subjects were seated comfortably, asked to remain as still as possible, and to swallow as infrequently as was comfortable during recording. Following SFOAE measurements, 4 of the 20 subjects were excluded from the analysis, based on the signal-to-noise ratio (SNR) of the measurements (Sec. III B), leaving 16 subjects in the analysis.

All experiments were carried out in accordance with the guidelines of the Declaration of Helsinki, and were approved by the ISVR Human Experimentation Safety and Ethics Committee.

B. Equipment

The equipment used is specifically designed for the measurement of distortion product otoacoustic emissions (DPOAEs) and SFOAEs. It comprises a PC containing a Loughborough Sound Images DSP card that controlled a remote converter module, which is connected to the two earphones and microphone of an Etymotic ER-10B+probe assembly.

The Etymotic ER-10B+probe assembly comprised a low-noise microphone, which was connected through a preamplifier to the input of the remote converter module, and two ER-2 insert earphones delivering tones to the ear via coupling tubes running through the body of the probe. The two earphones were connected to two output channels on the remote converter module. This allowed two tones to be presented to the ear simultaneously via the two earphones, thus minimizing nonlinear artifacts due to the measurement system during 2TS experiments.

The remote converter module comprised stereo 16 bit DACs, stereo 20 bit ADCs, and a sample rate generator. Tones were generated digitally in the DSP card with a sample rate of 32.768 kHz at frequencies which were multiples of 16 Hz. Each stereo output sample pair was sent to the remote converter module via a fast serial connection on the DSP and output under the timing of the sample rate generator. The input sample pair was acquired synchronously and sent via the serial connection to the DSP. The ADCs employed 8× oversampling obviating the need for antialiasing filters. Only the most significant 16 bits of the ADCs were used by the DSP for the present work. The tones were segmented into epochs of 62.5 ms (2048 sample points) containing a whole number of periods of the stimulus tones. The sampled microphone signal was recorded via one ADC in corresponding epochs and loaded into buffers of 2048 points. On completion of the epoch and before the next epoch begins, an in-place fast Fourier transform (FFT) with 16 Hz resolution is calculated. The FFT is then transferred to the PC for further processing. For simple recording of SFOAEs, only one DAC channel was used, while for suppressed SFOAEs both channels were used. The first epoch was always discarded since it contained the transient response of the transducers and the ear. The tone for which the SFOAE is calculated is designated the "probe" tone, while the second tone is designated the "suppressor."

Averaging of the complex FFT was performed across subsequent epochs to improve SNR. Noisy epochs, as determined by the noise power in a number of spectral lines adjacent to the SF, were excluded from the average. Noisy epochs were rejected when the average level in the ten spectral lines either side of the SFOAE frequency (but excluding the spectral line at the frequency of the suppressor tone, if present) exceeded a preset rejection level set at 10 dB SPL. Averaging of the epochs continued until 20 epochs had been accepted.

The measurement system was calibrated with the OAE probe fitted with an ear tip, and inserted into the DB2012 accessory (external ear simulator) of a Bruel and Kjaer Type 4157 ear simulator (IEC 711). The earphones were calibrated

at 16 Hz intervals across the measured frequency range. Throughout this paper, the earphone level is defined as the level that would be recorded at the reference microphone of the ear simulator at the same applied voltage level. The OAE probe microphone was calibrated at 1 kHz only, such that, for a 1 kHz tone presented via the probe earphone, the probe microphone gave the same reading as the ear simulator microphone.

C. Measurement procedures

Two series of measurements were performed, the first being SFOAE probe-frequency sweeps for estimating $BW_{SF\text{-}GD}$, and the second being 2TS-isoinput tuning characteristics for estimating $BW_{SF\text{-}2TS}$.

1. Probe-frequency sweep measurements

In the SFOAE probe-frequency sweep measurements, the probe frequency, f_1, was increased from 992 to 1984 Hz in 16 Hz intervals, giving 63 points, while the probe-level was held constant at 40 dB SPL. In order to obtain an estimate of the SFOAE pressure, p_{SF}, at each point, a version of the high-level suppressor method (e.g., Kemp and Brown, 1983; Kemp *et al.*, 1990) was used whereby recordings are made in two different conditions (termed the *unsuppressed* and *maximally suppressed* conditions) at each probe frequency. In the unsuppressed condition, the probe tone is presented alone to give a measurement of the complex ear canal pressure, which is assumed to comprise the SFOAE pressure added to the "passive" ear canal pressure (i.e., the pressure that would be obtained if the cochlear amplifier were inactive). In the maximally suppressed condition, the probe tone is presented together with a high-level suppressor tone to give an estimate of the passive ear canal pressure alone. A suppressor tone at 60 dB SPL at a frequency 16 Hz above the probe tone was used for this condition. The two measurement conditions were completed within a short time interval (typically <2 s) in order to minimize the "drift" noise component, which appears in the derived value of p_{SF} due, for example, to gradual changes in probe-fit or middle ear pressure, leading to changes in the passive impedance of the ear canal occurring over the duration of the two measurements. Two replicate probe-frequency sweeps were performed in order to allow estimation of the test-retest reliability of $BW_{SF\text{-}GD}$.

2. SFOAE 2TS-isoinput tuning characteristic measurements

Previous studies of SFOAE 2TS have often used an isosuppression paradigm, in which the suppressor level is altered at each suppressor frequency so as to achieve a criterion degree of suppression of the SFOAE at the probe frequency (e.g., Kemp and Chum, 1980). However, an isoinput paradigm, in which the suppressor level is held constant as its frequency is swept, can also be used to obtain a bandpass characteristic whose bandwidth is expected to depend on the underlying cochlear mechanical tuning (e.g., Brass and Kemp, 1993). Currently, there is no strong argument for favoring one paradigm over the other. In the present study, the isoinput paradigm was adopted with both the suppressor and the probe level set to 40 dB SPL (the same level as for the probe-sweep measurements). The weak rationale for this is that both the traveling waves in the 2TS condition are as close as possible to those in the probe-sweep condition. This paradigm is also simpler to implement experimentally than the isosuppression paradigm.

In 2TS measurements, the probe tone was fixed in frequency and level, while the suppressor tone was swept in frequency at constant level. The probe frequency was set close to 1488 Hz (i.e., the center of the range swept above), but the exact probe frequency was chosen for each individual ear from within the range 1328–1552 Hz to ensure that a sufficiently high amplitude SFOAE existed prior to introducing the suppressor tone, based on visual inspection of the SFOAE probe-sweep measurements. The probe and suppressor levels were fixed at 40 dB SPL throughout.

The 2TS-isoinput tuning characteristic was constructed from two sweeps of the suppressor frequency. The first was a coarse sweep from approximately 0.67 to 1.5 times the probe frequency, in steps of 64 Hz. The second was a fine sweep from approximately 0.83 to 1.2 times the probe frequency, in steps of 16 Hz, to provide higher resolution in the region around peak suppression. The point in the sweep where the suppressor frequency was exactly equal to the probe frequency was omitted.

The measurement condition with both probe and suppressor present is called the *partially suppressed* condition. In order to minimize any effects on the 2TS measurements of drift in the probe-fit, after each partially suppressed measurement, an additional unsuppressed measurement of the ear canal pressure was made, with the probe presented in isolation. Since the probe remains unaltered over the 2TS measurements, these are simply repeated measurements in the same condition. However, they were included to provide an accurate reference condition made close in time (typically within 2 s) to the preceding partially suppressed condition, to allow the change in ear canal pressure due to the suppressor to be established. A second replicate set of 2TS measurements was obtained to allow estimation of the test-retest reliability of $BW_{SF\text{-}2TS}$.

III. RESULTS

A. Obtaining SFOAE pressures from the measured ear canal pressure

Over the two measurement paradigms, there are in all three conditions in which the complex acoustic ear canal pressure is measured (unsuppressed, partially suppressed, and maximally suppressed), which are denoted $p_{EC;Un}$, $p_{EC;PS}$, and $p_{EC;MS}$, and which are shown schematically in Fig. 1 for a given probe frequency.

1. Analysis of probe-frequency sweep measurements

In the probe-sweep paradigm, $p_{EC;Un}$ and $p_{EC;MS}$ are measured. From this, the (complex) unsuppressed SFOAE pressure is obtained from

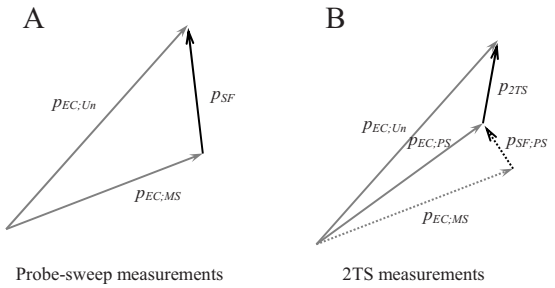


FIG. 1. Complex representation of the ear canal pressure and the SFOAE pressure in the two measurement paradigms. The dependence on frequency is omitted. (A) Probe-sweep measurements of the ear canal pressure, with two conditions: unsuppressed ($p_{EC;Un}$) and maximally suppressed ($p_{EC;MS}$), where the residual SFOAE is assumed negligible. Thus the unsuppressed SFOAE pressure (p_{SF}) can be calculated. (B) 2TS measurements of the ear canal pressure, with two conditions: unsuppressed ($p_{EC;Un}$) and partially suppressed ($p_{EC;PS}$). The 2TS pressure change (p_{2TS}) can then be calculated. The maximally suppressed ear canal pressure ($p_{EC;MS}$) and the partially suppressed SFOAE pressure (p_{SF}) are not obtained in this paradigm, but are shown (dotted arrows) for comparison.

$$p_{SF} = p_{EC;Un} - p_{EC;MS}, \tag{1}$$

where the explicit dependence on probe frequency, f_1, is omitted. It has been assumed in Eq. (1) that the ear canal pressure in the maximally suppressed condition is approximately equal to the passive ear canal pressure. It is also convenient to use the complex SFOAE pressure normalized with respect to the stimulus pressure (Shera and Zweig, 1993b), as defined by

$$p_{SF/stim} \equiv p_{SF}/p_{EC;MS} = p_{EC;Un}/p_{EC;MS} - 1. \tag{2}$$

2. Analysis of SFOAE 2TS-isoinput tuning characteristic measurements

In the 2TS paradigm, $p_{EC;PS}$ and $p_{EC;Un}$ are measured (but not $p_{EC;MS}$). The change in the ear canal pressure due to the suppressor is called the 2TS pressure change, and is defined as

$$p_{2TS} \equiv p_{EC;Un} - p_{EC;PS}, \tag{3}$$

which is used to generate 2TS-isoinput tuning characteristics, and to calculate $BW_{SF\text{-}2TS}$. The explicit dependence on probe and suppressor frequencies, f_1 and f_2, is omitted.

It is instructive to note that p_{2TS} can be expressed as the change in SFOAE pressure between the unsuppressed condition, and the partially suppressed condition

$$p_{2TS} \equiv p_{SF} - p_{SF;PS},$$

where

$$p_{SF;PS} \equiv p_{EC;PS} - p_{EC;MS} \tag{4}$$

and where $p_{SF;PS}$ is the SFOAE in the partially suppressed condition (Fig. 1).

B. Exclusion of subjects based on SNR estimates

Having obtained two replicate waveforms for $p_{SF}(f_1)$ from the probe-sweep measurements, and $p_{2TS}(f_2)$ from the 2TS measurements, the SNRs were assessed. To this end, the replicate waveform cross-correlation coefficients for $p_{SF}(f_1)$ and $p_{2TS}(f_2)$ were calculated across their measured frequency ranges. Subjects were then excluded from further analysis if the waveform cross-correlation coefficient was <0.5 for either measurement paradigm. This led to four subjects being excluded. For the remaining 16 subjects, the replicate waveform cross-correlation coefficients ranged from 0.89 to 0.99 (median of 0.98) for $p_{SF}(f_1)$, and from 0.51 to 0.93 (median of 0.71) for $p_{2TS}(f_2)$.

C. Obtaining $BW_{SF\text{-}GD}$ from the SFOAE probe-frequency sweep

Two typical examples of the normalized SFOAE pressure are shown in Fig. 2. According to the cochlear mechanical theory, the group delay of the SFOAE, which is proportional to the slope of the phase (Fig. 2), is related to $BW_{BM\text{-}Mech}$.

The measure used here to arrive at $BW_{SF\text{-}GD}$ is based on a weighted average of the group delay of the SFOAE across the measured frequency range. There are two stages in calculating the weighted average; first the SFOAE data are transformed to take account of the expected variation in group delay with SF, and second, a weighting is applied to the phase data in order to minimize the effects of regions where the SNR is poor, and thus the phase unreliable (e.g., S8 at 1650 Hz, Fig. 2). These stages in the analysis are explained in the remainder of this section.

The phase, φ, and group delay, τ, of the normalized SFOAE are defined by

$$\phi(f) \equiv \arg\{p_{SF/stim}(f)\},$$

$$\tau(f) \equiv -\frac{1}{2\pi}\frac{d\phi}{df}, \tag{5}$$

where f is the probe frequency f_1, the subscript being dropped for convenience. Note that Eq. (5) requires the use of the "unwrapped" phase. For an idealized scaling-symmetric cochlea, the group delay would vary directly with the reciprocal of frequency (Zweig and Shera, 1995). However, measurements of SFOAE (Shera *et al.*, 2002; Shera and Guinan, 2003) and tone-burst evoked OAEs (Norton and Neely, 1987; Neely *et al.*, 1988; Jedrzejczak *et al.*, 2004) suggest that the human cochlea deviates from ideal scaling-symmetry such that the group delay varies approximately as

$$\tau(f) \propto \sim \frac{1}{f^n}, \tag{6}$$

where n is approximately 0.5 (reported values range from 0.37 to 0.57). Note that $n=1$ would correspond to perfect scaling-symmetry. It can be shown that if Eq. (6) holds, then the slope of the phase becomes constant when the independent variable, f, is transformed using

$$\eta \equiv f^{1-n}/f_R^{-n}, \tag{7}$$

where η is "transformed frequency." The variable η has the dimensions of frequency, and equals the untransformed frequency f, when f equals the arbitrary reference frequency, f_R. The transformed group delay, τ', is then given by

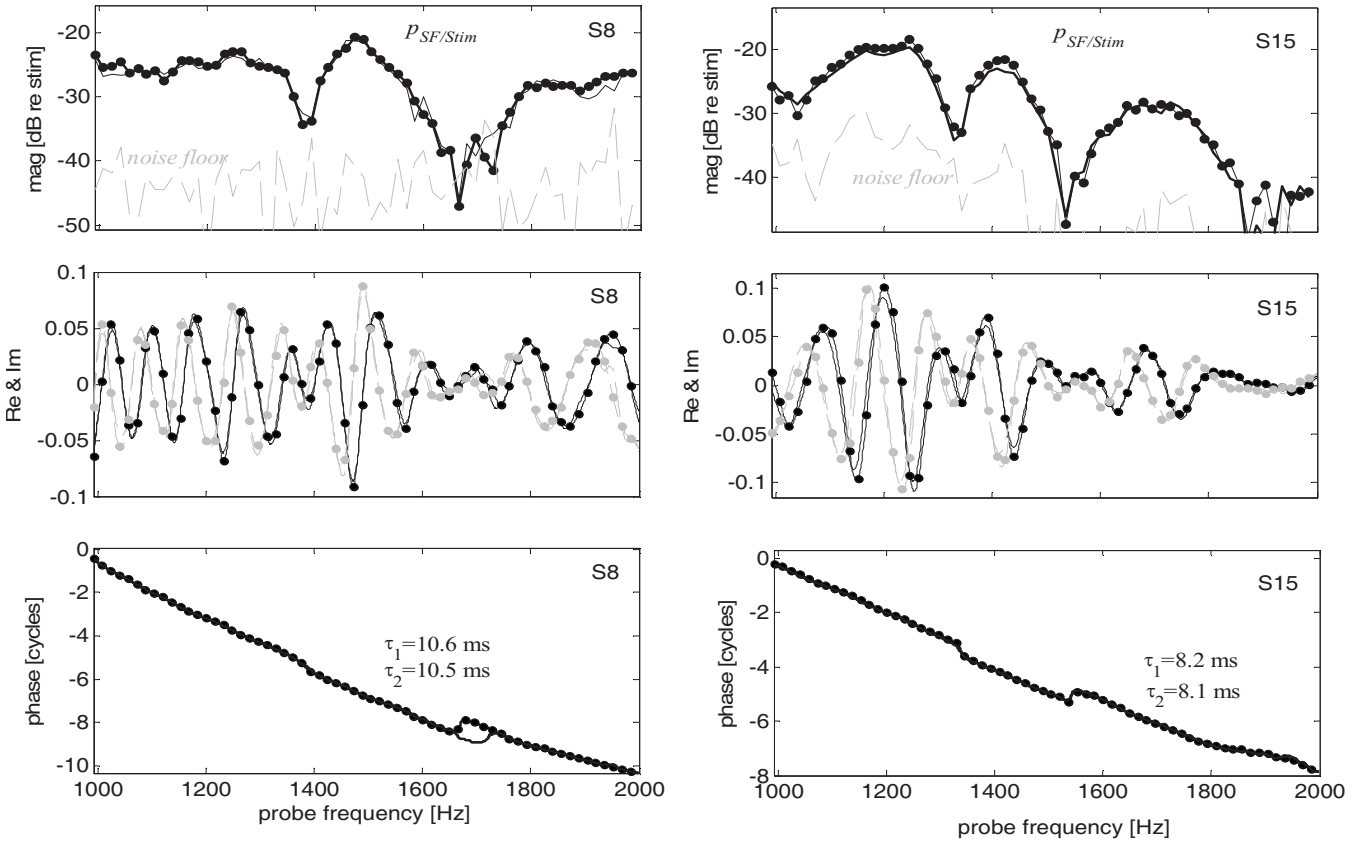


FIG. 2. Nondimensional SFOAE pressure from the probe-sweep measurements from subjects 8 (left-hand column) and 15 (right hand column). Top row: the absolute value of ($p_{SF/stim}$) (black lines) and the estimated noise floor (gray lines). Middle row: the real part (black lines) and the imaginary part (gray line). Bottom row: phase, and the estimated average group delay, τ_{av}. In all rows, for replicate 1, the measured data points are shown together with the smooth line, obtained from the measured points by low-pass interpolation. For replicate 2, only the smooth line is shown (not always visible, as it closely overlays replicate 1).

$$\tau'(\eta) \equiv -\frac{1}{2\pi}\frac{d\phi(\eta)}{d\eta}, \tag{8}$$

which is constant across the measured range, provided Eq. (6) holds true.

After transforming the SFOAE magnitude and phase into η coordinates, the transformed frequency averaged weighted group delay, τ_{av}, is calculated from

$$\tau_{av} \equiv \frac{A}{\eta_B - \eta_A}\int_{\eta_A}^{\eta_B} \tau'(\eta) w(\eta) d\eta, \tag{9}$$

where η_A and η_B are the transformed frequency values at the lower and upper values of the frequency sweep, $w(\eta)$ is the weighting factor, and A is a scale factor. The weighting factor is given by

$$w(\eta) \equiv |p_{SF}(\eta)|^2/p_{SF\text{-}av}^2,$$

$$p_{SF\text{-}av}^2 \equiv \frac{1}{\eta_B - \eta_A}\int_{\eta_A}^{\eta_B} |p_{SF}(\eta)|^2 d\eta. \tag{10}$$

The scale factor A is given by

$$A \equiv \frac{\eta_B - \eta_A}{f_B - f_A}, \tag{11}$$

where f_A and f_B are the frequency values at the lower and upper values of the frequency sweep. This scale factor simply accounts for the systematic overall scaling between f and η and ensures that the value of τ_{av} would not be affected by the choice of n in the idealized case where $w(\eta)=1$ for all η.

The value of n chosen for the analysis is $n=0.5$, and is the same for all subjects. This gives a transformed frequency intermediate between linear and logarithmic frequency scales. Note that the results of this analysis are highly insensitive to the choice of n. Indeed, it would have been possible to perform the analysis without first transforming the frequency [i.e., choosing $n=0$ in Eq. (7)]. However, since it is known that $n=0.5$ tends to "straighten out" the SFOAE phase variations, because of the underlying cochlear mechanics, it was reasoned that analyzing the transformed data would improve the correspondence with these underlying mechanisms. The value of f_R in Eq. (7) has no effect on the analysis.

The effect of the weighting factor, w, is more significant than the choice of n. Without the weighting factor, the frequency averaged group delay in Eq. (9) would simply depend on the overall change in the unwrapped phase across the measured range, as seen in Fig. 2, which includes steps in the phase where the signal is either zero or buried in noise. The effects of these steps are greatly de-emphasized by using the weighting in Eq. (10).

Equations (5) and (7)–(11) are implemented for the discrete data here by first using "low-pass" interpolation to pro-

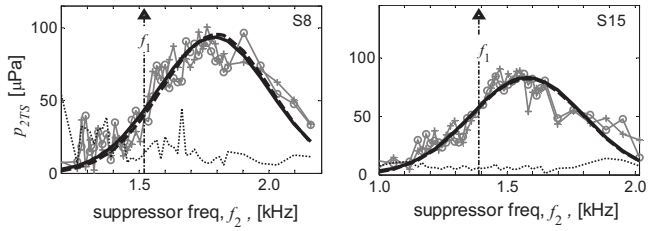


FIG. 3. SFOAE 2TS-isoinput tuning characteristic measurements from subjects 8 (left-hand column) and 15 (right-hand column). The measured points for two replicates are (circles and crosses) are joined with gray solid lines. The fitted Gaussian curves for replicates 1 and 2 are shown in black solid and dotted lines, respectively (but as they overlay each other closely, the dotted line is barely visible). The estimated noise floor is shown by the dotted line. The vertical dotted-dashed line indicates the frequency of the fixed probe tone at which the SFOAE is measured. The triangle indicates the unsuppressed SFOAE pressure obtained from the probe-sweep measurement. The bandwidths obtained from the fitted curves for subject 8 were 210 and 224 Hz (replicates 1 and 2); and for subject 15 were 222 and 221 Hz (replicates 1 and 2).

vide the smoothed curves seen in Fig. 2. The transformation in Eq. (7) was then performed with $n=0.5$; and the average group delay, τ_{av}, obtained by replacing the continuous integrals in Eqs. (9) and (10) with numerical integration by the trapezium rule. The value of $BW_{SF\text{-}GD}$ is then found from

$$BW_{SF\text{-}GD} \equiv 1/\tau_{av}. \tag{12}$$

D. Obtaining $BW_{SF\text{-}2TS}$ from the SFOAE 2TS-isoinput tuning characteristic

The 2TS-isoinput tuning characteristics were plotted as the absolute value of $p_{2TS}(f_2)$ as a function of the suppressor frequency f_2. The coarse and fine sweeps were merged into a single set of data points. The isoinput tuning characteristics appear as band-pass-like 2TS suppression curves seen for two typical subjects in Fig. 3.

There are several possible methods of obtaining a bandwidth from these curves. With no underlying theory for guidance, an analytical approach that is relatively robust to the presence of noise was chosen. The chosen method was to use the second moment of area, I_{xx}, about the center frequency, f_C (defined from the first moment of area), to obtain the bandwidth, $BW_{SF\text{-}2TS}$, as follows:

$$BW^2_{SF\text{-}2TS} \equiv I_{xx}/\text{Area},$$

$$I_{xx} \equiv \int_{-\infty}^{\infty} (f_2 - f_C)^2 y(f_2) df_2,$$

$$\text{Area} \equiv \int_{-\infty}^{\infty} y(f_2) df_2,$$

$$f_C \equiv \frac{1}{\text{Area}} \int_{-\infty}^{\infty} f_2 y(f_2) df_2,$$

$$y \equiv |p_{2TS}(f_2)|. \tag{13}$$

The values of $BW_{SF\text{-}2TS}$ and f_c were obtained from the measured 2TS-isoinput tuning characteristics using numerical integration by the trapezium rule over the measured frequency range, in place of the continuous integrals in Eq. (13).

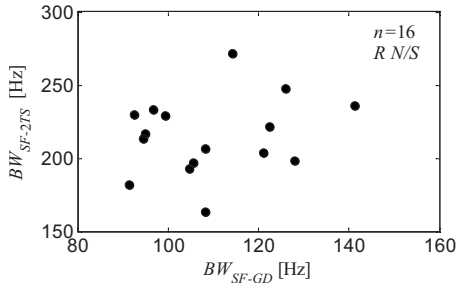


FIG. 4. Cross plot of measurements of $BW_{SF\text{-}2TS}$ against $BW_{SF\text{-}GD}$ across the 16 subjects included the analysis. The Pearson correlation coefficient, R, was 0.22, which was not statistically significant at the 5% level ($p>0.05$, 95% CI for R is −0.30 to 0.65).

These two values can be used to generate a fitted Gaussian curve to the data, using

$$y_{\text{fit}} = \frac{\text{Area}}{BW_{SF\text{-}2TS}\sqrt{2\pi}} \exp\left(\frac{-(f_2 - f_C)^2}{2BW^2_{SF\text{-}2TS}}\right), \tag{14}$$

which provides the visual comparison seen in Fig. 3, and illustrates that BW is the analog of the standard deviation of a Gaussian distribution. Note that the fitted value itself, y_{fit}, is purely calculated for visual illustration; only $BW_{SF\text{-}2TS}$ is used in the subsequent analyses.

E. Test-retest reliability of $BW_{SF\text{-}GD}$ and $BW_{SF\text{-}2TS}$ estimates

To assess the reliability of the estimates, the test-retest correlation coefficients across the sample of 16 subjects for both $BW_{SF\text{-}GD}$ and $BW_{SF\text{-}2TS}$ were calculated from the two replicate measurement sets, and found to be 0.98 and 0.85, respectively. A one-sample Kolmogorov–Smirnov test indicated that the sample distributions for $BW_{SF\text{-}GD}$ and $BW_{SF\text{-}2TS}$ did not differ from a normal distribution at the 5% level of significance.

F. Correlation between $BW_{SF\text{-}GD}$ and $BW_{SF\text{-}2TS}$

The average of the two replicate values was calculated to provide a single value of $BW_{SF\text{-}GD}$ and $BW_{SF\text{-}2TS}$ for each ear. Note that the absolute values of the BW estimates are not of interest here, and are not expected to be similar either to each other, or to any direct mechanical or neural BW estimate. Instead, it is the covariation in the two estimates across subjects that is of interest. The correlation coefficient was thus calculated between $BW_{SF\text{-}2TS}$ and $BW_{SF\text{-}GD}$ (Fig. 4) and found to be 0.22 ($p>0.05$, 95% confidence interval (CI) for R is −0.30 to 0.65). Thus, contrary to the predictions of the theory proposed in Sec. I, there is no hint in the data that the two estimates are correlated in the population.

It should be noted that $BW_{SF\text{-}GD}$ and $BW_{SF\text{-}2TS}$ were necessarily obtained from a sweep over a range of frequencies (here, up to one octave), rather than at a single frequency, and thus they will characterize some form of average bandwidth over this swept range. For a positive correlation to be predicted, it has been assumed that the mechanical sharpness of tuning is a characteristic that extends over this range of frequencies, such that a cochlea that is more sharply tuned

than the population average at one frequency is also more sharply tuned than average at neighboring frequencies, across the frequency range of interest.

IV. DISCUSSION

The possible explanations for the absence of any correlation seen in the data are several. It may be that there is a flaw in the underlying theory, in that in reality there is no correlation in the population between $BW_{BM\text{-}Mech}$ and either $BW_{SF\text{-}2TS}$, or $BW_{SF\text{-}GD}$, or both. Or it may be that there is a correlation, but it is too weak to detect with the given sample size (i.e., the study is underpowered). With a sample size of 16, the correlation would have to be at least 0.6 in order for the statistical power to be 80%, with a level of significance of 5%.

In order to investigate whether these empirical findings contradict current cochlear mechanical theories, a series of simulations of SFOAEs was performed with a cochlear model based on these theories. A more detailed description of these simulations and their results is presented in the Appendix. The keys result is that the model does predict positive pairwise correlations between $BW_{SF\text{-}2TS}$, $BW_{SF\text{-}GD}$, and $BW_{BM\text{-}Mech}$, but that these correlations are likely to be rather weak. This is because both $BW_{SF\text{-}2TS}$ and $BW_{SF\text{-}GD}$ are partly dependent on the particular realization of the random inhomogeneities arrayed along the basilar membrane, which leads to a large component in both these measures that is unrelated to $BW_{BM\text{-}Mech}$. Thus the models suggest that there may be a slight positive correlation between $BW_{SF\text{-}2TS}$ and $BW_{SF\text{-}GD}$ in the otologically normal population, but that a much larger sample size may be required to find it. If the model is correct, then one implication of its predictions is that neither $BW_{SF\text{-}2TS}$ nor $BW_{SF\text{-}GD}$, as currently defined, is likely to be able to usefully indicate whether a subject's cochlear tuning is significantly better or worse than the population-average. However, the experimental data do not allow this prediction to be confirmed. Furthermore, no attempt has been made to model the effects of mild sensory hearing loss, which (provided that the SFOAEs remain measurable) would be expected to cause a broadening in cochlear tuning.

V. CONCLUSIONS

In a sample of 16 normally hearing adults, this study found no statistically significant correlation between a measure of cochlear tuning bandwidth based on SFOAE group delay, and a measure based on the SFOAE 2TS-isoinput tuning characteristic.

A nonlinear cochlear model based on current theories of cochlear mechanics and OAE generation predicted some degree of correlation between these two measures, but also suggested that any correlation may well be weak. Thus, the absence of any correlation seen in the measurements may be due to the limited sample size rather than a complete absence in the target population. The model results also suggest that neither of the two OAE measures is likely to be a sensitive indicator of between-subject differences in cochlear tuning bandwidth.

APPENDIX: THEORETICAL SUPPORT FOR VALIDITY OF $BW_{SF\text{-}2TS}$ AS A CORRELATE OF $BW_{BM\text{-}Mech}$ BASED ON COCHLEAR MODEL SIMULATIONS

Unlike for the group delay, the theoretical link between $BW_{SF\text{-}2TS}$ and $BW_{BM\text{-}Mech}$ has not been clearly demonstrated. To investigate this link further, a series of simulations was performed using the nonlinear cochlear model described in detail by Lineton and Lutman (2003a). Briefly, this model is a variant of the quasilinear long-wave model developed by Kanis and de Boer (1994), into which random inhomogeneities in the basilar membrane impedance have been introduced in order to generate SFOAEs. The hydromechanics follows the standard long-wave model, and satisfies the requirement of Zweig and Shera's (1995) coherent scattering theory that the traveling wave exhibits a "tall-and-broad" peak, and that the scattering impedance should be "fine-grained" or "densely spaced." The nonlinearity is implemented at each point along the cochlea as an active amplification force, which follows a saturating nonlinear function (hyperbolic tangent) of the basilar membrane displacement.

As in Lineton and Lutman (2003a), the model formulation and parameters are as given by Kanis and de Boer (1994), with the following alterations. First random inhomogeneities were generated from a normally distributed perturbation of the passive damping term in the point-impedance of the basilar membrane along the length of the cochlea to give the scattering impedance. The amplitude of this random perturbation was set to have a standard deviation of 1% of the passive damping term. This relatively small impedance inhomogeneity has little effect on the primary mechanical response of the cochlea, but is sufficient to generate a backward traveling wave that appears as a SFOAE of realistic amplitude (typically 30–50 dB below the stimulus level). Second, the middle ear model of Kringlebotn (1988) was used in place of Kanis and de Boer's (1994) reflectionless middle ear, to better account for the reverse transmission through the middle ear. Third, the gain parameter, γ, was set to values other than $\gamma=1$, in order to create a population of cochlear models with differing mechanical tuning. As the model incorporates nonlinear saturation of the cochlear amplifier, it allows both SFOAE probe-sweeps and SFOAE 2TS-isoinput tuning characteristics to be simulated. The models were implemented in the MATLAB® software package (The Mathworks), and the full listings of the MATLAB code may be obtained from Lineton (2001).

This cochlear model has been shown to successfully match (at least qualitatively) several nonlinear features seen in measurements of cochlear mechanics in animal studies, and of SFOAEs in humans. The matches between the model and experimental findings include the basic shape of the basilar membrane mechanical frequency response (Kanis and de Boer, 1993); broadening of the mechanical frequency response with stimulus level (Kanis and de Boer, 1993); the approximate half-octave shift of the peak of the traveling wave with increased stimulus level (Kanis and de Boer, 1993); aspects of mechanical 2TS, such as phase shifts (Kanis and de Boer, 1994); SFOAEs with a spectral period (related to group delay) similar to that in humans (Lineton and Lutman, 2003a, 2003b); increases of SFOAE-spectral period

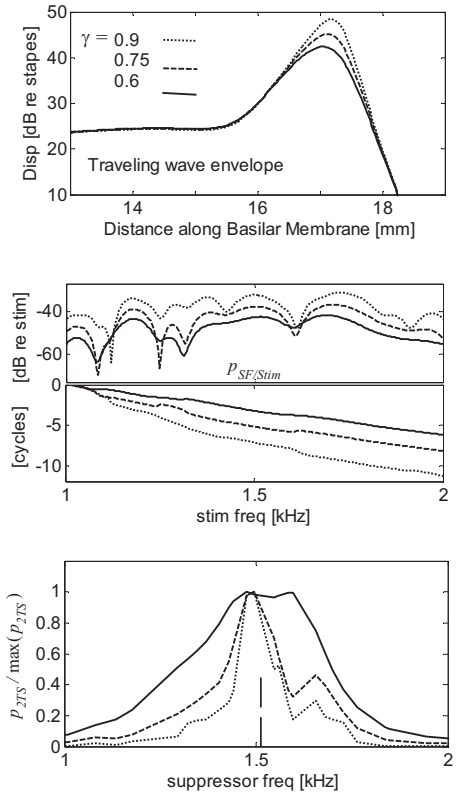


FIG. 5. Results of cochlear model simulations with three values of the cochlear amplifier gain, γ, for a single realization of the impedance inhomogeneities along the basilar membrane. Top panel: the displacement of the basilar membrane (i.e., traveling wave envelope) at a SF of 1.5 kHz for three values of γ, from which the variation of $BW_{BM\text{-}Mech}$ with γ was obtained. Middle panel: the simulated SFOAE probe-frequency sweep for three values of γ, from which the variation of $BW_{SF\text{-}GD}$ with γ was obtained. Bottom panel: the simulated SFOAE 2TS-isoinput tuning characteristics (p_{2TS} normalized with its peak value) for three values of γ, from which the variation of $BW_{SF\text{-}2TS}$ with γ was obtained.

with increased stimulus level similar to that in humans (Lineton and Lutman, 2003a, 2003b); and changes in SFOAE-spectral period with the introduction of a suppressor tone similar to that in humans (Lineton and Lutman, 2003a, 2003b).

In order to investigate whether, according to this model, $BW_{SF\text{-}2TS}$ is correlated with $BW_{BM\text{-}Mech}$ and $BW_{SF\text{-}GD}$, simulations were performed for different values of the cochlear amplifier gain parameter, which ranged from γ=0.6 to 0.9. This gain parameter, γ, controls the amplitude and sharpness of the envelope of the traveling wave, and thus the value of $BW_{BM\text{-}Mech}$, which is defined in the model from the traveling wave envelope using a 6 dB down criterion (i.e., from the difference in the characteristic frequencies at the two points along the basilar membrane where the traveling wave envelope is 6 dB below the peak). These simulations confirmed that in this model, as γ is increased, $BW_{SF\text{-}2TS}$, $BW_{SF\text{-}GD}$, and $BW_{BM\text{-}Mech}$ all increase (Fig. 5).

Further simulations were performed to investigate the random variability in the estimates of $BW_{SF\text{-}2TS}$ and $BW_{SF\text{-}GD}$ that would determine the actual correlation between these two quantities in a population of cochlear models. Thus, for each value of γ in the range from 0.6 to 0.9, an ensemble of models was generated, each model having a different realization of the impedance inhomogeneities that comprise the scattering impedance. From this ensemble, the distributions of $BW_{SF\text{-}2TS}$ and $BW_{SF\text{-}GD}$ can be obtained for each value of $BW_{BM\text{-}Mech}$. [Note that $BW_{BM\text{-}Mech}$ is almost unaffected by the scattering impedance, and can thus be taken as a function of γ only. This is because the scattering impedance is relatively small (1% of the passive damping term) and thus only a relatively small proportion of the energy in the traveling wave is reflected backward.] The resulting means and standard deviations of $BW_{SF\text{-}2TS}$ and $BW_{SF\text{-}GD}$ in the model are shown in Fig. 6.

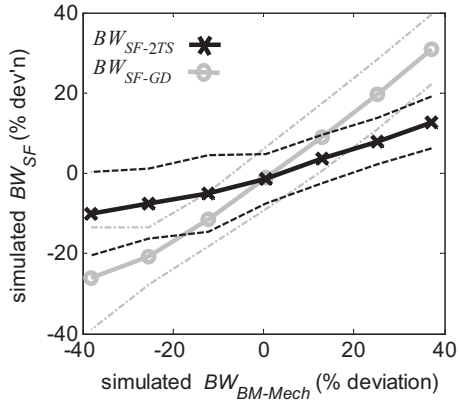


FIG. 6. Results of cochlear model simulations showing the effect of varying $BW_{BM\text{-}Mech}$ (achieved by varying γ) on $BW_{SF\text{-}2TS}$ (black lines) and $BW_{SF\text{-}GD}$ (gray lines). Bandwidths are expressed as a percentage change from the mean at the reference value of γ or 0.75. For each value of γ, simulations were performed for 64 realizations of the impedance inhomogeneities along the basilar membrane to obtain an ensemble of results. The solid lines show the mean effect while the dashed and dotted-dashed lines are at ± 1 standard deviation from the mean for $BW_{SF\text{-}2TS}$ and $BW_{SF\text{-}GD}$, respectively.

What these simulations reveal is that the model predicts some positive correlation between $BW_{SF\text{-}2TS}$ and $BW_{SF\text{-}GD}$ in a population showing some intersubject variation in γ. However, due to the variability in $BW_{SF\text{-}2TS}$ and $BW_{SF\text{-}GD}$ for a given γ, the degree of correlation is expected to be weak. The value of the actual correlation will depend on the degree of variation of γ, and hence $BW_{BM\text{-}Mech}$, in the population. From the results in Fig. 6, the correlations between $BW_{SF\text{-}2TS}$, $BW_{SF\text{-}GD}$, and $BW_{BM\text{-}Mech}$ can be found assuming a Gaussian distribution of $BW_{BM\text{-}Mech}$ in the population with a given standard deviation. Figure 7 shows how the correla-

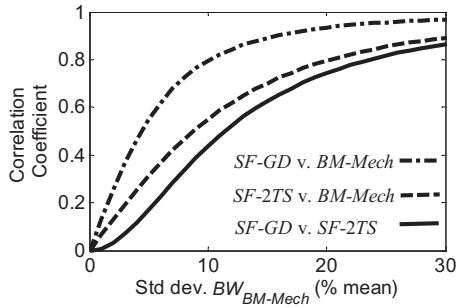


FIG. 7. Simulated variation of the cross correlation between pairwise combinations of the three bandwidths ($BW_{BM\text{-}Mech}$, $BW_{SF\text{-}GD}$, and $BW_{SF\text{-}2TS}$) as a function of the assumed variability in $BW_{BM\text{-}Mech}$ (arising from the variability in γ) in the population. The curves are obtained from the estimated means and standard deviations for the cochlear model shown in Fig. 6.

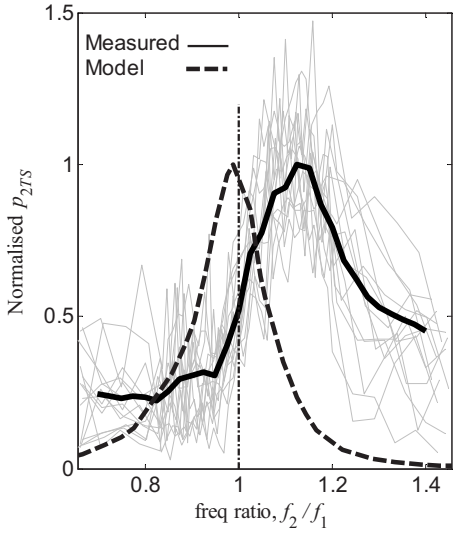


FIG. 8. Comparison of measured 2TS-isoinput tuning characteristics with the results of the cochlear model. The value of p_{2TS} was normalized with the maximum of the fitted curve (Eq. (14)). The measured isoinput tuning characteristics (thin solid lines) were averaged across the 16 subjects to give the average (thick solid line). The isoinput tuning characteristics from the cochlear model simulations were averaged across the 64 realizations (thick dotted-dashed line).

tion between $BW_{SF\text{-}2TS}$ and $BW_{SF\text{-}GD}$ in the population increases as the standard deviation of $BW_{BM\text{-}Mech}$ in the population increases.

These model simulations are not expected to be quantitatively accurate predictions of measured results in the human population. After all, the models are hugely simplified caricatures of real cochleas, and furthermore, the only source of variability considered in the population is due to γ. However, what these simulations do suggest is that any correlation between $BW_{SF\text{-}2TS}$ and $BW_{SF\text{-}GD}$ may be rather weak, and only detectable using large samples. For example, if the standard deviation of $BW_{BM\text{-}Mech}$ were 10% of the mean, then the expected correlation between $BW_{SF\text{-}2TS}$ and $BW_{SF\text{-}GD}$ would be <0.5, even in this highly idealized model population. A rough idea of the intersubject variability of the mechanical tuning can be gained from psychoacoustic measures of auditory filter bandwidth, where the standard deviation in a sample of normally hearing adults is typically between 8% and 20% of the mean (e.g., Wright, 1996). Figure 7 also shows that, in the model, $BW_{SF\text{-}GD}$ is a better predictor of $BW_{BM\text{-}Mech}$ than is $BW_{SF\text{-}2TS}$, though neither is a particularly good predictor.

As an aside, while in general the simulated SFOAEs from the cochlear models show several of the nonlinear features seen in measured SFOAEs (Lineton and Lutman, 2003a, 2003b), one glaring discrepancy was found in the present study: The maximum 2TS in the model occurred when the frequency of the suppressor tone equaled that of the probe tone (i.e., $f_2=f_1$). In contrast, in the measured 2TS-isoinput tuning characteristics, the maximum suppression occurred when f_2 was between $1.1f_1$ and $1.2f_1$ (Fig. 8). This indicates that, for reasons still unknown, the formulation of the nonlinear mechanics in this particular model does not accurately capture this key feature of 2TS of SFOAEs.

Brass, D., and Kemp, D. T. (**1993**). "Suppression of stimulus frequency otoacoustic emissions," J. Acoust. Soc. Am. **93**, 920–939.

Jedrzejczak, W. W., Blinowska, K. J., Konopka, W., Grzanka, A., and Durka, P. J. (**2004**). "Identification of otoacoustic emissions components by means of adaptive approximations," J. Acoust. Soc. Am. **115**, 2148–2158.

Kanis, L. J., and de Boer, E. (**1993**). "Self-suppression in a locally active non-linear model of the cochlea: A quasilinear approach," J. Acoust. Soc. Am. **94**, 3199–3206.

Kanis, L. J., and de Boer, E. (**1994**). "Two-tone suppression in a locally active nonlinear model of the cochlea," J. Acoust. Soc. Am. **96**, 2156–2165.

Keefe, D. H., Ellison, J. C., Fitzpatrick, D. F., and Gorga, M. P. (**2008**). "Two-tone suppression of stimulus frequency otoacoustic emissions," J. Acoust. Soc. Am. **123**, 1479–1494.

Kemp, D. T., Brass, D. N., and Souter, M. (**1990**). in *The Mechanics and Biophysics of Hearing*, edited by P. Dallos, C. D. Geisler, J. W. Matthew, M. A. Ruggero, and C. R. Steele (Springer-Verlag, Berlin), pp. 202–209.

Kemp, D. T., and Brown, A. M. (**1983**). "A comparison of mechanical nonlinearities in the cochleae of man and gerbil from ear canal measurements," in *Hearing—Physiological Bases and Psychophysics*, edited by R. Klinke and R. Hartmann (Springer, Berlin), pp. 82–88.

Kemp, D. T., and Chum, R. A. (**1980**). "Observations on the generation mechanism of stimulus frequency acoustic emissions—Two tone suppression," in *Psychophysical, Physiological and Behavioural Studies in Hearing*, *Proceedings of the Fifth International Symposium on Hearing*, edited by G. van den Brink and F. A. Bilsen (Delft University Press, The Netherlands), pp. 34–42.

Konrad-Martin, D., and Keefe, D. H. (**2005**). "Transient-evoked stimulus-frequency and distortion-product otoacoustic emissions in normal and impaired ears," J. Acoust. Soc. Am. **117**, 3799–3815.

Kringlebotn, M. (**1988**). "Network model for the human middle ear," Scand. Audiol. **17**, 75–85.

Lineton, B. (**2001**). "Testing a model of stimulus frequency otoacoustic emissions," Ph.D. thesis, University of Southampton, United Kingdom.

Lineton, B., and Lutman, M. E. (**2003a**). "Modeling the effect of suppression on the periodicity of stimulus frequency otoacoustic emissions," J. Acoust. Soc. Am. **114**, 859–870.

Lineton, B., and Lutman, M. E. (**2003b**). "A parametric model of the spectral periodicity of stimulus frequency otoacoustic emissions," J. Acoust. Soc. Am. **114**, 871–882.

Neely, S. T., Norton, S. J., Gorga, M. P., and Jesteadt, W. (**1988**). "Latency of auditory brain-stem responses and otoacoustic emissions using tone-burst stimuli," J. Acoust. Soc. Am. **83**, 652–656.

Norton, S. J., and Neely, S. T. (**1987**). "Tone-burst-evoked otoacoustic emissions from normal-hearing subjects," J. Acoust. Soc. Am. **81**, 1860–1872.

Robles, L., and Ruggero, M. A. (**2001**). "Mechanics of the mammalian cochlea," Physiol. Rev. **81**, 1305–1352.

Schairer, K. S., Ellison, J. C., Fitzpatrick, D., and Keefe, D. H. (**2006**). "Use of stimulus-frequency otoacoustic emission latency and level to investigate cochlear mechanics in human ears," J. Acoust. Soc. Am. **120**, 901–914.

Shera, C. A., and Guinan, J. J., Jr. (**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., Guinan, J. J., Jr., and Oxenham, A. J. (**2002**). "Revised estimates of human cochlear tuning from otoacoustic and behavioural measurements," Proc. Natl. Acad. Sci. U.S.A. **99**, 3318–3323.

Shera, C. A., and Zweig, G., (**1993a**). in *Biophysics of Hair-Cell Sensory Systems*, edited by H. Duifhuis, J. W. Horst, P. van Dijk, and S. M. van Netten (World Scientific, Singapore), pp. 54–63.

Shera, C. A., and Zweig, G. (**1993b**). "Noninvasive measurement of the cochlear traveling-wave ratio," J. Acoust. Soc. Am. **93**, 3333–3352.

Tognola, G., Grandori, F., and Ravazzani, P. (**1997**). "Time-frequency distributions of click-evoked otoacoustic emissions," Hear. Res. **106**, 112–122.

Wright, B. (**1996**). "Auditory filter asymmetry at 2000 Hz in 80 normal-hearing ears," J. Acoust. Soc. Am. **100**, 1717–1721.

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.

Stochastic properties of auditory brainstem coincidence detectors in binaural perception

Ram Krips and Miriam Furst[a)]
Department of Electrical Engineering-Systems, Faculty of Engineering, Tel-Aviv University, Tel-Aviv 69978, Israel

(Received 25 May 2008; revised 10 December 2008; accepted 11 December 2008)

In the mammalian auditory brainstem, two types of coincidence detector cells are involved in binaural localization: excitatory-excitatory (EE) and excitatory-inhibitory (EI). Using statistics derived from EE and EI spike trains, binaural discrimination abilities of single tones were predicted. The minimum audible angle (MAA), as well as the just noticeable difference of interaural time delay (ITD) and interaural level difference (ILD) were analytically derived for both EE and EI cells on the basis of two possible neural coding patterns, rate coding that ignores a spike's timing information and all-information coding (AIN), which considers a spike's timing occurrences. Simulation results for levels below saturation were qualitatively compared to experimental data, which yielded the following conclusions: (1) ITD is primarily estimated by EE cells with AIN coding when the ipsilateral auditory input exhibits phase delay between 40° and 65°. (2) In ILD, both AIN and rate coding provide identical performances. It is most likely that ILD is primarily estimated by EI cells according to rate coding, and for ILD the information derived from the spikes' timing is redundant. (3) For MAA estimation, the derivation should take into account ambiguous directions of a source signal in addition to its true value.
 [DOI: 10.1121/1.3068446]

PACS number(s): 43.64.Bt, 43.66.Pn, 43.66.Ba, 43.66.Qp [BLM] Pages: 1567–1583

I. INTRODUCTION

Three main acoustic cues are known to contribute to sound localization: interaural time difference (ITD), interaural level difference (ILD), and spectral cues (e.g., review Blauert, 1997; Searle *et al.*, 1976). Systematic psychoacoustical experiments have determined the ability of human beings to spatially localize sounds as well as the capability for detecting acoustical cues. In such experiments, for example, subjects were introduced with two discrete and temporally nonoverlapping sounds from different positions; subjects were required to indicate whether the second sound was to the left or right of the first. Since the mid-20th century, just noticeable difference (JND) of ITD and ILD and the minimum audible angle (MAA) of localization were obtained (Klumpp and Eady, 1956; Zwislocki and Feldman, 1956; Mills, 1958; Mills, 1960; Grantham, 1986).

Most studies that tried to model the ability to lateralize a perceived signal were based on the existence of coincidence detectors (CDs) (Jeffress, 1948). The CD based model that Jeffress developed was founded on three main assumptions: bilateral time-locked (phase-locked) inputs into the ITD-processing system, coincidence detection by ITD detector neurons, and an arrangement of delay lines to adjust CD neurons to different preferred ITDs, creating azimuthally a topographic representation of space.

Physiological studies have shown that many cells in the auditory midbrain are sensitive to the direction of sound sources (King and Palmer, 1983; Knudsen, 1982; Knudsen and Konishi, 1978; Semple *et al.*, 1983; see Irvine, 1992, for a review).

The initial processing of interaural timing cues occurs in the medial superior olive (MSO) in which neurons tuned for low-frequency sounds are relatively overrepresented in comparison to neurons tuned to high frequencies. MSO neurons receive excitatory input from the large spherical bushy cells of both (right and left) antro ventral cochlear nucleus (AVCN), which preserve and even enhance the timing accuracy seen in the auditory nerve providing exquisitely timed inputs (Warr, 1966, 1969; Palmer *et al.*, 2002; Joris *et al.*, 1994).

The initial processing of ILDs occurs in the lateral superior olive (LSO) where high-frequency neurons are relatively over-represented compared to low-frequency neurons. The small spherical bushy cells of the ipsilateral AVCN form excitatory synapses on LSO principal neurons (Warr, 1966; Glendenning *et al.*, 1985; Cant, 1991). Additionally, LSO neurons receive inhibitory inputs from neurons in the ipsilateral medial nucleus of the trapezoid body (MNTB), which, in turn, receive excitatory input from the globular bushy cells of the contralateral cochlear nucleus (Glendenning *et al.*, 1985; Cant, 1991; Warr, 1972; Boudreau and Tsuchitani, 1968). The pathway from VCN to MNTB is characterized by synapses producing secure short-latency responses and, therefore, near coincident arrival at the LSO of the ipsilateral excitation and the contralateral inhibition. Neurons in the LSO are sensitive to the balance of intensity at the ears because the excitation due to ipsilateral sounds is reduced by

[a)]Author to whom correspondence should be addressed. Electronic mail: mira@eng.tau.ac.il

increasing levels of contralateral sounds (Boudreau and Tsuchitani, 1970; Guinan *et al.*, 1972; Caird and Klinke, 1983; Caspary and Finlayson, 1991).

Cells that are sensitive to localization cues such as ITD and ILD were identified in both the superior olivary complex (SOC) and the inferior colliculus (IC) (Goldberg and Brown, 1969; Rose *et al.*, 1967; Irvine, 1986, 1992; Delgutte *et al.*, 1999; Fitzpatrick *et al.*, 2000). There are mainly two types of cells that are sensitive to binaural cues: excitatory-excitatory (EE) cells and excitatory-inhibitory (EI) cells (Rose *et al.*, 1967; Goldberg and Brown, 1969; Yin *et al.*, 1987; Yin and Chan, 1990; Joris and Yin, 1995; Joris, 1996; Batra *et al.*, 1997; Palmer *et al.*, 1997; McAlpine *et al.*, 1998; Brand *et al.*, 2002).

Mammalian EE and EI cells are natural candidates for CDs in a "Jeffress-type" system (Moushegian *et al.*, 1975; Yin and Chan, 1990; Joris *et al.*, 1998). The evidence to support this view includes the connection patterns with excitatory inputs from both sides, the cyclic nature of the ITD sensitivity when tested with pure tones, and the fact that the best ITD could be predicted on the basis of the time delay of the response to the monaural inputs from both ears (Goldberg and Brown, 1969; Crow *et al.*, 1978; Yin and Chan, 1990; Spitzer and Semple, 1995). Moreover, results from many human psychophysical studies can be explained by the concept of binaural cross correlation, as indicated by the Jeffress model (Stern and Trahiotis, 1995; Palmer *et al.*, 2002). Most published models for ITD are based on EE cells (review by Colburn, 1996). Yet, recent models account for some aspects of the binaural ITD phenomena with EI cells (e.g., McAlpine *et al.*, 1998; Colburn *et al.*, 2004). ILD processing is frequently considered and modeled by EI mechanisms (Lindemann, 1986; Breebaart, 2001; Yue and Johnson, 1997; Park *et al.*, 2004).

In modeling experimental JND, it is assumed that the neural system estimates the measured parameter (α), and if the brain is acting as an optimal estimator, then the JND is equal to the standard deviation of the estimator (Siebert, 1968). Moreover, if the estimator is unbiased, its variance is bounded by the Cramer–Rao lower bound (CRLB).

A CRLB has been derived for an auditory nerve fiber that behaves as a nonhomogeneous Poisson process (NHPP) where the timing of the spikes was considered (Bar David, 1969). Even though the CRLB derivation yielded significantly lower values than human experimental performance (Siebert, 1968; Colburn, 1973; Heinz *et al.*, 2001), it was generally found to be very useful in indicating constraints in models of the auditory system. For example, calculation of frequency JND as a function of the stimulus frequency and signal duration indicated that rate coding is insufficient for explaining experimental data (Heinz *et al.*, 2001).

Some previous studies predicted binaural psychoacoustical properties by deriving lower bounds of the neural activity. In some studies, CRLBs of the two monaural auditory nerves followed by theoretical coincident detectors were derived (Colburn, 1973; Stern and Colburn, 1978; Huettel and Collins, 2004; Cohen *et al.*, 2004). In other studies, such as Yue and Johnson (1997), specific LSO cells' response was presented as a function of the binaural stimuli. Those cells are mainly response to high frequencies where there is no phase synchronization. An optimal decision model was applied to the LSO's cells response that yielded a biased bound.

We have recently proven that the outputs of both EE and EI cells, whose inputs behave as NHPP, also behave as NHPP, and their corresponding instantaneous rates (IRs) were analytically derived (Krips, 2008). In the present paper, we are using our recent theoretical result in order to predict psychoacoustical binaural discrimination abilities. None of the previous studies could determine analytically the lower bounds from the neural activity of EE and EI cells. These analytic derivations allow us to compare EE and EI performances in detecting binaural cues according to both types of coding, rate coding and all-information coding (AIN).

II. THEORETICAL METHODOLOGY

A. Derivation of JND lower bounds on the basis of neural activity

In a psychoacoustical experiment that was designed to detect the minimal perceived difference of a parameter α, JND(α) is obtained. In modeling such an experiment, we assume that the brain acts as an unbiased optimal estimator, which yields for $\alpha=\alpha^*$ the estimator $\hat{\alpha}$ that satisfies the following conditions (Hartmann and Rakerd, 1989):

$$E[\hat{\alpha}] = \alpha^*,$$

$$\mathrm{JND}(\alpha^*) \approx \sigma(\hat{\alpha}) \geqslant \mathrm{CRLB}(\alpha^*), \tag{1}$$

where CRLB(α^*) is the CRLB at $\alpha=\alpha^*$.

When the estimation is based on a neural activity that behaves as NHPP, there are two possible ways to analyze the performance. The first way is referred to as "rate coding," which means that the performance is analyzed on the basis of the number of spikes that yields a stimulus during an interval of length T s. The probability density function (pdf) is obtained by

$$P_{\mathrm{rate}}(N(0,T)=n) = \frac{1}{n!}\left[\int_0^T \lambda(t,\alpha)dt\right]^n \times \exp\left\{-\int_0^T \lambda(t,\alpha)dt\right\}, \tag{2}$$

where $\lambda(t,\alpha)$ is the IR of the nerve fiber. The resulting CRLB is obtained by (Snyder and Miller, 1991)

$$\mathrm{CRLB}_{\mathrm{rate}}(\alpha^*) = \frac{1}{\sqrt{\frac{T}{\bar{\lambda}(\alpha^*)}\left[\left.\frac{\partial\bar{\lambda}(\alpha)}{\partial\alpha}\right|_{\alpha=\alpha^*}\right]^2}}, \tag{3}$$

where $\bar{\lambda}(\alpha)=(1/T)\int_0^T\lambda(t,\alpha)dt$ is the average rate.

For AIN the timing of discharge spikes is considered in addition to the number of spikes in the interval, which yields the following pdf:

$$P_{\mathrm{AIN}}(t_1,\ldots,t_n)=\frac{1}{n!}\prod_{k=1}^{n}\lambda(t_k,\alpha)\exp\left\{-\int_0^T\lambda(t,\alpha)dt\right\}, \tag{4}$$

where $t_1,\ldots,t_n$ are the discharge times. The resulting CRLB was derived by Bar David (1969), which yields

$$\mathrm{CRLB}_{\mathrm{AIN}}(\alpha^*)=\frac{1}{\sqrt{\int_0^T\frac{1}{\lambda(t,\alpha^*)}\left[\left.\frac{\partial\lambda(t,\alpha)}{\partial\alpha}\right|_{\alpha=\alpha^*}\right]^2dt}}. \tag{5}$$

CRLB is useful when the estimator uses information from the vicinity of the true value α^*. However, there are cases where the estimator might use other values besides the true one. For example, a subject will similarly perceive a continuous 2 kHz tone that originated from his right side, left side, or the front of his head (e.g., Kuhn, 1987). Thus, when the brain estimates the direction of such a tone, all these possibilities should be considered. CRLB in such cases is inadequate (Krips and Furst, 2006a).

Barankin (1949) developed a lower bound that takes into account different possible values of the estimated parameter other than those located in the proximity of the true one (ambiguous locations). If the true value is α^* and there are L other possible values, $\alpha_1,\ldots,\alpha_L\neq\alpha^*$, then the Barankin lower bound (BLB) was derived (McAulay and Hofstetter, 1971) and is obtained by

$$\begin{aligned}\sigma(\hat{\alpha})&\geqslant\mathrm{BLB}(\alpha^*)\\&=\sqrt{[\mathrm{CRLB}(\alpha^*)]^2+\underline{\Gamma}(\alpha^*)\Delta^{-1}(\alpha^*)\underline{\Gamma}^T(\alpha^*)},\end{aligned} \tag{6}$$

where $\underline{\Gamma}(\alpha^*)=\underline{\alpha}-\mathrm{CRLB}^2(\alpha^*)\underline{A}$ and $\Delta(\alpha^*)=B-\underline{A}^T\mathrm{CRLB}^2(\alpha^*)\underline{A}$. $\mathbf{B}$ is a symmetric matrix whose size is $L\times L$ and the $\underline{\alpha}$ and $\underline{A}$ are vectors whose size are L ($\underline{\alpha}=[\alpha_1,\ldots,\alpha_L]$, $\underline{A}=[A_1,\ldots,A_L]$). The vector $\underline{A}$ and the matrix $\mathbf{B}$ are derived in Appendix A (Krips, 2008) for both rate coding [Eq. (2)] and AIN [Eq. (4)], which yields

$$\begin{aligned}A_i^{\mathrm{rate}}&=T\left.\frac{\partial\bar{\lambda}(\alpha)}{\partial\alpha}\right|_{\alpha=\alpha^*}\left[\frac{\bar{\lambda}(\alpha_i)}{\bar{\lambda}(\alpha^*)}-1\right],\quad i=1,\ldots,L,\\A_i^{\mathrm{AIN}}&=\int_0^T\left.\frac{\partial\lambda(t,\alpha)}{\partial\alpha}\right|_{\alpha=\alpha^*}\left[\frac{\lambda(t,\alpha_i)}{\lambda(t,\alpha^*)}-1\right]dt,\quad i=1,\ldots,L\end{aligned} \tag{7}$$

and

$$\begin{aligned}B_{i,j}^{\mathrm{rate}}&=\exp\left\{-T\left[\bar{\lambda}(\alpha_i)+\bar{\lambda}(\alpha_j)-\bar{\lambda}(\alpha^*)-\frac{\bar{\lambda}(\alpha_i)\bar{\lambda}(\alpha_j)}{\bar{\lambda}(\alpha^*)}\right]\right\},\\B_{i,j}^{\mathrm{AIN}}&=\exp\left\{-\int_0^T\left[\lambda(t,\alpha_i)+\lambda(t,\alpha_j)-\lambda(t,\alpha^*)-\frac{\lambda(t,\alpha_i)\lambda(t,\alpha_j)}{\lambda(t,\alpha^*)}\right]dt\right\}\end{aligned} \tag{8}$$

for $i,j=1,\ldots,L$.

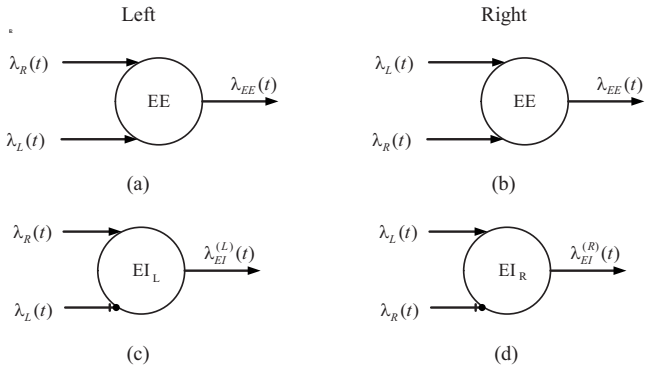


FIG. 1. Schematic description of (a) left SOC EE cell, (b) right SOC EE cell, (c) left SOC EI cell, and (d) right SOC EI cell. The inputs are from the right and left auditory nerves that are indicated by their IRs $\lambda_R(t)$ and $\lambda_L(t)$, respectively. Excitatory input is indicated by an arrow, and an inhibitory input by a black circle.

B. Derivation of EE and EI cell instantaneous rates

In this paper we investigate the ability of CD cells to predict psychoacoustical binaural abilities. We particularly refer to two typical CD cells, EE and EI, that are found in both the right and left mammalian SOCs (see Fig. 1). Each input represents a single primarylike auditory nerve that statistically behaves as NHPP.

EE cells receive two inputs [Figs. 1(a) and 1(b)] that arrive from the two ears. The IRs of the right and left inputs are denoted by $\lambda_R(t,\alpha)$ and $\lambda_L(t,\alpha)$, and their correspondent refractory periods are represented by τ_R and τ_L, respectively. We showed that if the two inputs behave as NHPP and if the time interval (Δ) in which the two inputs can interact satisfies the condition $\Delta\ll\min\{\tau_R,\tau_L\}$, then EE output also behaves as NHPP and its IR is obtained by

$$\begin{aligned}\lambda_{\mathrm{EE}}(t,\alpha)&=\lambda_L(t,\alpha)\int_{t-\Delta}^{t}\lambda_R(t',\alpha)dt'\\&\quad+\lambda_R(t,\alpha)\int_{t-\Delta}^{t}\lambda_L(t',\alpha)dt'.\end{aligned} \tag{9}$$

Since both right and left EE cells receive similar inputs, their output IR is also identical, i.e.,

$$\lambda_{\mathrm{EE}}^{(R)}(t,\alpha)=\lambda_{\mathrm{EE}}^{(L)}(t,\alpha)=\lambda_{\mathrm{EE}}(t,\alpha). \tag{10}$$

A possible coincidence window is $\Delta=20$ μs (Agmon-Snir *et al.* 1998). It was previously used in theoretical models (e.g., Colburn, 1973; Heinz *et al.*, 2001; Cohen, *et al.*, 2004). This value satisfies the condition $\Delta\ll\min\{\tau_R,\tau_L\}$ since the refractory period at the auditory nerve is of the order of 500 μs to 1 ms (Miller *et al.*, 2001; Dynes, 1996; Bruce *et al.*, 1999; Brown and Abbas, 1990).

EI cells receive two types of inputs, an excitatory input and an inhibitory input. The EI in the right SOC [Fig. 1(d)] receives an excitatory input from the left side and an inhibitory input from the right side, which yield

$$\lambda_{\mathrm{EI}}^{(R)}(t,\alpha)=\lambda_L(t,\alpha)\left(1-\int_{t-\Delta}^{t}\lambda_R(t',\alpha)dt'\right). \tag{11}$$

On the other hand, EI cells in the left SOC receive the antisymmetric inputs, i.e., an excitatory input from the right side and an inhibitory input from the left side [Fig. 1(c)], which yield

$$\lambda_{\mathrm{EI}}^{(L)}(t,\alpha)=\lambda_R(t,\alpha)\left[1-\int_{t-\Delta}^{t}\lambda_L(t',\alpha)dt'\right]. \tag{12}$$

A possible coincidence window is $\Delta=200\ \mu$s (Park, 1998). It is ten times longer than found in EE cells. However it satisfies the condition $\Delta<\min\{\tau_R,\tau_L\}$, which guarantees that EI cells behave as NHPP if their inputs also behave as NHPP.

C. Prediction of JND on the basis of both SOCs

In order to predict the psychoacoustical results, for every type of cell EE or EI, both sides are combined. This yields bounds of each side that provides

$$\begin{aligned}\mathrm{JND}_{\mathrm{EE}}(\alpha^*)&=1/\sqrt{[\mathrm{CRLB}_{\mathrm{EE}}^{(R)}(\alpha^*)]^{-2}+[\mathrm{CRLB}_{\mathrm{EE}}^{(L)}(\alpha^*)]^{-2}},\\ \mathrm{JND}_{\mathrm{EI}}(\alpha^*)&=1/\sqrt{[\mathrm{CRLB}_{\mathrm{EI}}^{(R)}(\alpha^*)]^{-2}+[\mathrm{CRLB}_{\mathrm{EI}}^{(L)}(\alpha^*)]^{-2}},\end{aligned} \tag{13}$$

where $\mathrm{CRLB}_{\mathrm{EE}}^{(R)}(\alpha^*)$, $\mathrm{CRLB}_{\mathrm{EI}}^{(R)}(\alpha^*)$, $\mathrm{CRLB}_{\mathrm{EE}}^{(L)}(\alpha^*)$, and $\mathrm{CRLB}_{\mathrm{EI}}^{(L)}(\alpha^*)$ are the CRLBs obtained by EE and EI cells in the right and left SOCs, respectively.

D. Derivation scheme for binaural simple tones

In this paper we test localization abilities that are evoked by audio signal $S(t,\theta)$ that enters both ears from direction θ. The incoming sounds to each ear are transformed in ways which depend on the shape and size of the head, torso, and the outer ears' pinna. These anatomical features are known as the head related transfer function (HRTF) that can be measured and synthesized in the form of linear time-invariant filters. Therefore, the signals that are conveyed to the left and right cochleae are

$$\begin{aligned}S_L(t,\theta)&=S(t,\theta)*\mathrm{HRTF}_L(t,\theta),\\ S_R(t,\theta)&=S(t,\theta)*\mathrm{HRTF}_R(t,\theta),\end{aligned} \tag{14}$$

where $*$ represents a convolution, and $\mathrm{HRTF}_L(t,\theta)$ and $\mathrm{HRTF}_R(t,\theta)$ are the left and right HRTFs.

This paper presents only the tests of signals that are composed of simple tones, i.e., $S(t)=A\sin(2\pi ft)$. The effects of the HRTF on such a signal are a phase shift and amplitude alteration, which yield

$$\begin{aligned}S_L(t,\theta)&=A_L(\theta)\sin(2\pi ft+\varphi_L(\theta)),\\ S_R(t,\theta)&=A_R(\theta)\sin(2\pi ft+\varphi_R(\theta)).\end{aligned} \tag{15}$$

Therefore the resulting interaural difference is an interaural phase difference (IPD), which is obtained by

$$\mathrm{IPD}(\theta)=\varphi_R(\theta)-\varphi_L(\theta), \tag{16}$$

which corresponds to ITD by

$$\mathrm{ITD}(\theta)=\frac{\mathrm{IPD}(\theta)}{2\pi f}, \tag{17}$$

and the ILD in dB is given by

$$\mathrm{ILD}(\theta)=20\log_{10}\left(\frac{A_R(\theta)}{A_L(\theta)}\right)=20\log_{10}(\delta), \tag{18}$$

where $\delta=A_R(\theta)/A_L(\theta)$.

The IRs of EE and EI cells [Eqs. (9), (11), and (12)] depend on the input IRs $\lambda_L(t)$ and $\lambda_R(t)$. The input fibers to the SOC's EE and EI cells arrive from the CN. We assume that those are primarylike nerves and that their statistical properties are similar to those of the auditory nerve.

The auditory nerve IR for simple tone stimuli is commonly expressed with exponential functions (e.g., Siebert, 1968; Colburn, 1973; Heinz *et al.*, 2001; Cohen *et al.*, 2004) as follows:

$$\begin{aligned}\lambda_L(t)&=\Lambda(S_L(t,\theta))\exp\{B(S_L(t,\theta))\sin(2\pi ft+\varphi_L(\theta)+\Theta_L(f))\},\\ \lambda_R(t)&=\Lambda(S_R(t,\theta))\exp\{B(S_R(t,\theta))\sin(2\pi ft+\varphi_R(\theta)+\Theta_R(f))\}.\end{aligned} \tag{19}$$

Generally, $\Lambda(S(t,\theta))$ is a function of the stimulus level. It has a minimum value that is equal to the fiber's spontaneous rate and a maximum value that represents the fiber's saturation rate. For stimuli whose levels are in the midrange ($20\leqslant A\leqslant 50$ dB SPL), $\Lambda(S(t,\theta))$ is proportional to the stimulus level, i.e., $\Lambda(A\sin(2\pi ft))=\gamma A$. The proportional coefficient (γ) is different for every fiber as determined by the location along the cochlear partition that the fiber innervates. In this paper we refer only to stimuli in the midrange levels, where the IR is proportional to the stimuli level. The function $B(S(t,\theta))$ governs the synchronization of the fiber response, which decreases with the increase in both frequency and level of simple tone stimuli. In this paper we refer only to the dependence of the synchronization on frequency. The synchronization is typically measured by its vector strength (VS), which is obtained by

$$\mathrm{VS}(f)=\frac{\Upsilon(f)}{\Upsilon(0)}, \tag{20}$$

where $\Upsilon(f)$ is the absolute value of the IR's Fourier transform. VS data (Palmer and Russell, 1986; Johnson, 1980; Rose *et al.*, 1967) can be modeled by sigmoid functions such as

$$\mathrm{VS}(f)=2\frac{e^{-\beta\cdot f}}{1+e^{-\beta\cdot f}}. \tag{21}$$

A family of VS function is plotted as a function of frequency in Fig. 2. Each curve was obtained with a different β. The leftmost (unsynchronized) curve in Fig. 2 was obtained with $\beta=0.0007$ and corresponds to species where their auditory nerve fibers start losing their synchrony below 1 kHz (Liu *et al.*, 2006). There are other species that lose their synchrony at higher frequencies; e.g., cats and squirrel monkeys start losing synchrony above 3 kHz (Johnson, 1980; Palmer and Russell, 1986; Rose *et al.*, 1967; Reyes *et al.*, 1996). Thus their VS can be fitted with higher values of β (for example, the smallest value used was of $\beta=0.0001$).

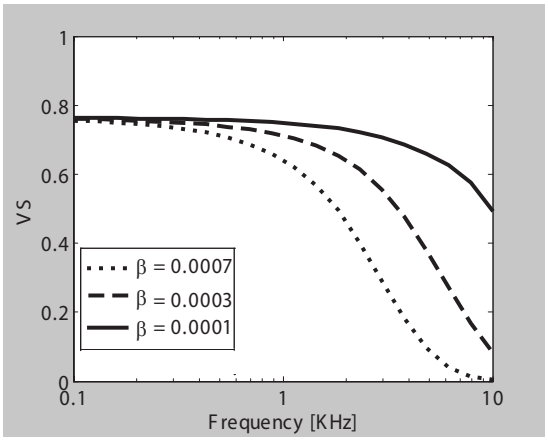


FIG. 2. VS as a function of frequency. Each curve represents a different function [Eq. (21)] as indicated by the parameter β.

III. SIMULATION RESULTS

A. Interaural phase difference

In order to evaluate the ability of EE and EI cells to predict properties that are related to ITD or IPD, the dependence on both the stimulus direction and level is ignored. Therefore,

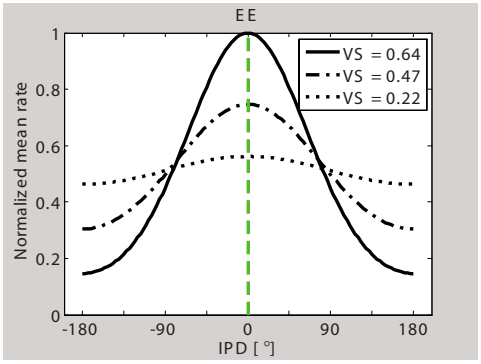
$$S_L(t) = \sin(2\pi f t), \qquad S_R(t) = \sin(2\pi f t + \mathrm{IPD}). \tag{22}$$

The IRs $\lambda_{EE}^{(R)}(t,\mathrm{IPD})$, $\lambda_{EE}^{(L)}(t,\mathrm{IPD})$, $\lambda_{EI}^{(R)}(t,\mathrm{IPD})$, and $\lambda_{EI}^{(L)}(t,\mathrm{IPD})$ were obtained by substituting Eq. (22) in Eq. (19). Initially we assumed that $\Theta_L(f)=\Theta_R(f)=0$.

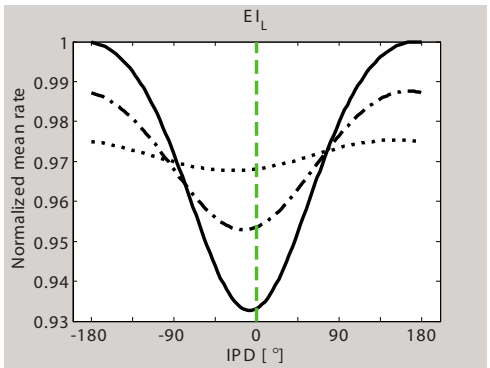
Figure 3 represents the normalized average rates $\bar{\lambda}_{EE}$ and $\bar{\lambda}_{EI}$ as a function of IPD for different input frequencies. The input rates were obtained according to Eq. (19) by substituting $\Lambda=200$ spikes/s. The normalization was obtained to the signal period (i.e., for every frequency f, $T=1/f$). We chose to indicate the VS of the fiber rather than its frequency because IR dependence on frequency is uniquely determined by a nerve's synchronization functions $B(S(t,\theta))$ [Eq. (19)]. Moreover, different synchronization functions can yield different frequencies for the same VS, as shown in Fig. 2.

Since both right and left EE cells are identical, the normalized average rate of the output of an EE cell is presented in a single figure [Fig. 3(a)] as a function of IPD. It yields a maximum when both inputs are in phase (IPD=0°) and decreases with an increase in |IPD|. In general, a similar behavior was obtained for all input frequencies. However, for large values of VS, which correspond to low frequencies, a distinct peak is observed; for low VS values that correspond to high frequencies, the mean rate as a function of IPD reveals shallower curves.

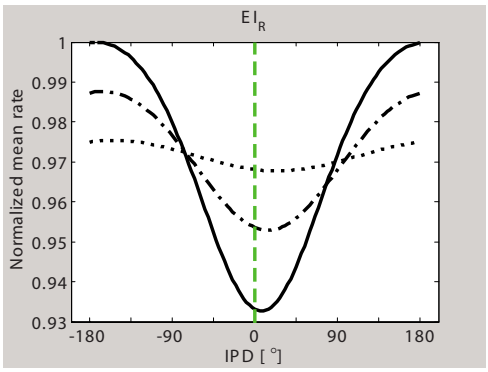
In the simulation shown in Fig. 3(a), we used a coincidence window $\Delta=20$ μs. Increasing Δ will cause steeper curves as a function of IPD, but the general behavior holds. It is therefore clear from Fig. 3(a) that EE cells describe an



(a)



(b)



(c)

FIG. 3. (Color online) Normalized mean spiking rate as a function of IPD for different frequencies as indicated by their VS. Each panel represents a different cell: (a) EE, (b) left EI, and (c) right EI cells. Derivation was obtained with the following parameters: $\Lambda=200$ spikes/s for all inputs, $\Delta=20$ μs for EE cell, and $\Delta=200$ μs for EI cell.

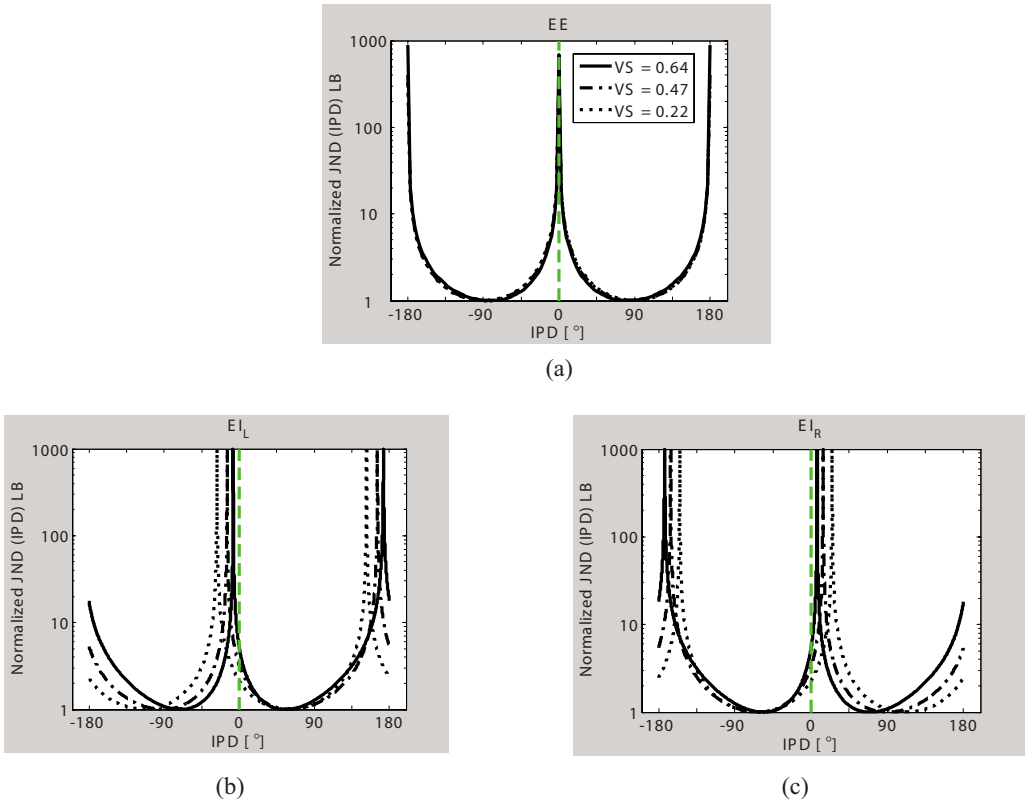


FIG. 4. (Color online) Normalized lower bound of JND of IPD obtained by rate coding as a function of IPD derived by (a) EE cell, (b) left EI cell, and (c) right EI cell. Each curve represents a different input frequency as indicated by its VS.

efficient correlator mechanism between the two ears, especially for frequencies with high VS (low frequencies).

The left and right EI cells are antisymmetric, as can be seen in Fig. 1. The inhibitory input arrives from the ipsilateral ear, while the excitatory input arrives from the contralateral ear. $\bar{\lambda}_{\mathrm{EI}}^{(R)}$ yields a minimum at $\phi=\mathrm{IPD}>0$ [Fig. 3(c)], while $\bar{\lambda}_{\mathrm{EI}}^{(L)}$ reveals a minimum at $-\phi$ [Fig. 3(b)]. In both left and right cases, the minimum rate was obtained when the inhibitory input preceded the excitatory input by $\phi°$. In the simulation for EI cells, we used a coincidence window $\Delta=200\ \mu\mathrm{s}$.

The IRs $\lambda_{\mathrm{EE}}^{(R)}(t,\mathrm{IPD})$, $\lambda_{\mathrm{EE}}^{(L)}(t,\mathrm{IPD})$, $\lambda_{\mathrm{EI}}^{(R)}(t,\mathrm{IPD})$, and $\lambda_{\mathrm{EI}}^{(L)}(t,\mathrm{IPD})$ were used to calculate the CRLB of JND (IPD) according to the rate coding [Eq. (3)] and the AIN coding [Eq. (5)].

Figures 4 and 5 present CRLB (IPD) as a function of IPD, according to both rate coding and AIN, respectively. Since $\lambda_{\mathrm{EE}}^{(R)}(t,\mathrm{IPD})=\lambda_{\mathrm{EE}}^{(L)}(t,\mathrm{IPD})$, both right and left EE cells provide identical behaviors that correspond to EE in Figs. 4(a) and 5(a). Right and left EI cells, on the other hand, provide different behaviors, as can be seen in Figs. 4(b), 4(c), 5(b), and 5(c). Since the maximum value reached infinity, the results were normalized to the minimum value in each figure.

For the rate coding (Fig. 4), EE cells yielded a minima at about $\mathrm{IPD}=\pm 90°$, where the average rate had a maximum derivative [Eq. (3) and Fig. 3]. An infinite bound was obtained for IPD=0°. This behavior was found for all frequencies [Fig. 4(a)]. Both right and left EI cells provided frequency-dependent minima. For example, for frequencies whose $0.22\leqslant \mathrm{VS}\leqslant 0.64$, the minimum CRLB was obtained at $90°\geqslant \mathrm{IPD}\geqslant 60°$, respectively.

The AIN yielded quite a different behavior as a function of IPD (Fig. 5). For EE it yielded for all frequencies, minimum at about $\mathrm{IPD}\approx\pm 50°$ and a local maximum at IPD $=0°$. The right EI cell yielded, for all frequencies, a minimum at about $\mathrm{IPD}\approx 0°$. On the other hand, the left EI cell yielded a global minimum at $\mathrm{IPD}=\pm 180°$ and a local minimum at $\mathrm{IPD}\approx 0°$.

In order to predict the psychoacoustical results, we combined the information from both sides as expressed in Eq. (13). We expect the best performance in the frontal position and symmetric performance in both sides, i.e., a minimum JND at IPD=0° and a symmetric increase in JND with an absolute increase in IPD (Mills, 1958; Durlach and Colburn, 1978). In view of the results in Figs. 3–5, the psychoacoustical measurements are unpredictable from either rate coding or all-information coding. We thus propose to introduce a phase delay in the ipsilateral input of both EE and EI cells, which means introducing $\Theta_L(f)\neq 0$ or $\Theta_R(f)\neq 0$ in Eq. (19) for left or right CD cells, respectively. In order to obtain the actual phase delay that will yield a minimum JND at IPD $=0°$, the lower bound of the binaural JND was computed according to Eq. (13) with different values of $\Theta_L(f)$ and $\Theta_R(f)$. The phase delay that yielded the minimum JND at IPD=0° was chosen as the ipsilateral input phase delay.

Physiological experiments in cats and guinea pigs that measured the average neural response of different cells in the

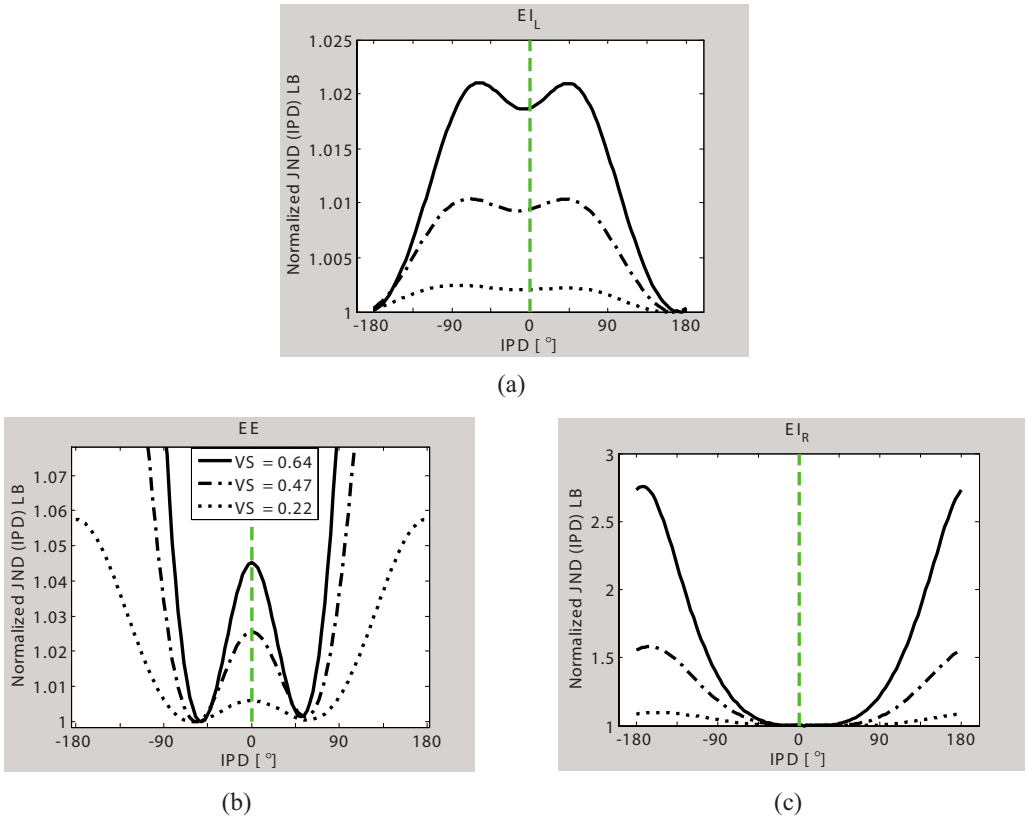


FIG. 5. (Color online) Normalized lower bound of JND of IPD obtained by AIN coding as a function of IPD derived by (a) EE cell, (b) left EI cell, and (c) right EI cell. Each curve represents a different input frequency as indicated by its VS.

IC, whose characteristic frequency was equal to the stimulus frequency, yielded a maximum response at $40° \leqslant \text{IPD} \leqslant 65°$ (Yin and Chan, 1990; Palmer *et al.*, 2002; McAlpine and Grothe, 2003; Hancock and Delgutte, 2004; Joris and Yin, 2006). Low frequencies (below 200 Hz) yielded a maximum response at $\text{IPD} \approx 40°$, while the maximum response at high frequencies (1000–1100 Hz) was obtained at $\text{IPD} \approx 65°$. There was a small monotonic increase in IPD as a function of frequency. One would expect a maximum response at $\text{IPD} \approx 0°$; thus these data are interpreted as having a phase shift between 40° and 65° in the ipsilateral input.

A comparison between the derived ipsilateral phase delay and experimental results is shown in Fig. 6. Derivations were obtained for EE and EI cells according to both types of coding, rate and AIN. The results in Fig. 6 are plotted as a function of VS. The mean experimental data were fitted to a linear regression line, and the measurements' standard deviations were indicated by error bars (McAlpine and Grothe, 2003; Hancock and Delgutte, 2004). The fitted regression lines ± standard deviation are plotted as gray shadowed areas in Fig. 6. There are six panels in Fig. 6; the left panels represent AIN coding, and the right panels represent rate coding. The derivations for EE cells are presented in the upper row, and EI cell derivations are presented in the second and third rows.

Panel (a) in Fig. 6 represents the best input phase delay as derived by EE cell according to AIN coding (solid line). The computed phase delay resembles the experimental results (gray shadowed areas), a moderate decrease in the absolute phase as a function of VS, which means moderate increase as a function of frequency. In none of the other panels, the theoretical derivations (solid lines) resemble the experimental results.

The derived optimal ipsilateral input phase delay was substituted in Eq. (19), the lower bounds of the JND were computed according to Eq. (13), and the normalized values are plotted in Fig. 7 as a function of IPD. The normalization factor was the maximum value of each derivation as a function of IPD. The minimum value in some cases was so small that it seems unreasonable to be used as a normalization factor.

Two types of architectures were tested: combined right and left EE cells and combined right and left EI cells. Each of the architectures was tested by both types of coding rate and AIN. As expected, all the different models yielded a similar behavior, a minimum at $\text{IPD} = 0°$ and a symmetric increase with $|\text{IPD}|$.

The results presented in Fig. 7 cannot indicate which of the four models are most appropriate to predict the psychoacoustical measurements. However, the results shown in Fig. 7 were obtained by using an optimum ipsilateral phase delay. Since only EE cells according to AIN coding provide a reasonable match between the physiological experimental re-

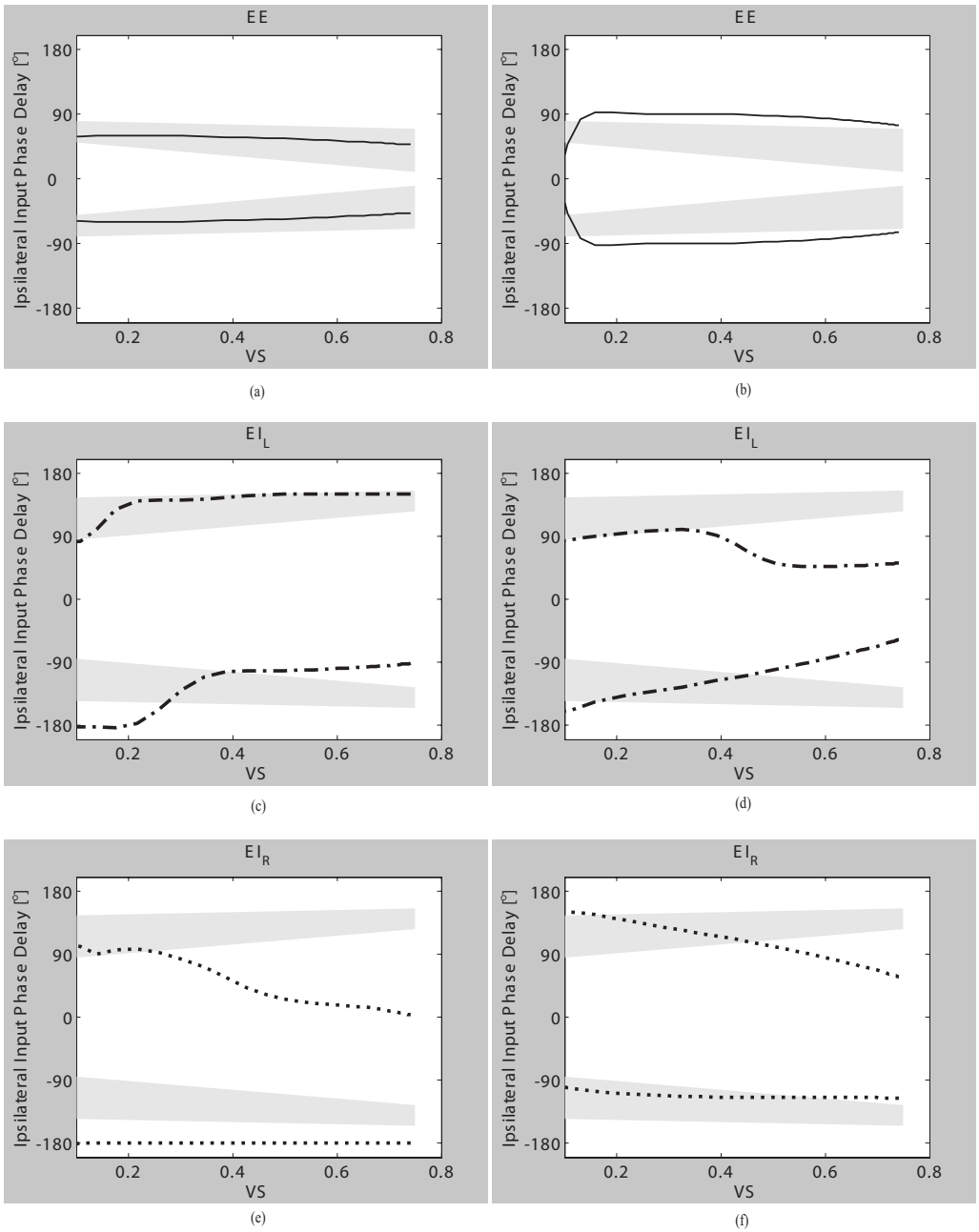


FIG. 6. The derived optimal ipsilateral input phase delay as a function of VS. The theoretical calculations are denoted by the solid lines, and the physiological data are denoted by the gray shadowed areas. Each panel represents a different model as follows: (a) EE cell and AIN coding, (b) EE cell and rate coding, (c) left EI cell and AIN coding, (d) left EI cell and rate coding, (e) right EI cell and AIN coding, and (f) right EI cell and rate coding.

sults (Fig. 6), we conclude that the combined EE cells with AIN coding are the most appropriate to predict the psychoacoustical data.

B. Prediction of interaural level difference

Figure 1 represents the architecture of EE and EI cells that are used for predicting ILD. For simplicity, we define the binaural stimulus with no IPDs, which yields

$$S_L(t) = A_L \sin(2\pi f t),$$
$$S_R(t) = A_R \sin(2\pi f t). \tag{23}$$

Deriving $\lambda_{EE}(t,\mathrm{ILD})$ and $\lambda_{EI}(t,\mathrm{ILD})$ by substituting Eqs. (23) and (19) in Eqs. (11) and (12) reveals that both IRs can be expressed as a multiplication of two functions, where one depends on ILD only and the other one on the time only, i.e.,

$$\lambda_{EE}(t,\mathrm{ILD}) = f_1(\mathrm{ILD}) f_2(t),$$
$$\lambda_{EI}(t,\mathrm{ILD}) = f_3(\mathrm{ILD}) f_4(t). \tag{24}$$

Substituting Eq. (24) in the expressions for the rate CRLB (Eq. (3)) and AIN CRLB (Eq. (5)) reveals an equal value, i.e.,

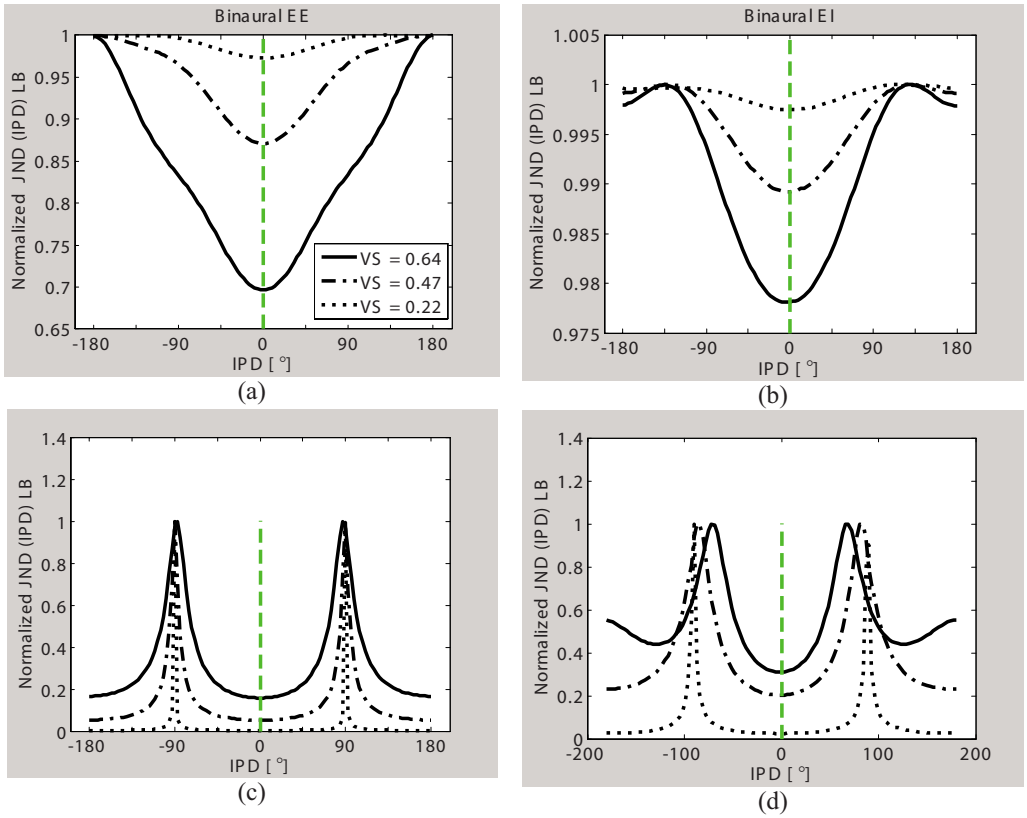


FIG. 7. (Color online) Normalized lower bound of JND as a function of IPD according to the following models: (a) combined right and left EE cells and AIN coding, (b) combined right and left EI cells and AIN coding, (c) combined right and left EE cells and rate coding, and (d) combined right and left EI cells and rate coding.

$$\left.\begin{aligned}\mathrm{CRLB}_{\mathrm{RATE}}^{\mathrm{EE}}(\mathrm{ILD}) &= \mathrm{CRLB}_{\mathrm{AIN}}^{\mathrm{EE}}(\mathrm{ILD}) \\ &= \left\{\frac{1}{f_1(\mathrm{ILD})}\left(\frac{\partial f_1(\mathrm{ILD})}{\partial \mathrm{ILD}}\right)^2 \int_0^T f_2(t)dt\right\}^{-1/2} \\ \mathrm{CRLB}_{\mathrm{RATE}}^{\mathrm{EI}}(\mathrm{ILD}) &= \mathrm{CRLB}_{\mathrm{AIN}}^{\mathrm{EI}}(\mathrm{ILD}) \\ &= \left\{\frac{1}{f_3(\mathrm{ILD})}\left(\frac{\partial f_3(\mathrm{ILD})}{\partial \mathrm{ILD}}\right)^2 \int_0^T f_4(t)dt\right\}^{-1/2}\end{aligned}\right\}. \tag{25}$$

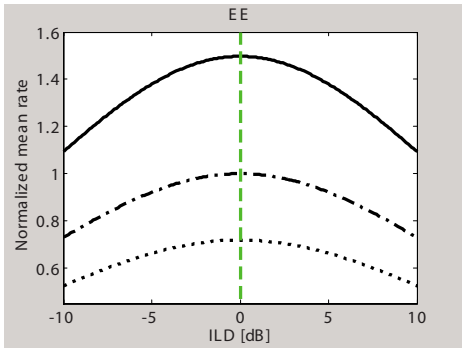
Figure 8 represents the normalized average rates $\bar{\lambda}_{\mathrm{EE}}$ and $\bar{\lambda}_{\mathrm{EI}}$ as functions of ILD for different input frequencies. The spiking rate is normalized so that the sum of the average spiking rate in both AN is equal to 400 spikes/s (e.g., when the amplitude is equal, the average spiking rate for both ears equals 200 spikes/s). Since both right and left EE cells are identical, the normalized average rate of an EE cell is presented in a single figure [Fig. 8(a)] as a function of ILD, which yields a maximum when both inputs are in phase (ILD=0) and decreases as |ILD| increases. In general, a similar behavior was obtained for all input frequencies.

In the simulation shown in Fig. 8, we used $\Delta=20$ μs for EE cells and $\Delta=200$ μs for EI cells similar to the ITD simulation. The left and right EI cells demonstrate a monotonic change with ILD. Right and left SOCs are antisymmetric relative to each other [Figs. 8(b) and 8(c)]. The inhibitory input arrives from the ipsilateral ear, while the excitatory input arrives from the contralateral ear. $\bar{\lambda}_{\mathrm{EI}}^{(R)}$ and $\bar{\lambda}_{\mathrm{EI}}^{(L)}$ yield a minimum when the ipsilateral signal is the strongest and the contralateral is the weakest.

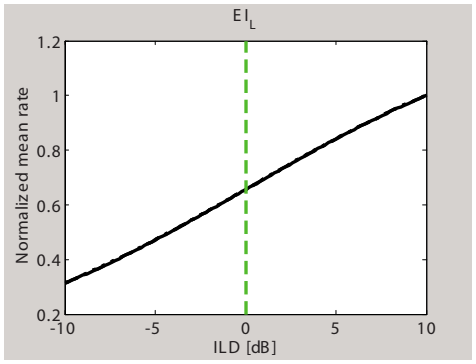
The behavior of the EI mean output spiking rate resembles experimental data in the midlevel range, a monotonic linear increase as a function of ILD in the contralateral EI. EE, on the other hand, produces an artificial ambiguity (signals approaching with either positive or negative ILD obtain equal spiking rates), which is not solved as for ITD since there is no parallel biological mechanism in the ILD processing to the phase shift at the ITD neural mechanisms. EI cells produce an antisymmetric response which is ambiguity-free even when using a single-side SOC. Moreover, the result of the EI cells' mean rate as a function of VS is negligible.

Equation (25) shows that the performance bound (CRLB ILD) when using either rate or AIN model is identical; thus there is no need for the brain to use timing information for ILD extraction. With this information there is no need to plot the CRLB(ILD) using both models. CRLB(ILD) predictions are presented as a function of ILD in Fig. 9. Since $\lambda_{\mathrm{EE}}^{(R)}(t,\mathrm{ILD})=\lambda_{\mathrm{EE}}^{(L)}(t,\mathrm{ILD})$, both right and left EE cells provide identical behaviors that correspond to EE. Right and left EI cells, on the other hand, provide different behaviors, as can be seen in Figs. 9(a) and 9(b).

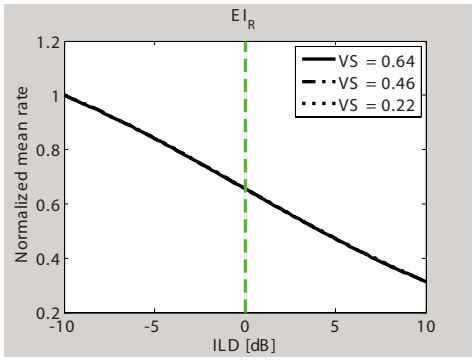
EE cells yield a maxima at ILD=0 dB. This behavior was found for all frequencies (see Eq. (25)). Right and left EI cells provided an antisymmetric behavior between left and



(a)

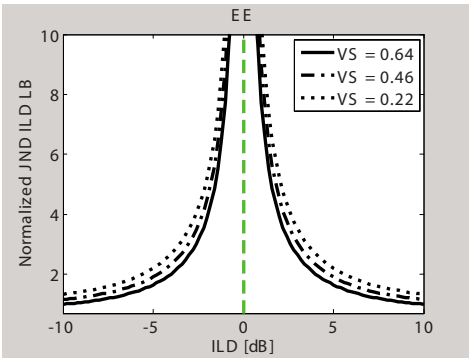


(b)

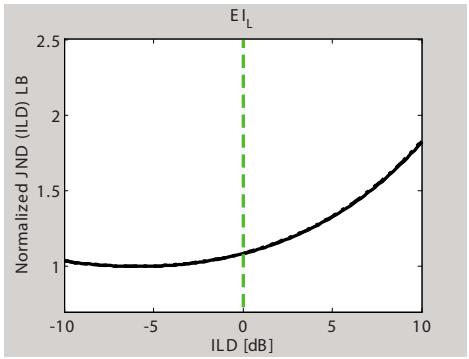


(c)

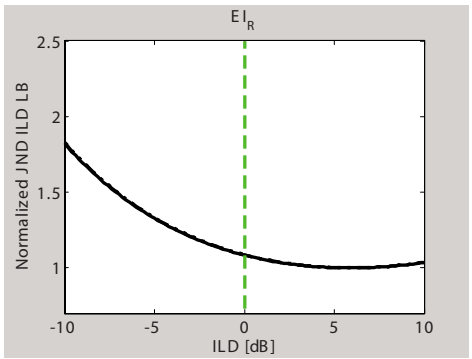
FIG. 8. (Color online) Normalized mean rate as a function of ILD for different frequencies (indicated by their VS) as obtained by the following types of cells: (a) EE, (b) left EI, and (c) right EI. In each panel, the normalization factor is equal to the maximum value.



(a)



(b)



(c)

FIG. 9. (Color online) Normalized lower bound of JND of ILD as a function of ILD derived by (a) EE cell, (b) left EI cell, and (c) right EI cell. Each curve represents a different input frequency as indicated by its VS. In each panel, the normalization factor is equal to the minimum value

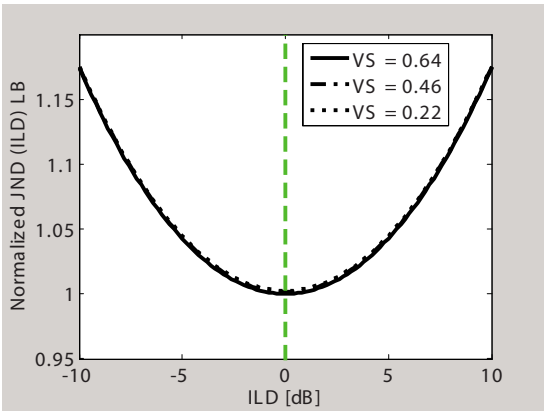


FIG. 10. (Color online) Normalized lower bound of JND as a function of ILD according to combined right and left EI cells for different frequencies as indicated by their VS.

right. This was expected in view of the mean rate behavior (Fig. 8), which fully determines the performance for ILD stimuli.

In order to predict the psychoacoustical results, we combined the information from both sides as expressed in Eq. (13). We expect a minimum JND at ILD=0 dB and a symmetric increase in JND with an absolute increase in ILD (Furst and Algom, 1995). The resulting JND(ILD) is plotted in Fig. 10 for EI cells, whereas the normalized JND(ILD) bound for EE cells is identical to that plotted for monaural EE cells at Fig. 9(a) and found as not adequate. Since the JND of EE cell reached infinity at ILD=0, The JND was normalized to the minimum value. As can be seen, EI predictions resemble the experimental results, a minimum at ILD=0 dB and a symmetric increase with |ILD|. The identical behavior of the bound both for rate and all-information models demonstrates the robustness of the mechanisms especially to injuries that might even harm the synchronization of the signals. The response across frequencies is essential for creating a robust mechanism that can integrate data across frequencies and cope with high-frequency signals with a large bandwidth.

C. Minimal audible angle

In order to evaluate the ability of EE and EI cells to predict the minimum perceived azimuth direction of sine waves, the input signal should be convolved with HRTF [Eq. (14)]. In the following simulation, a typical set of HRTF was used. It was obtained from a "Knowles Electronic Manikin for Auditory Research" (KEMAR) on a dummy head (Algazi *et al.*, 2001).

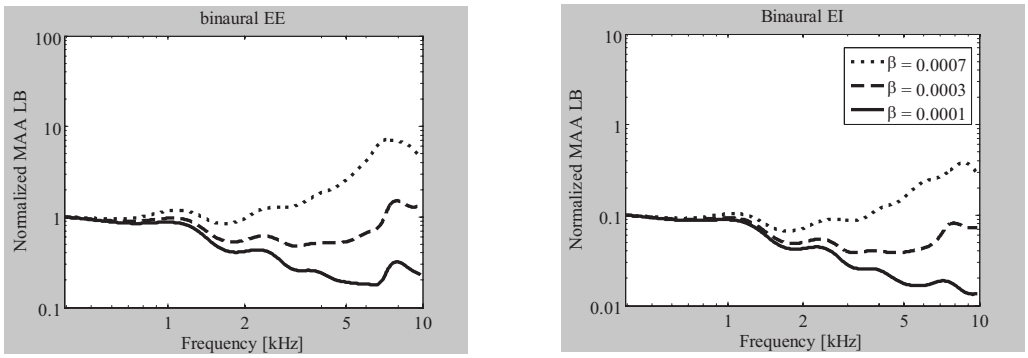
As suggested by Eq. (15), due to HRTF, both ITD and ILD are affected. Thus, in evaluating the MAA, both ITD and ILD should be considered. According to the previous sections, ITD was best described by EE and AIN coding, while ILD was best coded by EI cells where both rate and AIN codings provided identical behaviors. Therefore when MAA is considered, we shall refer to AIN coding only, which will cover both ITD and ILD.

MAA will be tested as a function of frequency. As was mentioned earlier, the frequency dependence is uniquely determined by VS [Eq. (21)]. Figure 11 represents CRLB-(MAA) as a function of frequency for frontal direction (facing the nose) according to AIN coding for both EE and EI cells. All the derivations were obtained from the two sides of the brain (right and left) according to Eq. (13) and are indicated as binaural EE and binaural EI in Fig. 11.

Each curve in Fig. 11 represents a different VS, as indicated by the parameter β in Eq. (21). The curve that represents an input that lost its synchrony at relatively high frequencies (β=0.0001, solid line) reveals a decrease in MAA as a function of frequency. On the other hand, synchrony loss at low frequencies (β=0.0007, dotted line) reveals an increase in MAA as a function of frequency by both EE and EI cells. However, experimental data demonstrate generally improving performance at low frequencies and deteriorating performance at high frequencies with minimal performances at about 2 and 8 kHz (Mills, 1958). None of those experimental results were predicted by CRLB.

It is quite possible that CRLB failed to predict the MAA experimental results due to the fact that it uses only local information of the parameter under test. Human sound localization, on the other hand, is inherently ambiguous; this requires consideration of other possible parameters other than the true.

The BLB seems to be the most appropriate for the purpose of evaluating performance while taking into consideration other possibilities of the estimated parameter in addition to the true one. BLB was derived in Appendix A, and



(a) (b)

FIG. 11. Normalized CRLB of MAA as a function of frequency obtained by binaural EE and binaural EI cells. The different curves represent various VS functions [Eq. (21)] that are denoted by the parameter β.

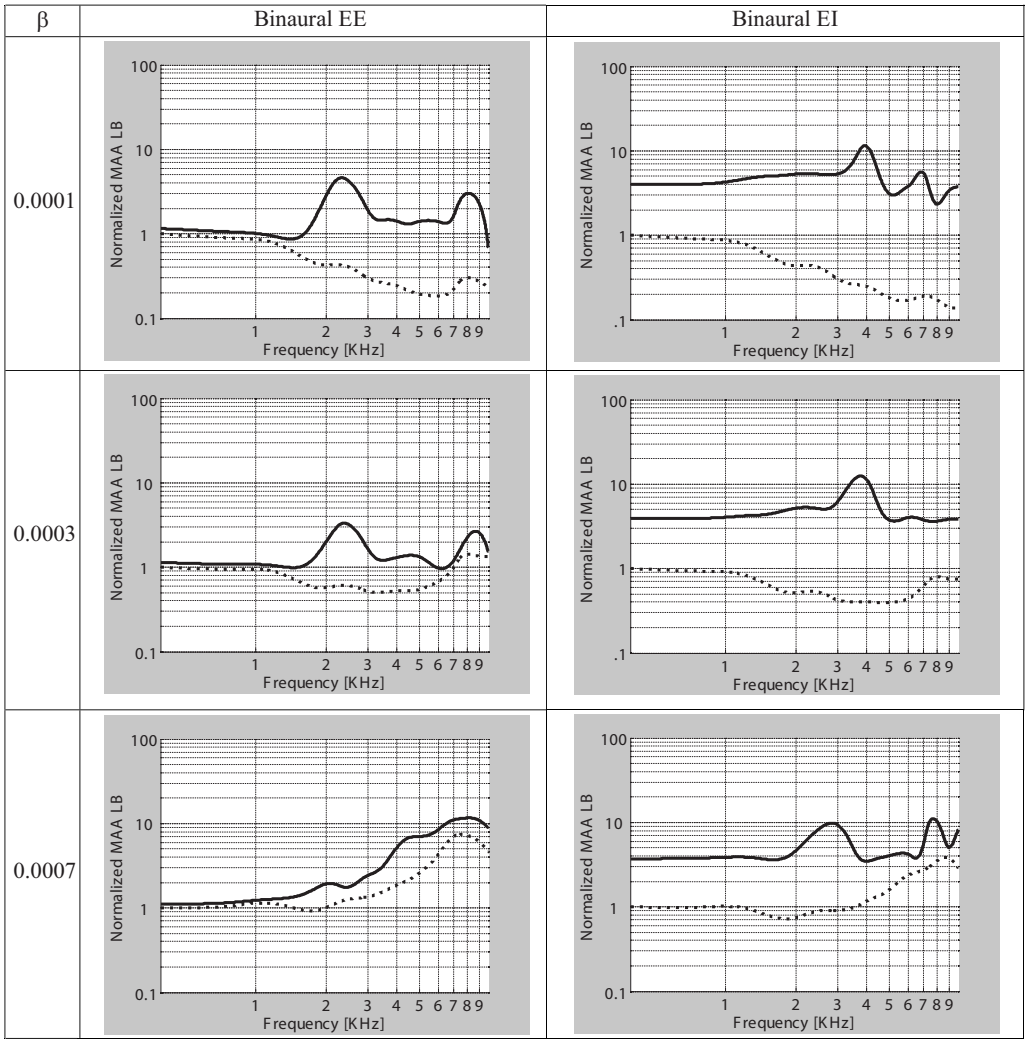


FIG. 12. Normalized BLB (solid line) and CRLB (dotted line) of MAA as a function frequency, obtained from binaural EE cells (left panels) and binaural EI cells (right panels). Each row represents a different VS function [Eq. (21)] and is indicated by the parameter β.

the final results are given in Eqs. (6)–(8). For every tested frequency other directions than the frontal reference direction were taken into account. For all the frequencies, the back-front confusion was included by adding to the true reference $\theta=0°$, the opposite direction $\theta=180°$. Moreover, there were frequencies that yielded more ambiguous directions as was obtained from the HRTF. For example, a signal of 2 kHz that arrives from the front of a subject's face ($\theta=0°$) can be confused with signals coming from the back ($\theta=180°$) and from both sides ($\theta=\pm 90°$). Therefore, in Eqs. (6)–(8), $L=3$, and $\underline{\alpha}=[-90°, 90°, 180°]$.

Figure 12 represents both lower bounds, CRLB (dotted lines) and BLB (solid lines), as functions of frequency as derived by both binaural EE and EI cells. There are six panels in Fig. 12. The left panels represent the binaural EE cells, and the right panels represent the binaural EI cells. Each row represents a different VS function corresponding to those presented in Fig. 11. As expected, BLB yielded higher values than CRLB for all the cases. The greatest difference between BLB and CRLB was obtained for $\beta=0.0001$ (synchrony loss at high frequencies). For EE cells, BLB yielded local maxima at frequencies that are perceived similarly when coming from several directions such as 2.5 and 8 kHz. CRLB, on the other hand, yielded a monotonic decrease for frequencies below 2 kHz. The oscillatory behavior was also obtained by EI cells, but the local maxima were obtained at around 4 and 7 kHz.

Decreasing the frequency for which the input auditory nerve loses its synchrony (increasing β) reveals an increase in MAA predictions by both CRLB and BLB for EE cells. Yet, EI cells maintained an oscillatory behavior for this case as well.

From this simulation it is quite clear that the oscillatory behavior in MAA experimental data (Mills, 1958) is due to the physical acoustical ambiguity of sine waves that affect perception. Both EE and EI cells produced such behavior. The amount of oscillation and the frequencies that yield the worst performances depend on the function that describes the synchrony loss in the auditory nerve. Since this function is unknown in the human auditory nerve, we cannot uniquely determine if EE cells, EI cells, or both are adequate for describing human abilities in MAA experiments.

IV. DISCUSSION

We have investigated the ability of EE and EI CD cells to provide the required acoustical cues (ITD and ILD) and

the source angle for binaural localization. Our study was based on the analytical derivation of the IRs of EE and EI cells whose inputs are NHPP. The ability of EE and EI cells to estimate the input of acoustical cues was analyzed by analytically computing the lower bounds of estimator variances and comparing them to binaural behavioral performances. We compared two types of possible neural coding: (1) rate coding that ignores spike timing information and (2) AIN (all-information) coding that considers spike timing occurrences.

Our simulation revealed the following main results:

(1) Behavioral ITD can be estimated by either EE or EI cells while using any type of coding method only if the ipsilateral input exhibits a phase delay.
(2) Only when EE cells use AIN coding does the optimal derived ipsilateral phase delay match physiological results (Yin and Chan, 1990; Palmer *et al.*, 2002; McAlpine and Grothe, 2003; Hancock and Delgutte, 2004; Joris and Yin, 2006), i.e., a phase delay between 40° and 65°. We thus concluded that ITD is primarily estimated by EE cells when firing timing information is considered.
(3) In deriving the lower bounds for ILD, both types of coding yielded identical results. We thus concluded that for ILD, firing timing information is redundant and rate coding is sufficient.
(4) Only analysis of EI cells revealed an asymmetric response for positive and negative ILDs as found in physiological experiments (Yin and Chan, 1990; Palmer *et al.*, 2002; Park, 1998). Analysis of EE cells yielded a symmetric response that does not match experimental results. We thus conclude that ILD is primarily estimated by EI cells according to rate coding.
(5) In the MAA derivation, the BLB was found as a better estimate than the CRLB. This is due to the fact that the BLB also takes into account ambiguous directions other than the true direction, while CRLB relates only to the true direction.
(6) Following the BLB derivation of MAA, both EE and EI cells revealed an oscillatory performance of MAA as a function of frequency as was found behaviorally (Stevens and Newman, 1934; Mills, 1958; Harris, 1972).

We have tested EE and EI cell performances only for single tones. For ITD estimation, we have shown in the simulation that the CD cells efficiently perform correlation for single tones. However, the analytical expressions for the IRs of EE and EI cells reveal that correlation will be obtained for any input signal. In particular, wide-band signals will yield even better performance since they include more information. Comparing localization of wide-band signals and high-frequency single tones, the former seems to be an easier task than the latter. High-frequency localization suffers from a significant ambiguity problem, but in wide-band signals, EE and EI cells can use additional information such as the signal's envelope that includes low frequencies (Joris and Yin, 1995; Joris, 1996).

Our main results are in general agreement with previous models that included CD cells and modeled binaural abilities, either those that referred to EE cells or those that refer to EI cells and modeled both ITD and ILD (e.g., Lindemann, 1986; Stern and Trahiotis, 1995; Colburn, 1996; Yue and Johnson, 1997; McAlpine *et al.*, 1998; Breebaart *et al.*, 2001; Palmer *et al.*, 2002; Colburn *et al.*, 2004; Park *et al.*, 2004). However, the methodology presented in this paper was proven to be very efficient in determining brainstem neural coding constraints. The use of analytical expressions for the IRs of both EE and EI cells makes it possible to straightforwardly compare the performances of both rate and AIN codings.

The synchrony function of the human auditory nerve (Rose *et al.*, 1967; Johnson, 1980; Palmer and Russell, 1986; Joris *et al.*, 1994; McAlpine *et al.*, 2001; Fitzpatrick and Kuwada, 2001) is still an open question. Using MAA predictions, we have shown that localization performances depend on the synchrony function. The methodology presented in this paper, which takes into account the ambiguity in source localization along with experimental MAA data from different reference positions, might help resolve this issue.

The analysis presented in this paper highlights some interesting points regarding the efficiency of the brain's signal processing. For instance, ITD is estimated by EE cells when all information (AIN model) is considered; it is particularly relevant for low frequencies in the midrange levels where the auditory nerve response is phase locked. For high frequencies, the auditory nerve response is not synchronized, and thus ITD evaluation is not efficient. In examining the MAA predictions as presented in this paper, we saw that the acoustic ambiguity due to human head size causes deterioration in MAA performance even if the auditory nerve would have been phase locked (Krips and Furst, 2006b). Thus it seems that the human brain does not "waste energy" on encoding high-frequency timing information since the performance deteriorates due to ambiguity.

Another interesting point is that for ILD estimation the timing information can be ignored. Therefore AIN and rate codings provide identical performances. Even for low frequencies, AIN is not required for estimating ILD. It seems that the human brain does not restore information that does not provide any significant advantages from the perspective of timing and level of performance.

All the calculations in the present paper used a single cell in each side of the brain. Therefore, the data were normalized, and only a qualitative comparison compared to experimental results was obtained. A quantitative comparison will only be possible when a complete model of SOC and its preceding brainstem neural nuclei will be available. In particular, it should include a number of EE and EI cells and their frequency phase-locking characteristics, including level-dependent properties. However, the methodology used in this paper is adequate for further investigations that will focus on other auditory brainstem source phenomena.

ACKNOWLEDGMENT

This research was partly supported by the Adams Brain Center at Tel Aviv University.

APPENDIX A: BARANKIN LOWER BOUND DERIVATION OF A NONHOMOGENEOUS POISSON PROCESS

Barankin (1949) developed a lower bound that takes into account different possible values of the estimated parameter other than those located in the proximity of the true one (ambiguous locations). If the true value is α^* and there are L other possible values $\alpha_1,\ldots,\alpha_L \neq \alpha^*$, then the BLB was derived and is obtained by

$$\sigma(\hat{\alpha}) \geqslant \mathrm{BLB}(\alpha^*) = \sqrt{[\mathrm{CRLB}(\alpha^*)]^2 + \underline{\Gamma}(\alpha^*)\Delta^{-1}(\alpha^*)\underline{\Gamma}^T(\alpha^*)}, \quad \text{(A1)}$$

where $\underline{\Gamma}(\alpha^*)=\underline{\alpha}-\mathrm{CRLB}^2(\alpha^*)\underline{A}$ and $\Delta(\alpha^*)=B-\underline{A}^T\mathrm{CRLB}^2(\alpha^*)\underline{A}$. $\mathbf{B}$ is a symmetric matrix whose size is $L\times L$, and $\underline{\alpha}$ and $\underline{A}$ are vectors with size L ($\underline{\alpha}=[\alpha_1,\ldots,\alpha_L]$, $\underline{A}=[A_1,\ldots,A_L]$).

McAulay and Hofstetter (1971) showed that for a single parameter vector $\underline{\mathbf{A}}$ elements are given by

$$A_p = \sum_{n=0}^{N}\int_{t_1=0}^{T}\int_{t_2=0}^{T}\cdots\int_{t_n=0}^{T}\frac{\partial \ln[P(t_1,\ldots,t_n|\alpha^*)]}{\partial\alpha} \times P(t_1,\ldots,t_n|\alpha_p)dt_1,\ldots,dt_n, \quad 1\leqslant p\leqslant L, \quad \text{(A2)}$$

and the matrix $\mathbf{B}$ elements are given by

$$B_{ij} = \sum_{n=0}^{N}\int_{t_1=0}^{T}\int_{t_2=0}^{T}\cdots\int_{t_n=0}^{T}\frac{P(t_1,\ldots,t_n|\alpha_i)}{P(t_1,\ldots,t_n|\alpha^*)} \times P(t_1,\ldots,t_n|\alpha_j)dt_1,\ldots,dt_n, \quad \begin{matrix}1\leqslant i\leqslant L\\ 1\leqslant j\leqslant L\end{matrix}. \quad \text{(A3)}$$

In rate coding, the probability $P(t_1,\ldots,t_n|\alpha)=P_{\mathrm{rate}}(N(0,T)=n|\alpha)$, as defined in Eq. (2), which yields

$$\frac{\partial \ln(P(N(0,T)=n|\alpha))}{\partial\alpha} = \frac{n\int_0^T\frac{\partial\lambda(t,\alpha)}{\partial\alpha}dt}{\int_0^T\lambda(t,\alpha)dt} - \int_0^T\frac{\partial\lambda(t,\alpha)}{\partial\alpha}dt. \quad \text{(A4)}$$

Substituting Eqs. (A3) and (2) in Eq. (A2) yields

$$A_p^{\mathrm{rate}} = \sum_{n=0}^{N}\left[\frac{n\int_0^T\frac{\partial\lambda(t,\alpha)}{\partial\alpha^*}dt}{\int_0^T\lambda(t,\alpha^*)dt} - \int_0^T\frac{\partial\lambda(t,\alpha)}{\partial\alpha^*}dt\right]\frac{1}{n!} \times\left[\int_0^T\lambda(t,\alpha_p)dt\right]^n e^{-\int_0^T\lambda(t,\alpha_p)dt}, \quad 1\leqslant p\leqslant L \quad \text{(A5)}$$

Note that $\Sigma_{n=0}^{N}P_{\mathrm{rate}}(N(0,T)=n|\alpha_p)=1$, and for large N, $\Sigma_{n=0}^{N}P_{\mathrm{rate}}(N(0,T)=n-1|\alpha_p)\to 1$. Thus Eq. (A5) becomes

$$A_p^{\mathrm{rate}} = \int_0^T\frac{\partial\lambda(t,\alpha^*)}{\partial\alpha}dt\left[\frac{\int_0^T\lambda(t,\alpha_p)dt}{\int_0^T\lambda(t,\alpha^*)dt}-1\right], \quad 1\leqslant p\leqslant L. \quad \text{(A6)}$$

Similarly the elements of matrix $\mathbf{B}$ are obtained by substituting Eq. (2) in Eq. (A2), which yields

$$B_{ij} = \sum_{n=0}^{N}\left\{\frac{\int_0^T\lambda(t,\alpha_i)dt}{\int_0^T\lambda(t,\alpha^*)dt}\int_0^T\lambda(t,\alpha_j)dt\right\}^n \times\exp\left\{-\int_0^T\lambda(t,\alpha_i)dt-\int_0^T\lambda(t,\alpha_j)dt + \int_0^T\lambda(t,\alpha^*)dt\right\}, \quad \begin{matrix}1\leqslant i\leqslant L\\ 1\leqslant j\leqslant L\end{matrix} \quad \text{(A7)}$$

For large N, the sum in Eq. (A7) can be expressed as an exponent ($\Sigma_{n=0}^{\infty}(a^n/n!)=e^a$), which yields

$$B_{i,j}^{\mathrm{rate}} = \exp\left(\int_0^T[-\lambda(t,\alpha_i)-\lambda(t,\alpha_j)+\lambda(t,\alpha^*)]dt + \frac{\int_0^T\lambda(t,\alpha_i)dt}{\int_0^T\lambda(t,\alpha^*)dt}\int_0^T\lambda(t,\alpha_j)dt\right). \quad \text{(A8)}$$

In AIN coding, the probability $P(t_1,\ldots,t_n|\alpha)=P_{\mathrm{AIN}}(t_1,\ldots,t_n|\alpha)$, as defined in Eq. (4), which yields

$$\frac{\partial \ln[P(t_1,\ldots,t_n|\alpha^*)]}{\partial\alpha} = \sum_{i=1}^{n}\frac{\partial\lambda(t_i,\alpha^*)}{\partial\alpha}\frac{1}{\lambda(t_i,\alpha^*)} - \int_0^T\frac{\partial\lambda(t,\alpha^*)}{\partial\alpha}dt. \quad \text{(A9)}$$

Substituting Eqs. (A9) and (4) in Eq. (A2) yields an expression for A_p, $A_p^{\mathrm{AIN}}=A_p^{(1)}-A_p^{(2)}$, where

$$A_p^{(1)} = \sum_{n=1}^{N}\sum_{i=1}^{n}\int_{t_1=0}^{T}\int_{t_2=0}^{T}\cdots\int_{t_n=0}^{T}\Upsilon(t_1,\ldots,t_n)dt_1,\ldots,dt_n, \quad \text{(A10)}$$

$$\Upsilon(t_1,\ldots,t_n) = \frac{\partial\lambda(t_i,\alpha^*)}{\partial\alpha}\frac{1}{\lambda(t_i,\alpha^*)} \times\left[\frac{\prod_{k=1}^{n}\lambda(t_k,\alpha_p)}{n!}\exp\left\{-\int_0^T\lambda(t,\alpha_p)dt\right\}\right].$$

Note that the sum in $A_p^{(1)}$ starts at $n=1$ since $n=0$ does not exist in Eq. (A9) and

$$A_p^{(2)} = \sum_{n=0}^{N} \int_{t_1=0}^{T} \int_{t_2=0}^{T} \cdots \int_{t_n=0}^{T} \left[\int_0^T \frac{\partial \lambda(t,\alpha^*)}{\partial \alpha} dt \right] \times \left[\frac{\prod_{k=1}^{n} \lambda(t_k,\alpha_p)}{n!} \exp\left\{ - \int_0^T \lambda(t,\alpha_p) dt \right\} \right] dt_1, \ldots, dt_n. \quad \text{(A11)}$$

Let us denote $P_{\mathrm{AIN}}(n-1\,|\,\alpha)$ as the AIN probability when $N(0,T)=n-1$ and substituting it in Eq. (A10), which yields

$$A_p^{(1)} = \sum_{n=1}^{N} \sum_{i=1}^{n} \int_{t=0}^{T} \frac{\partial \lambda(t,\alpha^*)}{\partial \alpha} \frac{\lambda(t,\alpha_p)}{\lambda(t,\alpha^*)} dt \frac{P_{\mathrm{AIN}}(n-1|\alpha_p)}{n}. \quad \text{(A12)}$$

Replacing the order of the sums and integration yields

$$A_p^{(1)} = \sum_{n=1}^{N} \frac{P_{\mathrm{AIN}}(n-1|\alpha_p)}{n} \sum_{i=1}^{n} \int_{t=0}^{T} \frac{\partial \lambda(t,\alpha^*)}{\partial \alpha} \frac{\lambda(t,\alpha_p)}{\lambda(t,\alpha^*)} dt. \quad \text{(A13)}$$

Since $\int_{t=0}^{T} (\partial \lambda(t,\alpha^*)/\partial \alpha)(\lambda(t,\alpha_p)/\lambda(t,\alpha^*))dt$ is independent both of i and n, $A_p^{(1)}$ can be expressed as

$$A_p^{(1)} = \int_{t=0}^{T} \frac{\partial \lambda(t,\alpha^*)}{\partial \alpha} \frac{\lambda(t,\alpha_p)}{\lambda(t,\alpha^*)} dt \sum_{n=1}^{N} \frac{P_{\mathrm{AIN}}(n-1|\alpha_p)}{n} n. \quad \text{(A14)}$$

For large N, $\Sigma_{n=1}^{N} P(n-1\,|\,\alpha) \to 1$; thus

$$A_p^{(1)} = \int_{t=0}^{T} \frac{\partial \lambda(t,\alpha^*)}{\partial \alpha} \frac{\lambda(t,\alpha_p)}{\lambda(t,\alpha^*)} dt. \quad \text{(A15)}$$

Similarly $A_p^{(2)}$ can be expressed as

$$A_p^{(2)} = \int_0^T \frac{\partial \lambda(t,\alpha^*)}{\partial \alpha} dt \sum_{n=0}^{N} \left(\int_{t_1=0}^{T} \int_{t_2=0}^{T} \cdots \int_{t_n=0}^{T} \frac{\prod_{k=1}^{n} \lambda(t_k,\alpha_p)}{n!} \times \exp\left\{ - \int_0^T \lambda(t,\alpha_p) dt \right\} dt_1, \ldots, dt_n \right). \quad \text{(A16)}$$

The second element in Eq. (A16) is the sum over all the possible discharge possibilities, i.e., $(\Sigma_{n=0}^{N} P_{\mathrm{AIN}}(n|\alpha)=1)$. Hence $A_p^{(2)}$ is given by

$$A_p^{(2)} = \int_0^T \frac{\partial \lambda(t,\alpha^*)}{\partial \alpha} dt. \quad \text{(A17)}$$

Substituting Eqs. (A15) and (A17) in the expression for A_p^{AIN} yields

$$A_p^{\mathrm{AIN}} = \int_0^T \frac{\partial \lambda(t,\alpha^*)}{\partial \alpha} \frac{\lambda(t,\alpha_p)}{\lambda(t,\alpha^*)} dt - \int_0^T \frac{\partial \lambda(t,\alpha^*)}{\partial \alpha} dt. \quad \text{(A18)}$$

Similarly, the elements in matrix **B** each are given by

$$B_{i,j}^{\mathrm{rate}} = \exp\left\{ - \int_0^T [\lambda(t,\alpha_i) + \lambda(t,\alpha_j) - \lambda(t,\alpha^*)] dt \right\} \times \sum_{n=0}^{N} \frac{1}{n!} \left(\int_{t=0}^{T} \frac{\lambda(t,\alpha_i)\lambda(t,\alpha_j)}{\lambda(t,\alpha^*)} dt \right)^n. \quad \text{(A19)}$$

For large N, the sum in Eq. (A19) can be expressed as an exponent $(\Sigma_{n=0}^{\infty}(a^n/n!)=e^a)$, which yields

$$B_{i,j}^{\mathrm{AIN}} = \exp\left(\int_0^T \left[- \lambda(t,\alpha_i) - \lambda(t,\alpha_j) + \lambda(t,\alpha^*) + \frac{\lambda(t,\alpha_i)\lambda(t,\alpha_j)}{\lambda(t,\alpha^*)} \right] dt \right). \quad \text{(A20)}$$

Agmon-Snir, H., Carr, C. E., and Rinzel, J. (**1998**). "The role of dendrites in auditory coincidence detection," Nature (London) **393**, 268–272.

Algazi, V. R., Duda, R. O., Thompson, D. M., and Avendano, C. (**2001**). "The CIPIC HRTF database," IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics, pp. 99–102.

Bar David, I. (**1969**). "Communication under Poisson regime," IEEE Trans. Inf. Theory **IT-15**, 31–37.

Barankin, E. W. (**1949**). "Locally best unbiased estimates," Ann. Math. Stat. **20**, 477–501.

Batra, R., Kuwada, S., and Fitzpatrick, D. C. (**1997**). "Sensitivity to interaural temporal disparities of low- and high-frequency neurons in the superior olivary complex. I. Heterogeneity of responses," J. Neurophysiol. **78**, 1222–1236.

Blauert, J. (**1997**). *Spatial Hearing: The Psychophysics of Human Sound Localization* (MIT, London).

Boudreau, J. C., and Tsuchitani, C. (**1968**). "Binaural interaction in the cat superior olive S segment," J. Neurophysiol. **31**, 442–454.

Boudreau, J., and Tsuchitani, C. (**1970**). "Cat superior olive S-segment cell discharge to tonal stimulation," Contrib Sens Physiol. **4**, 143–213.

Brand, A., Behrend, O., Marquardt, T., McAlpine, D., and Grothe, B. (**2002**). "Precise inhibition is essential for microsecond interaural time difference coding," Nature (London) **417**, 543–547.

Breebaart, J., Van de Par, S., and Kohlrausch, A. (**2001**). "Binaural processing model based on contralateral inhibition. I. Model structure," J. Acoust. Soc. Am. **110**, 1074–1088.

Brown, C. J., and Abbas, P. J. (**1990**). "Electrically evoked whole-nerve action poten tials: Parametric data from the cat," J. Acoust. Soc. Am. **88**, 2205–2210.

Bruce, I. C., Irlicht, L. S., White, M. W., O'Leary, S. J., Dynes, S., Javel, E., and Clark, G. M. (**1999**). "A stochastic model of the electrically stimulated auditory nerve: Pulse-train response," IEEE Trans. Biomed. Eng. **46**, 630–637.

Caird, D. M., and Klinke, R. (**1983**). "Processing of binaural stimuli by cat superior olivary complex neurons," Exp. Brain Res. **52**, 385–399.

Cant, N. B. (**1991**). "Projections to the lateral and medial superior olivary nuclei from the spherical and globular bushy cells of the anteroventral cochlear nucleus," in *Neurobiology of Hearing: The Central Auditory System*, edited by R. A. Altschuler, R. P. Bobbin, B. M. Clopton, and D. W. Hoffman (Raven, New York), pp. 99–119.

Caspary, D. M., and Finlayson, P. G. (**1991**). "Superior olivary complex—Functional neuropharmacology of the principal cell types," in *Neurobiology of Hearing: The Central Auditory System*, edited by R. A. Altschuler, R. P. Bobbin, B. M. Clopton, and D. W. Hoffman (Raven, New York), pp. 141–161.

Cohen, O., Furst, M., and Krips, R., (**2004**). "ITD and ILD estimation based on neural stochastic analysis," Proceedings of the 23rd IEEE Convention of Electrical and Electronics Engineers in Israel, pp. 185–188.

Colburn, S. H. (**1973**). "Theory of binaural interaction based on auditory-nerve data. I. General strategy and preliminary results on interaural discrimination," J. Acoust. Soc. Am. **54**, 1458–1470.

Colburn, H. S. (**1996**). "Computational models of binaural processing," in *Springer Handbook of Auditory Research, Vol. VI: Auditory Computation*, edited by H. L. Hawkins, T. A. McMullen, A. N. Popper, and R. R. Fay (Springer-Verlag, New York).

Colburn, H. S., Zhou, Y., and Dasika, V. K. (**2004**). "Inhibition in models of

coincidence detection," *Auditory Signal Processing: Physiology, Psychoacoustics, and Models* (Springer-Verlag, New York).
Crow, G., Rupert, A. L., and Moushegian, G. (**1978**). "Phase locking in monaural and binaural medullary neurons: Implications for binaural phenomena," J. Acoust. Soc. Am. **64**, 493–501.
Delgutte, B., Joris, P. X., Litovsky, R. Y., and Yin, T. C. (**1999**). "Receptive fields and binaural interactions for virtual-space stimuli in the cat inferior colliculus," J. Neurophysiol. **81**, 2833–2851.
Durlach, N. I., and Colburn, H. S. (**1978**). "Binaural phenomena," in *Handbook of Perception*, edited by E. C. Carterette and M. P. Freidman (Academic, New York), Vol. **4**, pp. 365–466.
Dynes, S. (**1996**). "Discharge characteristics of auditory nerve fibers for pulsatile electrical stimuli," Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, MA.
Fitzpatrick, D. C., and Kuwada, S. (**2001**). "Tuning to interaural time differences across frequency," J. Neurosci. **21**, 4844–4851.
Fitzpatrick, D. C., Kuwada, S., and Batra, R. (**2000**). "Neural sensitivity to interaural time differences: Beyond the Jeffress model," J. Neurosci. **20**, 1605–1615.
Furst, M., and Algom, D. (**1995**). "Lateralization and discrimination of dichotic clicks: Evidence from patients with brainstem lesions and normal cohorts," J. Basic Clin. Physiol. Pharmacol. **6**, 149–171.
Glendenning, K. K., Hutson, K. A., Nudo, R. J., and Masterton, R. B. (**1985**). "Acoustic chiasm. II: Anatomical basis of binaurality in lateral superior olive of cat," J. Comp. Neurol. **232**, 261–285.
Goldberg, J. M., and Brown, P. B. (**1969**). "Response of binaural neurons of dog superior olivary complex to dichotic tonal stimuli: Some physiological mechanisms of sound localization," J. Neurophysiol. **32**, 613–636.
Grantham, D. W. (**1986**). "Detection and discrimination of simulated motion of auditory targets in the horizontal plane," J. Acoust. Soc. Am. **79**, 1939–1949.
Guinan, J. J., Guinan, S. S., and Norris, B. E. (**1972**). "Single auditory units in the superior olivary complex. I. Responses to sounds and classifications based on physiological properties," Int. J. Neurosci. **4**, 101–120.
Hancock, K. E., and Delgutte, B., (**2004**). "A physiologically based model of interaural time difference discrimination," J. Neurosci. **24**, 7110–7117.
Harris, J. D. (**1972**). "A florilegium of experiments on directional hearing," Acta Oto-Laryngol. **298**, 3–26.
Hartmann, W. M., and Rakerd, B. (**1989**). "On the minimum audible angle: A decision theory approach," J. Acoust. Soc. Am. **85**, 2031–2041.
Heinz, M. G., Colburn, H. S., and Carney, L. H. (**2001**). "Evaluating auditory performance limits: I. One-parameter discrimination using a computational model for the auditory nerve," Neural Comput. **13**, 2273–2316.
Huettel, L. G., and Collins, L. M. (**2004**). "Predicting auditory tone-in-noise detection performance: The effects of neural variability," IEEE Trans. Biomed. Eng. **51**, 282–293.
Irvine, D. R. F. (**1986**). *The Auditory Brainstem* (Springer-Verlag, Berlin).
Irvine, D. R. F. (**1992**). "Physiology of the auditory brain stem," in *The Mammalian Auditory Pathway: Neurophysiology*, edited by A. N. Popper and R. R. Fay (Springer, New York), pp. 153–231.
Jeffress, L. A. (**1948**). "A place theory of sound localization," J. Comp. Physiol. Psychol. **41**, 35–39.
Johnson, D. (**1980**). "The relationship between spike rate and synchrony in responses of auditory-nerve fibers to single tones," J. Acoust. Soc. Am. **68**, 1115–1122.
Joris, P. X. (**1996**). "Envelope coding in the lateral superior olive. II. Characteristic delays and comparison with responses in the medial superior olive," J. Neurophysiol. **76**, 2137–2156.
Joris, P. X., Carney, L. H., Smith, P. H., and Yin, T. C. T. (**1994**). "Enhancement of neural synchronization in the anteroventral cochlear nucleus. I. Responses to tones at the characteristic frequency," J. Neurophysiol. **71**, 1022–1036.
Joris, P. X., Smith, P. H., and Yin, T. C. T. (**1998**). "Coincidence detection in the auditory system: 50 years after Jeffress," Neuron **21**, 1235–1238.
Joris, P. X., and Yin, T. C. T. (**1995**). "Envelope coding in the lateral superior olive. I. Sensitivity to interaural time differences," J. Neurophysiol. **73**, 1043–1062.
Joris, P. X., and Yin, T. C. T. (**2006**). "A matter of time: Internal delays in binaural processing," Trends Neurosci. **30**, 70–78.
King, A. J., and Palmer, A. R. (**1983**). "Cells responsive to free-field auditory stimuli in guinea-pig superior colliculus: Distribution and response properties," J. Physiol. (London) **342**, 361–381.
Klumpp, R. G., and Eady, H. R. (**1956**). "Some measurements of interaural time difference thresholds," J. Acoust. Soc. Am. **28**, 859–860.
Knudsen, E. I. (**1982**). "Auditory and visual maps of space in the optic tectum of the owl," J. Neurosci. **2**, 1177–1194.
Knudsen, E. I., and Konishi, M. (**1978**). "A neural map of auditory space in the owl," Science **200**, 795–797.
Krips, R. (**2008**). "Stochastic properties of coincidence-detector neural cells and their implications for binaural perception," Ph.D. dissertation, Tel-Aviv University, Tel-Aviv, Israel.
Krips, R., and Furst, M. (**2006a**). "Probability characteristics of neural coincidence detectors in the brainstem," J. Acoust. Soc. Am. **120**, 3258.
Krips, R., and Furst, M. (**2006b**). "Theoretical MAA estimation based on ambiguity information," ARO MWM.
Kuhn, G. F. (**1987**). "Physical acoustics and measurements pertaining to directional hearing," in *Directional Hearing*, edited by W. A. Yost and G. Gourevitch (Springer, New York), pp. 3–25.
Lindemann, W. (**1986**). "Extension of a binaural cross-correlation model by contralateral inhibition. I. Simulation of lateralization for stationary signals," J. Acoust. Soc. Am. **80**, 1608–1622.
Liu, L. F., Palmer, A. R., and Wallace, M. N. (**2006**). "Phase-locked responses to pure tones in the inferior colliculus," J. Neurophysiol. **95**, 1926–1935.
McAlpine, D., and Grothe, B. (**2003**). "Sound localization and delay lines—Do mammals fit the model?," Trends Neurosci. **26**, 347–350.
McAlpine, D., Jiang, D., and Palmer, A. R. (**2001**). "A neural code for low-frequency sound localization in mammals," Nat. Neurosci. **4**, 396–401.
McAlpine, D., Jiang, D., Shackleton, T. M., and Palmer, A. R. (**1998**). "Convergent input from brainstem coincidence detectors onto delay sensitive neurons in the inferior colliculus," J. Neurosci. **18**, 6026–6039.
McAulary, R. J., and Hofstetter, E. M. (**1971**). "Barankin bounds on parameter estimation," IEEE Trans. Inf. Theory **IT 17**, 669–676.
Miller, C. A., Abbas, P. J., and Robinson, B. K. (**2001**). "Response properties of the refractory auditory nerve fiber," J. Assoc. Res. Otolaryngol. **2**, 216–232.
Mills, A. W. (**1958**). "On the minimum audible angle," J. Acoust. Soc. Am. **30**, 237–246.
Mills, A. W. (**1960**). "Lateralization of high-frequency tones," J. Acoust. Soc. Am. **32**, 132–134.
Moushegian, G., Rupert, A. L., and Gidda, J. S. (**1975**). "Functional characteristics of superior olivary neurons to binaural stimuli," J. Neurophysiol. **38**, 1037–1048.
Palmer, A. R., McAlpine, D., and Jiang, D. (**1997**). "Processing of interaural delay in the inferior colliculus," in *Acoustical Signal Processing in the Central Auditory System*, edited by J. Syka (Plenum, New York), pp. 353–364.
Palmer, A. M., and Russell, I. J. (**1986**). "Phase-locking in the cochlear nerve of the guinea-pig and its relation to the receptor potential of inner hair-cells," Hear. Res. **24**, 1–15.
Palmer, A. R., Shackleton, T. M., and McAlpine, D. (**2002**). "The physiological basis of the binaural masking level difference," Acta. Acust. Acust. **88**, 312–319.
Park, T. J. (**1998**). "IID sensitivity differs between two principal centers in the interaural intensity difference pathway: The LSO and the IC," J. Neurophysiol. **79**, 2416–2431.
Park, T. J., Klug, A., Holinstat, M., and Grothe, B. (**2004**). "Interaural level difference processing in the lateral superior olive and the inferior colliculus," J. Neurophysiol. **92**, 289–301.
Reyes, A. D., Rubel, E. W., and Spain, W. J. (**1996**). "In vitro analysis of optimal stimuli for phase-locking and time-delayed modulation of firing in avian nucleus laminaris neurons," J. Neurosci. **16**, 993–1007.
Rose, J. E., Brugge, J. F., Anderson, D. J., and Hind, J. E. (**1967**). "Phase-locked response to low-frequency tones in single auditory nerve fibers of the squirrel monkey," J. Neurophysiol. **30**, 769–793.
Searle, C. L., Braida, L. D., Davis, M. F., and Colburn, H. S. (**1976**). "Model for auditory localization," J. Acoust. Soc. Am. **60**, 1164–1175.
Semple, M. N., Aitkin, L. M., Calford, M. B., Pettigrew, J. D., and Phillips, D. P. (**1983**). "Spatial receptive fields in the cat inferior colliculus," Hear. Res. **10**, 203–215.
Siebert, W. M. (**1968**). "Stimulus transformation in the peripheral auditory system," in *Recognizing Patterns*, edited by P. A. Kolers and M. Eden (MIT, Cambridge, MA), pp. 104–133.
Snyder, D. L., and Miller, M. I. (**1991**). *Random point Processes in Time and Space* (Springer-Verlag, Berlin).
Spitzer, M. W., and Semple, M. N. (**1995**). "Neurons sensitive to interaural phase disparity in gerbil superior olive: Diverse monaural and temporal

response properties," J. Neurophysiol. **73**, 1668–1690.
Stern, R. M., Jr., and Colburn, H. S. (**1978**). "Theory of binaural interaction based on auditory-nerve data. IV. A model for subjective lateral position," J. Acoust. Soc. Am. **64**, 127–140.
Stern, R. M., and Trahiotis, C. (**1995**). "Models of binaural perception," in *Hearing*, edited by B. C. J. Moore (Academic, New York), pp. 347–386.
Stevens, S. S., and Newman, E. B. (**1934**). "The localization of pure tones," Proc. Natl. Acad. Sci. U.S.A. **20**, 593–596.
Warr, W. B. (**1966**). "Fiber degeneration following lesions in the anteroventral cochlear nucleus of the cat," Exp. Neurol. **23**, 140–155.
Warr, W. B. (**1969**). "Fiber degeneration following lesions in the posteroventral cochlear nucleus of the cat," Exp. Neurol. **23**, 140–155.
Warr, W. B. (**1972**). "Fiber degeneration following lesions in the multipolar and globular cell areas in the ventral cochlear nucleus of the cat," Brain Res. **40**, 247–270.
Yin, T. C., and Chan, J. C. (**1990**). "Interaural time sensitivity in medial superior olive of cat," J. Neurophysiol. **64**, 465–488.
Yin, T. C. T., Chan, J. C. K., and Carney, L. H. (**1987**). "Effects of interaural time delays of noise stimuli on low-frequency cells in the cat's inferior colliculus. III. Evidence for cross-correlation," J. Neurophysiol. **58**, 562–582.
Yue, L., and Johnson, D. H. (**1997**). "Optimal binaural processing based on point process models of preprocessed cues," J. Acoust. Soc. Am. **101**, 982–992.
Zwislocki, J., and Feldman, R. S. (**1956**). "Just noticeable differences in dichotic phase," J. Acoust. Soc. Am. **28**, 860–864.

Considering distortion product otoacoustic emission fine structure in measurements of the medial olivocochlear reflex

Carolina Abdala[a)]
House Ear Institute, 2100 West Third Street, Los Angeles, California 90057

Srikanta K. Mishra
Institute of Sound and Vibration Research, University of Southampton, University Road, Southampton, S017 1BJ, United Kingdom

Tracy L. Williams
House Ear Institute, 2100 West Third Street, Los Angeles, California 90057

(Received 14 October 2008; revised 8 December 2008; accepted 9 December 2008)

In humans, when the medial olivocochlear (MOC) pathway is activated by noise in the opposite ear, changes in distortion product otoacoustic emission (DPOAE) level, i.e., the MOC reflex, can be recorded in the test ear. Recent evidence suggests that DPOAE frequency influences the direction (suppression/enhancement) of the reflex. In this study, DPOAEs were recorded at fine frequency intervals from 500 to 2500 Hz, with and without contralateral acoustic stimulation (CAS) in a group of 15 adults. The MOC reflex was calculated only at DPOAE frequencies corresponding to peaks in the fine structure. Additionally, inverse fast-Fourier transform was conducted to evaluate MOC effects on individual DPOAE components. Results show the following: (1) When considering peaks only, the mean MOC reflex was −2.05 dB and 97% of observations reflected suppression, (2) CAS reduced distortion characteristic frequency component levels more than overlap component levels, and (3) CAS produced an upward shift in fine structure peak frequency. Results indicate that when the MOC reflex is recorded at DPOAE frequencies corresponding to fine structure maxima (i.e., when DPOAE components are constructive and in phase), suppression is reliably observed and level enhancement, which probably reflects component mixing in the ear canal rather than strength of the MOC reflex, is eliminated. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3068442]

PACS number(s): 43.64.Jb, 43.64.Bt [BLM] Pages: 1584–1594

I. INTRODUCTION

Descending medial olivocochlear (MOC) fibers synapse almost exclusively on outer hair cells (OHCs) and through cholinergic channels, alter OHC function. This alteration in OHC function influences motion of the basilar membrane and by extension, impacts otoacoustic emissions (OAEs) (Guinan, 2006). The MOC reflex, as it has been termed, has been studied in laboratory animals and humans by recording distortion product otoacoustic emissions (DPOAEs) in one ear while presenting contralateral noise to the opposite ear (Collet *et al.*, 1990; Puel and Rebillard, 1990; Moulin *et al.*, 1993). Contralateral acoustic stimulation (CAS) is thought to activate the uncrossed MOC pathway. This activation can be quantified by measuring the difference between DPOAE levels recorded in no-CAS and +CAS conditions. Another means of activating the MOC reflex is to present a suppressor ipsilaterally. This produces a larger onset adaptation response in cats and other laboratory animals, presumably because it includes the larger bundle of crossed fibers (Liberman *et al.*, 1996). Work with humans, however, has not shown the same robust DPOAE onset adaptation (Agrama *et al.*, 1998; Kim *et al.*, 2001).

DPOAEs are generally *reduced* in level by MOC activity, but sometimes show amplitude enhancement as well (Maison and Liberman, 2000; Abdala *et al.*, 1999; Müller *et al.*, 2005). The amount of MOC-induced suppression typically ranges from 1 to 3 dB in humans (Abdala *et al.*, 1999; Bassim *et al.*, 2003; Moulin *et al.*, 1993). Some studies have reported larger changes when the MOC reflex value is calculated in absolute values (i.e., disregarding direction of the change) or when the *range* between suppression and enhancement is used as an index of the MOC reflex (Müller *et al.*, 2005; Wagner *et al.*, 2007).

The MOC reflex has been measured using other types of OAEs that arguably could provide a more effective assay than DPOAE-based paradigms. Guinan (2006, p. 597) warns that the most critical difficulty associated with use of DPOAEs for measurement of the MOC reflex is that "…the effect can be in either direction and can change greatly with small changes in stimulus parameter, thereby making a single measurement difficult to interpret." Recent modifications to the typical DPOAE-based MOC reflex protocol have attempted to control for, or at least consider, the bidirectionality of changes in DPOAE level produced by CAS (Sun, 2008b; Müller *et al.*, 2005; Zhang *et al.*, 2007). These modified paradigms take into account how the choice of DPOAE frequency (and where this frequency falls along the pattern of alternating level peaks and dips known as fine structure) influences measures of the MOC reflex.

a)Electronic mail: cabdala@hei.org

 0001-4966/2009/125(3)/1584/11/$25.00

The ear canal DPOAE is thought to be comprised of a nonlinear distortion component from around the primary tone overlap or f_2 region, and a reflection component from the DP characteristic frequency (CF) region at $2f_1-f_2$ (Dhar *et al.*, 2002; Kim, 1980; Shera and Guinan, 1999; Talmadge *et al.*, 1999). These two components mix in the ear canal through vector summation and form the composite DPOAE that is traditionally measured at the probe microphone. Because phase rotates as a function of frequency at different rates for each component, they can sum either constructively or destructively, depending on their relative phase. DPOAE fine structure is considered an "interference pattern," depicting the phase relationships between these two DPOAE sources across frequency. If the components sum while out of phase, they can cancel one another and produce reduced DPOAE level; that is, a "minimum" or dip in fine structure. If they combine in a constructive fashion, while in phase, they augment DPOAE level and produce a "maximum" or peak in fine structure. Thus, for any DPOAE measurement, the test frequencies selected determine the phase relationship between components.

If MOC activation influences the two DP components differentially, it could explain observations of DPOAE level enhancement. For example, if the DPOAE is measured at a frequency where the two components are out of phase, producing cancellation (at a fine structure dip or minimum) and the MOC reflex suppresses only one component, it will alter the phase relationship between them, consequently releasing phase cancellation. Thus, MOC activation in this scenario would result in a sudden increase in DPOAE level, i.e., enhancement. Conversely, if the DPOAE is measured at a frequency where the two components are in phase and adding constructively (at a peak or maximum in fine structure), a reduction in amplitude of either or both components by MOC activation will produce a reduction in ear canal DPOAE level, i.e., suppression. By controlling the phase relationship between components, it is possible to effectively address the greatest difficulty attributed to DPOAE-based measures of the MOC reflex and avoid instances of enhancement that are unrelated to MOC reflex strength.

Most earlier studies of the MOC reflex using the DPOAE-based CAS paradigm, including one from our own laboratory, did not control for the phase relationship between the two DPOAE sources. Generally, test frequency has been selected without regard for fine structure and, as a result, MOC findings cannot be interpreted without ambiguity. The purpose of the present study is to describe (1) a DPOAE-based measure of the MOC reflex that attempts to control the phase relationship between components by measuring only at fine structure peaks, and (2) the effects of CAS on the relationship between dual components contributing to the ear canal DPOAE.

II. METHODS

A. Subjects

Twenty-one potential adult subjects were screened for participation in this study. Of those, 15 were included and six were eliminated primarily due to low DPOAE level, history of noise exposure, elevated audiometric thresholds, and/or excessively high noise floor. Subjects were employees of the House Ear Institute (Los Angeles, CA) or friends of employees and were recruited via e-mail advertisement. They were screened for hearing thresholds $\leq$15 dB HL between 0.25 and 8 kHz and normal middle ear function, defined as a type A tympanogram with static compliance between 0.4 and 1.5 cm^3 and peak pressure between $\pm$150 daPa. Six left and nine right ears from 11 females and four males with a mean age of 31 years (range=21–43 years) were tested. All tests were conducted at the House Ear Institute in a sound-attenuated IAC booth. Each participant was tested twice with the same DPOAE protocol (trials 1 and 2). The two trials were separated by 30–45 min at most and did not typically include a refitting of the probe between trials.

B. Instrumentation and protocol

Signal generation and recording were controlled using custom software developed by Talmadge *et al.* (1999) and run on an Apple Macintosh G4 computer via a MOTU 828 Mk II (24 bits/44 100 Hz). Stimulus tones were presented to the subjects' ear canal via an Etymotic Research ER-10C DPOAE probe system. The output of the ER-10C microphone was preamplified and then passed through an analog high pass filter with 300 Hz cutoff frequency before being digitized by the MOTU and stored on disk.

DPOAE recordings were made between DP frequencies of ~500 Hz (f_2=782 Hz) and ~2500 Hz (f_2=3910 Hz) using stimulus levels of 65 (L_1) and 55 (L_2) dB sound pressure level (SPL) and a constant stimulus frequency ratio of (f_2/f_1) 1.22. The stimulus tones were swept at a rate of 8 s/octave, using a technique originally described by Long *et al.* (2004, 2008). Eight such sweeps were averaged in both no-CAS and +CAS conditions with conditions counterbalanced. Contralaterally presented broadband noise was generated by an external noise generator and presented at 60 dB SPL (unweighted) to all subjects, through an ER-2 insert phone. Preliminary work was conducted on ten normal-hearing adults to establish the optimal CAS level for MOC activation (between 30 and 70 dB SPL). In these ten subjects, 55–65 dB SPL produced maximum reductions in DPOAE level. Because 60 dB SPL is the recommended level used in most human studies of DPOAE contralateral suppression (Guinan, 2006) and it agreed with results of our preliminary testing, it was selected as the noise elicitor for all subjects in the current study. This CAS level is considered low enough that it does not evoke middle ear muscles in most cases (Gelfand, 1984) and high enough in level to produce MOC activation.

DPOAE level and noise-floor estimates were made using a least-square-fit algorithm (LSF) (Long *et al.*, 2008) and yielded estimates at every 2 Hz around 500–1000 Hz and every 6 Hz around 2500 Hz. The stimulus levels were calibrated and system distortion measured in a Zwislocki coupler. The two transducers in the ER-10C were individually equalized to produce flat constant drive voltage frequency responses up to 7000 Hz. System distortion was below −30 dB SPL for the stimulus levels used in these recordings.

A MATLAB-based analysis software (NIPR), also developed by Talmadge *et al.* (1999), uses an inverse fast-Fourier transform (IFFT) algorithm described in Dhar *et al.*, 2002 and Withnell *et al.*, 2003 to separate the DPOAE overlap and CF components based on their respective group delays. DPOAE complex amplitude measured in the frequency domain is multiplied by a moving Welch window (400 Hz window width) in 50 Hz step sizes. The IFFT converts each windowed data set into the time domain, after which a time-window filter is applied to extract the desired delay component in the time domain. These filtered windows of data are then transformed back to the frequency domain by FFT and the complex amplitudes of the overlap and CF components are reconstructed in the frequency domain.

C. Data analysis

1. Fine structure classification

a. Level. The median value for every three successive data points was computed to calculate DPOAE level and noise floor. Data points where the signal-to-noise ratio (SNR) between level and noise-floor medians was not at least 6 dB were eliminated. Although, the LSF procedure for estimating DPOAE level is a more efficient measure than provided by traditional FFT analyses and is less contaminated by the noise floor, a rather conservative 6 dB SNR criterion was used to ensure valid and reliable DPOAE estimates.

b. Maxima. The identification of maxima and measurement of fine structure spacing and depth were conducted with an automated algorithm developed by Dhar (Dhar and Abdala, 2007). Fine structure maxima were identified based on the first and second derivatives of the DPOAE level function and the relationship between them. Data points where the first derivative was equal to zero were identified as extrema (maxima and minima) and then further classified as a maximum or minimum based on the second derivative being negative (maxima) or positive (minima). Although maxima were identified using an automated algorithm, this process was checked manually by two observers familiar with the identification of DPOAE fine structure. This secondary level of analysis was implemented to eliminate minor ripples associated with noise that were erroneously identified by the program as maxima. These errors typically included peaks that were artificially tall and narrow and mimicked the noise-floor configuration in the low-frequency range. Sixty-five percent of the maxima eliminated during secondary analysis were in the frequency region below 1000 Hz. (Note: An updated analysis algorithm currently in use automatically eliminates maxima with less than 2.5 dB depth and fine structure spacing that is artificially narrow: $f/\Delta f > 25$.)

c. Prevalence. Prevalence of fine structure periods was calculated by counting maxima in 1/3 octave intervals.

d. Depth and spacing. Fine structure depth for each period was computed as $\mathrm{FS}_{\mathrm{depth}} = 20 \log 10\ (P_{\max}/P_{\mathrm{av(min)}})$, where $P_{\max}$ is the DPOAE level at a maximum and $P_{\mathrm{av(min)}}$ is the average DPOAE amplitude of the preceding and following minima. The frequency spacing of fine structure was computed both in Hz and as $f/\Delta f$, where f is the geometric mean between two adjacent minima and Δf is the frequency separation between them (Shera, 2003).

2. Component separation

The composite DPOAE level was separated into two components using an IFFT technique, as described in Sec. II. The two components are thought to reflect (a) the primary DPOAE generation source around f_2, corresponding to the point of maximum overlap between the two stimulus tones (overlap component); and (b) a source at $2f_1 - f_2$, the DP CF component.

3. Quantifying the MOC effect

The MOC effect was quantified at frequencies corresponding to fine structure maxima only. Two indices of the MOC effect were calculated.

(a) *MOC reflex.* This is the difference between DPOAE level at a given maxima frequency in the no-CAS condition and, at the same frequency, in the +CAS condition. Only the MOC reflex values that reflected DPOAE level suppression were included in analyses although the distribution of level enhancement evoked by CAS was reported as well.

(b) *MOC shift.* This is a shift in peak frequency observed between a DPOAE maximum in the no-CAS condition and the same maxima during CAS. A frequency ratio value, $f/\Delta f$, was calculated to reflect this frequency shift, similar to the index used in the analysis of fine structure spacing. In this case, f=peak frequency in the no-CAS condition and Δf=the difference between this reference frequency and frequency of the corresponding maxima in the +CAS condition. Only frequency shifts toward higher frequencies were analyzed but the distribution of both low- and high-frequency shifts was reported.

4. Statistical analyses

DPOAE level (including composite level and separate overlap and CF component levels), prevalence, depth, and spacing were calculated for each individual and compared across frequency, trial, and CAS condition in multiway repeated measures analyses of variance (ANOVAs) using the statistical computing package, STATVIEW (SAS Institute). MOC reflex and shift were calculated for each individual and compared across frequency and trial. For analysis and display of DPOAE level only, f_2 frequency was plotted on the abscissa, as is typical in much of literature. Most DP-grams (DP level × frequency), as they are referred to in clinical nomenclature, are plotted as a function of f_2 or the geometric mean of f_1 and f_2 since this is considered the frequency/cochlear region being evaluated. We wish to remain consistent with this convention and produce a recognizable graph of DPOAE level. All fine structure, DPOAE component, and MOC-related variables were displayed and analyzed as a function of DP frequency $(2f_1 - f_2)$.

Variables were averaged into discrete frequency bands for analysis and display. For each subject, DPOAE level (in dB SPL) was averaged into four f_2 frequency bands centered at 1000, 2000, 3000, and 4000 Hz, corresponding to 2 mm distance on the basilar membrane as per Greenwood's map of frequency (1 mm on each side of the center frequency

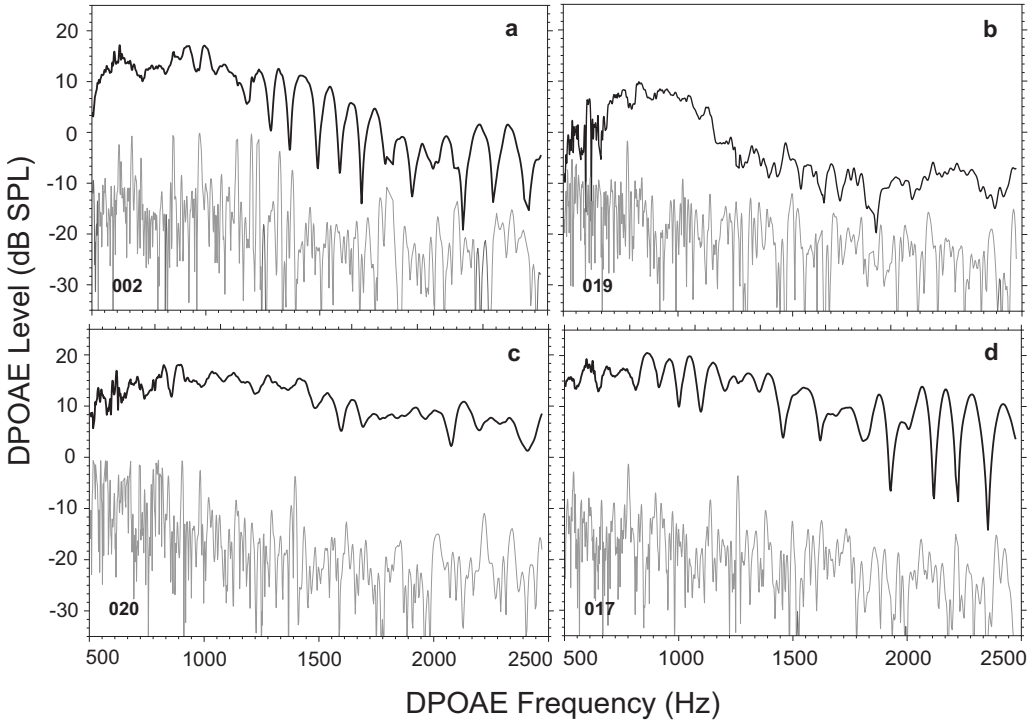


FIG. 1. DPOAE fine structure and noise floor from four randomly selected adult subjects. Panels (a) and (d) are from adults with strong fine structure and pronounced minima. Panels (b) and (c) in contrast show relatively poor-to-fair fine structure, with more shallow minima and more poorly defined peaks.

except for 4000 Hz, which had reduced range on the high-frequency end). DPOAE fine structure, component, and MOC variables were averaged into the following four 500-Hz-wide frequency bands for analysis: 500–1000, 1000–1500, 1500–2000, and 2000–2500 Hz. In all figures, these frequency bands are denoted by a value representing the upper limit of the frequency interval.

III. RESULTS

A. DPOAE fine structure

Prior to analyzing any effects of CAS on DPOAEs, initial analyses quantified fine structure characteristics from all subjects in the no-CAS condition. Figure 1 (a)–(d) show fine structure and corresponding noise floor from four randomly chosen adult subjects. Panels (a) and (d) show extremely deep, pronounced, strong fine structure throughout the frequency range. Panel (b) shows relatively poorly defined fine structure and low DPOAE levels beyond about 1000 Hz. Panel (c) shows moderate fine structure throughout the frequency range.

Mean *DPOAE level* ranged from an average of −3 dB SPL at 4000 Hz to a high of 11 dB SPL at 1000 Hz. Figure 2 shows mean level at each f_2 center frequency for trials 1 and 2 separately. A repeated measures ANOVA of DPOAE level (trial × f_2 frequency) showed a significant effect of frequency ($f=117.87$; $p=0.0001$) and no effect of trial. DP level decreased with increasing f_2 frequency.

Prevalence of fine structure maxima was calculated in six 1/3 octave intervals: 630–794, 794–1000, 1000–1260, 1260–1587, 1587–2000, and 2000–2500 Hz (not a full 1/3 octave interval). The corresponding prevalence of fine structure periods for these six DP frequency ranges was 0.9, 1.6, 2.5, 2.5, 3.2, and 2.8 with an average of 2.25 periods per 1/3 octave interval and an overall average of 13.5 periods over the entire approximately two-octave range analyzed. A repeated measures ANOVA (trial × DP frequency) showed a significant effect of frequency ($f=27.8$; $p=0.0001$) with no effect of trial. The prevalence of DPOAE fine structure periods increased with DP frequency and peaked in the mid- to high-frequency range (1587–2500 Hz).

Fine structure spacing (in Hz) ranged from 65 at 1000 Hz to 125 at 2500 Hz. These values were converted into an $f/\Delta f$ ratio for analysis and averaged into four DPOAE frequency bands with upper limits of 1000, 1500, 2000, and 2500 Hz. The overall mean ratio was 16.4, consistent with estimates of spontaneous OAE spacing in normal-hearing

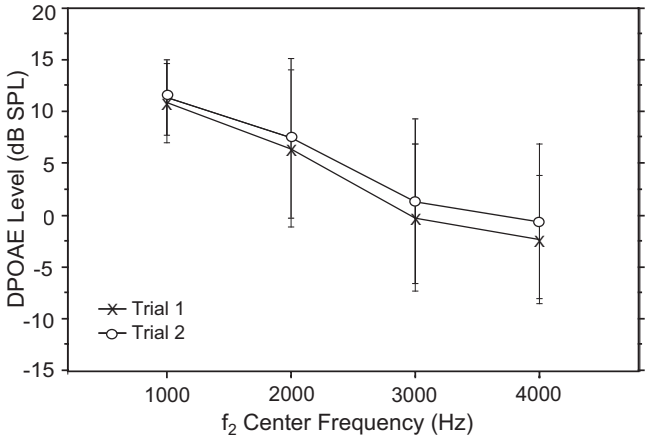


FIG. 2. Mean DPOAE level at four f_2 center frequencies in 15 normal-hearing adults. Data were collected for each subject during two separate runs (trials 1 and 2) within the same test session. Error bars reflect ±1 s.d.

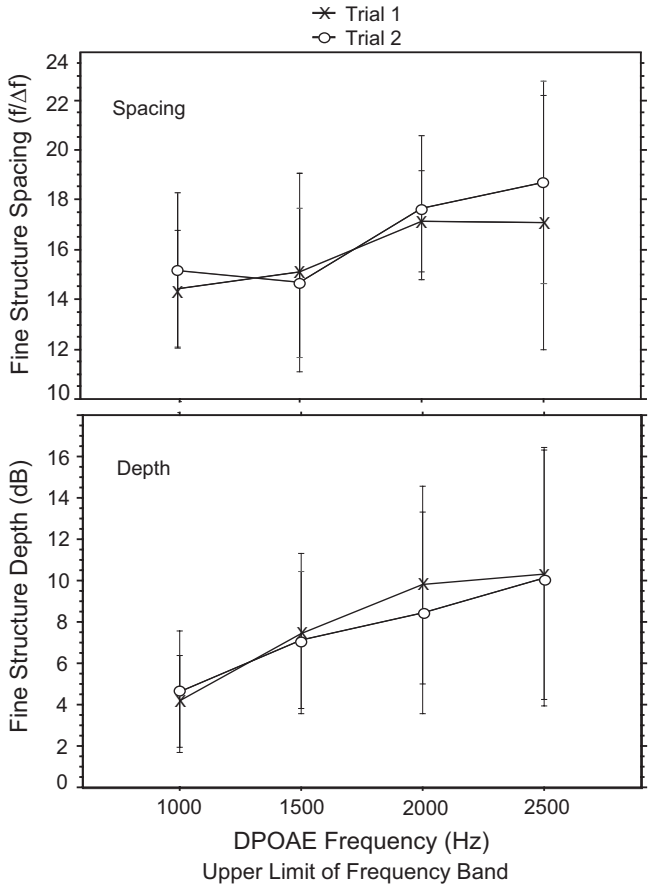


FIG. 3. Mean DPOAE fine structure spacing ratio and depth as a function of DP frequency in 15 adults. Data were collected for each subject during two separate runs (trials 1 and 2) within the same test session. Error bars reflect ±1 s.d.

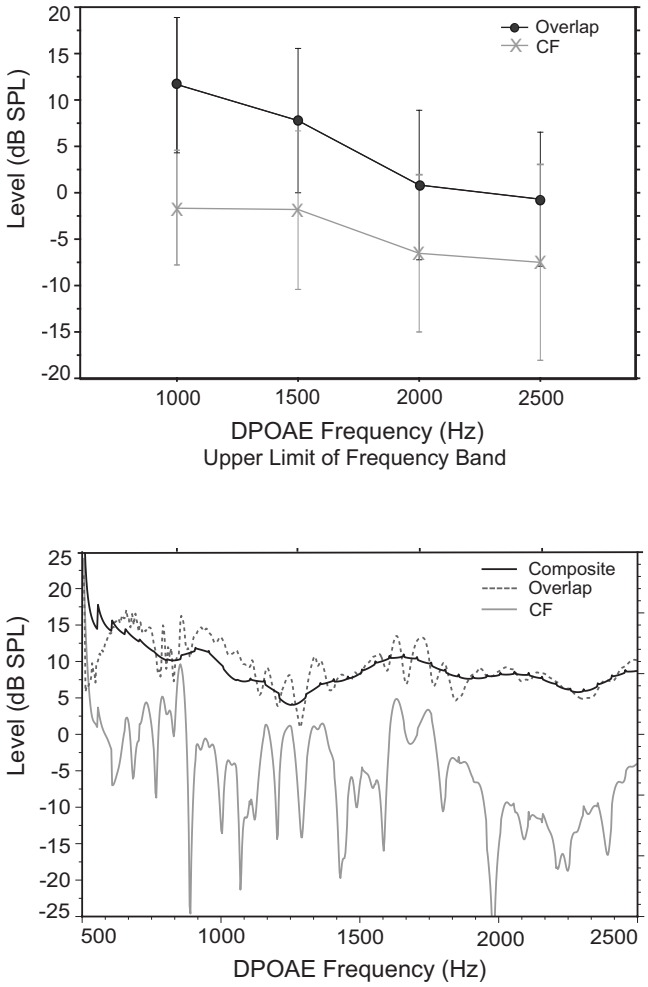


FIG. 4. Upper panel: Mean DP component level (overlap and CF) as a function of DP frequency in 15 adults. Data from trials 1 and 2 are combined. Lower panel: IFFT level output from one adult subject showing composite DPOAE level recorded in the ear canal, as well as overlap and CF component levels.

adult subjects (Shera, 2003). Mean spacing ratios ranged from 14.8 at 1000 Hz to 18.0 at 2500 Hz (Fig. 3, upper panel). A repeated measures ANOVA found no effect of trial on spacing but a significant effect of frequency ($f=7.0$; $p=0.0003$).

The *depth* of each fine structure period ranged from a mean of 4 dB at 1000 Hz to 10.3 dB at 2500 Hz (Fig. 3, lower panel). A repeated measures ANOVA of depth (frequency × trial) showed a significant effect of frequency on fine structure depth ($f=10.9$; $p<0.0001$) but no effect of trial. Fine structure periods became deeper as frequency increased.

Once the composite DPOAE was separated into two components using IFFT, DP level from both the overlap and CF regions was analyzed with a three-way repeated measures ANOVA (component × frequency × trial). This analysis sought to determine if response level was different between components and whether this difference was frequency or trial dependent. There was a main effect of component ($f=116$; $p<0.0001$) and frequency ($f=94$; $p<0.0001$) and an interaction between the two ($f=8.9$; $p<0.0001$). There was no effect of trial. The upper panel of Fig. 4 shows mean DP component levels (summed over trial) as a function of DPOAE frequency. Overlap component levels were higher than CF component levels by a mean value of 9.3 dB. Additionally, there was a general decrease in level as DPOAE frequency increased, for the overlap component only. The lower panel of Fig. 4 shows complete IFFT level output for one subject. Although much greater frequency resolution is apparent in this IFFT output graph, the same pattern is evident in both panels; the overlap component is higher in level than the CF component across frequency.

B. The effects of CAS on DPOAE fine structure

Because data from trials 1 and 2 were indistinguishable in previous analyses, they were combined here. Prevalence of DPOAE fine structure periods did not change with the presentation of CAS (CAS condition × frequency ANOVA). A similar test was conducted on *spacing* ($f/\Delta f$) and showed an effect of CAS ($f=14.6$; $p=0.0008$) and an effect of frequency ($f=4.8$; $p=0.007$). The upper panel of Fig. 5 depicts spacing ratio as a function of DPOAE frequency. As noted in Fig. 5, an interaction between CAS condition and frequency was also observed ($f=2.7$; $p=0.04$). *Post hoc* t-tests were conducted to further scrutinize the interaction using a Bonferoni adjusted alpha level (4 contrasts/0.05=0.012) and showed that the effect of CAS on fine structure spacing was only present in the lowest-

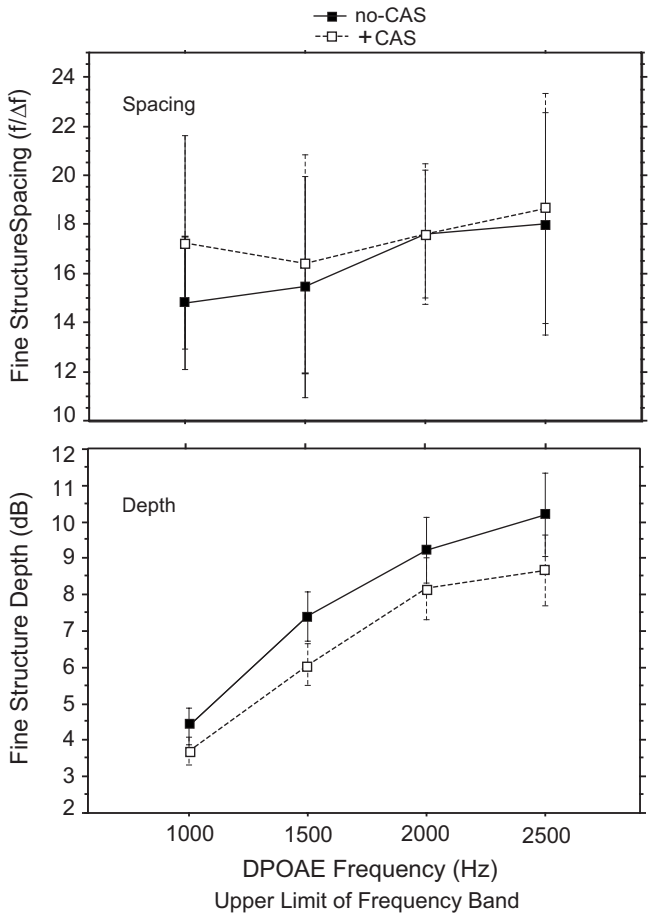


FIG. 5. Mean DPOAE fine structure spacing ratio (upper panel) and depth (lower panel) as a function of DP frequency in 15 adult subjects. The parameter is absence or presence of CAS. The filled squares represent no-CAS and the open squares represent +CAS. Trials 1 and 2 data are combined. Error bars reflect ±1 s.d.

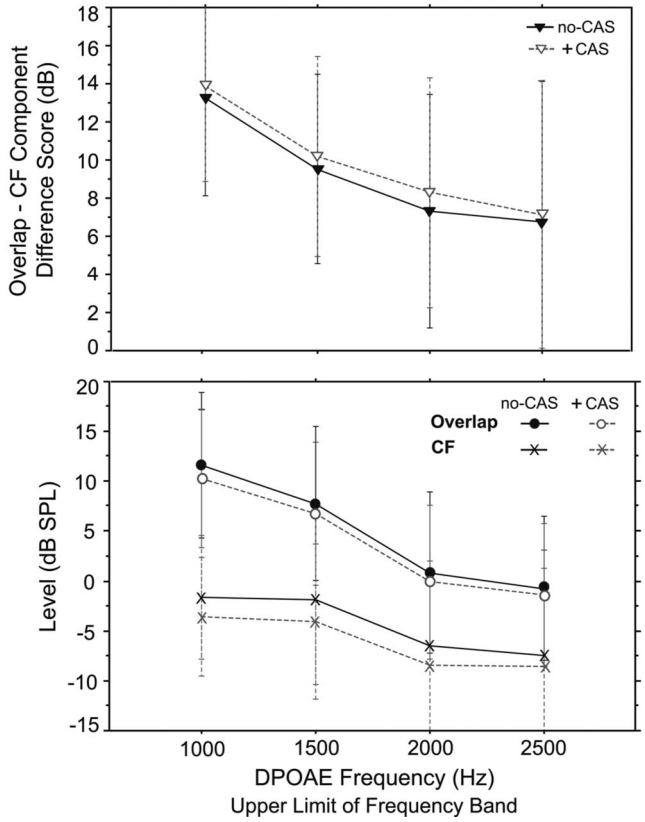


FIG. 6. Upper panel: Mean difference scores reflecting the level difference between overlap and CF components as a function of DP frequency. The filled triangles represent no-CAS and the open triangles represent +CAS. Lower panel: Mean level for the overlap and CF components in no-CAS and +CAS conditions. Error bars reflect ±1 s.d.

frequency band: 500–1000 Hz ($p=0.0003$). The mean spacing ratio in the low-frequency band was 14.7 (s.d.=2.72) in the no-CAS condition and 17.2 (s.d.=4.25) when CAS was presented. It may be noteworthy that the calculated average spacing value in this low-frequency interval was typically based on many fewer individual observations than averages in the other three frequency bands, due to a higher noise floor.

A repeated measures ANOVA (CAS condition × DP frequency) was conducted on fine structure depth and showed that depth values decreased and became more shallow when CAS was presented ($f=32.8$; $p<0.0001$). In the absence of CAS, overall mean depth across all frequencies and subjects was 7.45 dB and with CAS presented, mean depth was 6.45 dB (Fig. 5, lower panel).

Previous analyses showed a significant difference between DP component levels (overlap level > CF level). The effect of CAS on this difference was analyzed by computing a difference score between component levels and testing it with a repeated measures ANOVA (CAS condition × frequency). Results showed an effect of both CAS ($f=5.8$; $p=0.02$) and DP frequency ($f=9.7$; $p<0.0001$) on the component level difference score. (See Fig. 6, upper panel.) The presentation of CAS increased the level difference between components by reducing CF levels more than overlap component levels. (See Fig. 6, lower panel.) MOC activation produced a mean reduction of 1.2 dB in the level of the overlap component, compared to 1.8 dB for the CF component level.

C. The MOC effect

1. Reliability

Prior to quantifying the MOC reflex, initial tests were conducted to establish the reliability of DPOAE level changes attributed to CAS. A three-way repeated measures ANOVA (CAS condition × frequency × trial) was conducted on DPOAE level. However, unlike the MOC reflex that was quantified at maximum frequencies only, this initial measure was conducted with all level estimates ($n\sim900$) across the frequency range, regardless of their position in the fine structure pattern. The ANOVA showed a main effect of noise on DPOAE level ($f=37.1$; $p<0.0001$) but no main effect of trial ($p=0.66$). Overall, the mean difference between DPOAE level recorded during trials 1 and 2 was 0.24 dB, whereas the overall difference in DPOAE level between no-CAS and +CAS noise conditions was 0.9 dB. These results suggest that changes in DPOAE level produced by CAS exceeded natural changes in DPOAE level that occur for any given subject during multiple trials. Thus, the test-retest re-

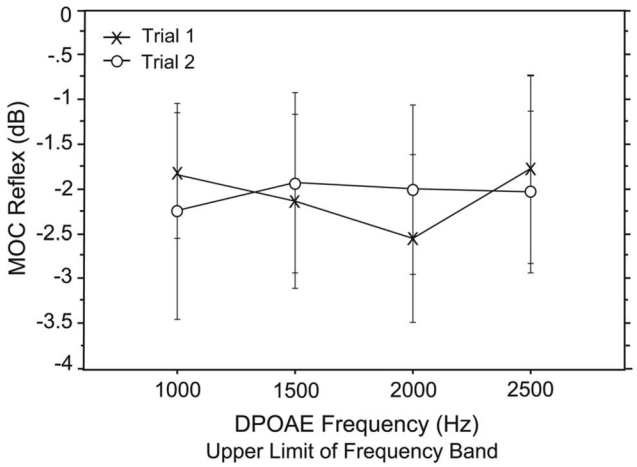


FIG. 7. Mean MOC reflex (defined as the difference between DPOAE level in no-CAS and +CAS conditions) calculated at DPOAE fine structure maximum frequencies only and shown as a function of DP frequency. Data were collected during two separate trials conducted within the same test session for 15 adults. Error bars reflect ±1 s.d.

liability of the MOC reflex is high and changes in DPOAE level attributed to CAS cannot be easily assigned to random intrasubject variability.

2. MOC reflex

When the MOC reflex was measured at only peak frequencies, 97% of data points reflected suppression. In contrast, if the MOC reflex was measured at all frequencies—not just fine structure maxima—only 68% of data points across all subjects and trials reflected level suppression and 32% reflected no change or level enhancement. MOC-induced suppression ranged from −1.93 dB at 1000 Hz to −2.47 dB at 2000 Hz with a mean of −2.05 dB as shown in Fig. 7. Figure 8 shows an example of DPOAE fine structure with and without CAS for four adult subjects. As is evident from this figure, very little response enhancement was observed at peak frequencies. Figure 9 shows fine structure in no-CAS and +CAS conditions for the same subjects shown in Figs. 8(a) and 8(d), but with the MOC reflex superimposed (and referenced to the right axis) to illustrate these findings in more detail. At fine structure minima, there was most often no change or DPOAE level enhancement with CAS, as evident from sharp spikes in the MOC reflex toward positive values around dips in fine structure. Conversely, around maxima, consistent decreases in DPOAE level were noted. No effect of trial or DP frequency was observed on the MOC reflex.

3. MOC shift

As can be seen in Fig. 8, CAS also produced a shift in fine structure peak frequency. When peaks shifted, they shifted upward in frequency 80% of the time. The shift ranged from 5.9 Hz at 1000 Hz to 10.6 Hz at 2000 Hz (Fig. 10). There was no effect of trial on MOC shift but a significant effect of frequency ($f=4.4$; $p=0.007$); shift increased with increasing DPOAE frequency. As a secondary analysis the shift was quantified as a ratio ($f/\Delta f$), the alpha level was adjusted, and the ANOVA was repeated with no significant effect of trial or frequency.

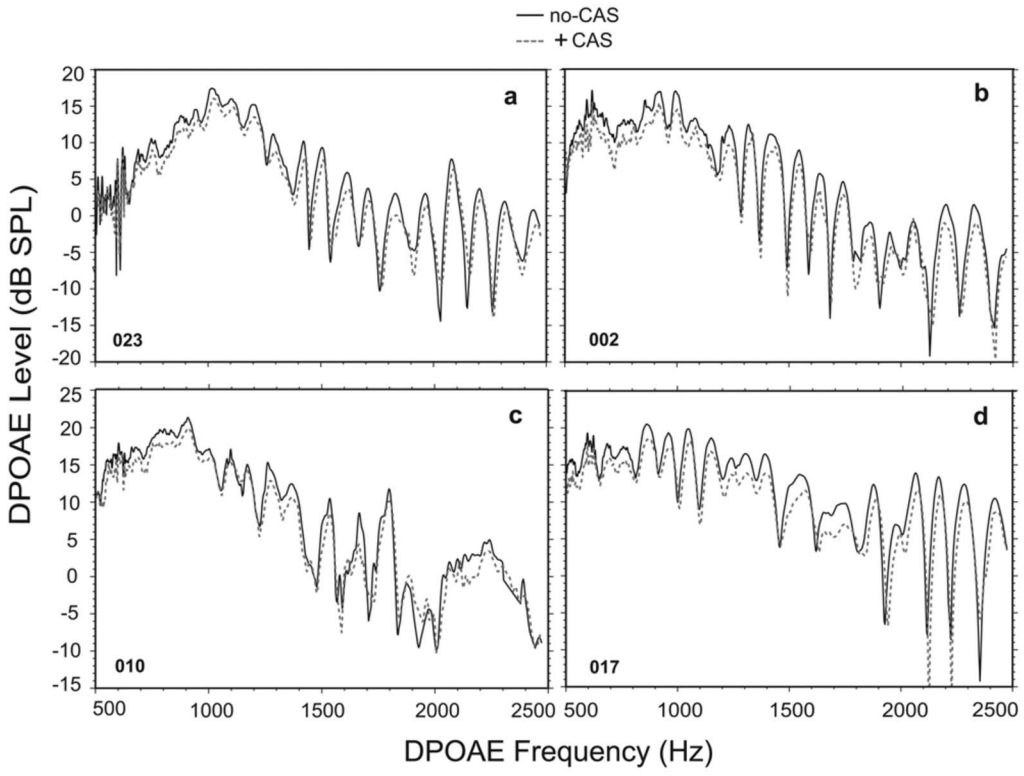


FIG. 8. DPOAE fine structure for four adult subjects with no-CAS, which is depicted with a solid black line, and in +CAS, which is depicted by a dashed gray line.

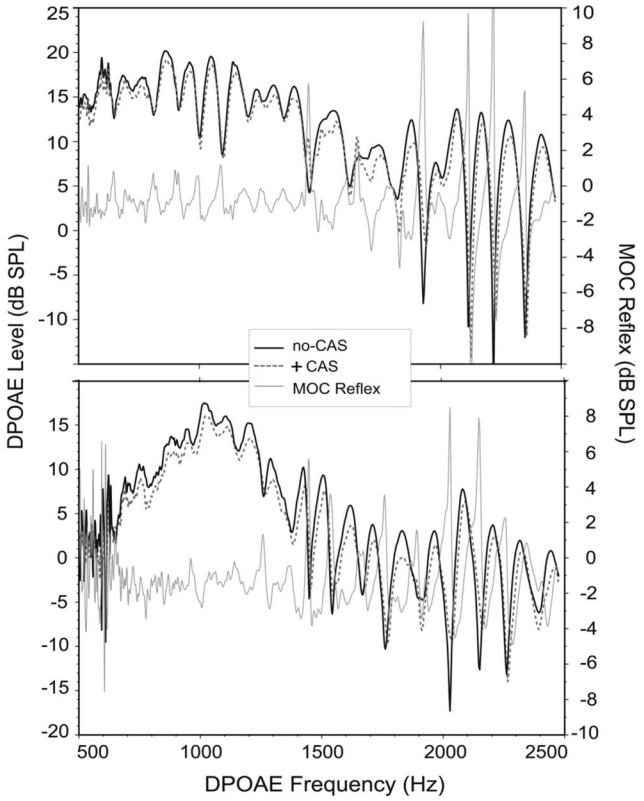


FIG. 9. DPOAE fine structure from the two adult subjects also shown in panels (d) and (a) of Fig. 8. The no-CAS condition is depicted with a solid black line and +CAS with a dashed line. The corresponding MOC reflex (gray solid line) is superimposed on fine structure and referenced to the right axis.

IV. DISCUSSION

A. Normative fine structure

The data from these normal-hearing, healthy young adults collected using a swept tone paradigm are comparable to normative fine structure data previously reported in literature (Reuter and Hammershøi, 2006). Reuter and Hammershøi (2006) reported two to three "ripples" or fine structure periods in each 1/3 octave interval and the present study reports an average of 2.25 per 1/3 octave. Additionally, Reuter and Hammershøi (2006) reported fine structure depth to be between 3 and 9 dB, depending to some extent on frequency, and we reported mean depth values of between 4 and 10 dB, with increasing depth in the mid- to high-frequency ranges.

The absolute spacing values reported in the present study are generally consistent with Reuter and Hammershøi (2006) and hover around 100–125 Hz in the midfrequency ranges. This is a bit lower than other reports, considering some have identified average spacing intervals of 157 Hz from 1.8–2.6 Hz (Engdahl and Kemp, 1996) and 180 Hz between 2000 and 4000 Hz. The differences in absolute spacing intervals may be due to slight differences in the algorithms used to calculate spacing. Additionally, the swept tone recording paradigm has much greater frequency resolution than those used in past studies and, therefore, this approach may have been able to define spacing intervals with greater accuracy. As expected, larger spacing intervals with increased DPOAE frequency were observed in the present study, consistent with past reports (Reuter and Hammershøi, 2006; He and Schmiedt, 1993; Engdahl and Kemp, 1996; Heitmann *et al.*, 1996). Changes in the spacing ratio index as a function of frequency were also observed and are somewhat surprising considering the presumed scaling symmetry along the basilar membrane. The observed increases in spacing ratio with increases in DPOAE frequency may suggest deviations from scaling symmetry and warrant further investigation to elucidate the source of this finding.

In addition to showing good agreement with the most current normative fine structure data, these findings also suggest that two trials from the same individual produce DPOAE level (component and composite) and fine structure features that do not differ significantly. This suggests stable fine structure within a subject. The two trials occurred within the same 2 h session and did not include a probe refitting but did include normal head and body movements during testing breaks. Reuter and Hammershøi (2006) also studied the repeatability of DPOAE fine structure in normal-hearing adults and found generally stable intrasubject fine structure patterns both within session and 4 weeks after the initial test.

B. Effects of CAS on DPOAE fine structure and components

Few reports exist on the effects of CAS (i.e., MOC activation) on DPOAE fine structure or individual DP components. In this study, effects of CAS on fine structure spacing were observed only in the low-frequency band (500–1000 Hz). However, as noted in Sec. III, the spacing average in this frequency band was based on only half the number of individual values compared to the other frequency-band averages. It is possible that this may have produced excessive variability in the estimate and influenced the result. Only one other study has addressed this issue and reported no change in periodicity of fine structure with presentation of contralateral noise (Sun, 2005). However, the data were reported on limited subjects, not analyzed statistically and unpublished, making comparison tenuous. We also found that the presentation of CAS made DPOAE fine structure more shallow (on average by 1 dB), which is consistent with results reported by Sun (2005).

The data shown in Fig. 6 show that CAS reduced the CF component level more than it reduced overlap levels, which suggests that the MOC reflex differentially affects the two DPOAE sources. This likelihood has been suggested by various researchers (Maison and Liberman, 2000; Müller *et al.*, 2005; Sun, 2008b) and is not unexpected. The MOC reflex is thought to act more strongly on the basilar membrane response to low-level stimulation. Consistent with this, indices of MOC function are often reported to be most robust at low stimulus levels (Abdala *et al.*, 1999; Moulin *et al.*, 1993; Janssen *et al.*, 2003). The component at DP CF is considered a low-level reflection source associated with cochlear amplifier function. The low-level nature of this component may make it particularly sensitive to changes in basilar membrane vibration and, thus, more susceptible to MOC activation. Others have hypothesized this same effect and have suggested that CAS should have its most significant

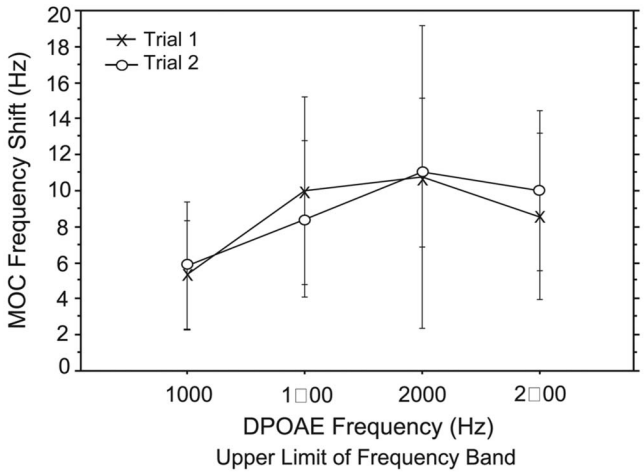


FIG. 10. Mean MOC shift (defined as an upward shift in fine structure peak frequency with the presentation of CAS) as a function of DP frequency for 15 adult subjects. Error bars reflect ±1 s.d.

effect on the reflection component of the DPOAE (Long *et al.*, 2004). Our findings are consistent with this hypothesis.

C. The MOC reflex at fine structure maxima

It is highly unlikely that the middle ear muscle reflex was activated or can account for DPOAE suppression evoked by 60 dB SPL of contralateral noise. A typical clinical instrument has a middle ear muscle threshold of 75 dB SPL for broadband noise (Gelfand, 1984) whereas a more sensitive piece of research equipment may detect thresholds around 65 dB SPL (Guinan *et al.*, 2003; Guinan, 2006). Thus, presenting a 60 dB SPL noise elicitor of MOC activity is unlikely to produce interference from middle ear muscle reflexes. An important caveat is that studies defining the middle ear muscle (MEM) reflex threshold to broadband noise cannot fully represent the paradigm used here since pure tones (f_1 and f_2) are not presented simultaneously to the opposite ear. This bilateral stimulation could lower MEM reflex threshold. It is not possible to be unequivocally certain that MEM reflexes are not activated with this combination of stimuli. A recent study argues against the likelihood of MEM reflex interference. Sun (2008a) found that contribution from the acoustic reflex to measures of DPOAE contralateral suppression is negligible, if present at all, when CAS is presented below the acoustic reflex threshold. Adult subjects in Sun's (2008a) study had a median reflex threshold for broadband noise of 85 dB SPL (range=70–100 dB SPL) and even when contralateral noise was presented at high enough levels to elicit the MEM reflex, the added effect on DPOAE level recorded in the opposite ear was insignificant.

In the present study, overall MOC-induced suppression averaged −2.05 dB (mean s.d.=0.965) when measured only at fine structure maxima. The magnitude of the MOC reflex recorded in the present study is somewhat larger than in our past work. Abdala *et al.* (1999) measured DPOAE contralateral suppression in normal-hearing adults without considering fine structure or the phase relationship between components and found MOC reflexes averaging 1.2 dB across frequency. Additionally, we reported greater intersubject variability (larger standard deviations) with respect to MOC reflex estimates than was observed here. Moulin *et al.* (1993) reported contralateral suppression ranging from 0.68 to 2.0 dB in normal-hearing adults, while Williams and Brown (1995) described mean values ranging from 1.1 to 1.9 dB in adult subjects. Thus, in the present study, by recording the MOC reflex at peak frequencies in fine structure only, the magnitude of the reflex was at least equal to and generally more robust than earlier studies that did not consider DPOAE fine structure. Additionally, the MOC reflex recorded in this study was a less variable assay than reflex measures reported in past studies, consistent with recent findings reported by Sun (2008b). Because episodes of enhancement were almost completely eliminated, the large swings between negative and positive reflex values were not present in our recordings and, therefore, variability was constrained. If all DPOAE frequencies spanning the entire test range of 500–2500 Hz (not just peak frequencies) were considered, DP level enhancement was observed in one-third of all data points. Müller *et al.* (2005) similarly found that 31% of their data points reflected enhancement of DPOAE level when measured at dips in fine structure. By eliminating these observations, we have lessened variability associated with estimates of the MOC reflex.

D. The "true" MOC reflex

Our results indicate that MOC activation does not suppress each component equally, explaining why DPOAE level will sometimes increase with CAS. In past studies of MOC activation in humans, including one from our own laboratory, target frequencies were selected without regard for DPOAE fine structure. Thus, when recording the MOC reflex at a given frequency, researchers were not certain of the phase relationship between the primary components contributing to the DPOAE. As briefly considered in Sec. I, if the DP components were out of phase at the test frequency selected, and one of the components was influenced by MOC activation more than the other, the phase relationship between them could shift, producing a release of cancellation. The result is often a marked enhancement of DPOAE level. DPOAE level enhancement around fine structure minima only is clearly noted in Fig. 9 and typifies the data collected in this study and reported elsewhere (Sun, 2005; Sun, 2008b; Zhang *et al.*, 2007). Like the present study, Sun (2008b) reported that CAS reduces DPOAE level at peak frequencies in fine structure, while producing little change or even enhancement in level at minima. Zhang *et al.* (2007) similarly found larger suppression values at fine structure peaks versus dips and reported less variability in measures of the MOC reflex when recorded at maxima.

In contrast to the strategy proposed here, that MOC effects be measured at DPOAE fine structure peak frequencies only, some investigators have specifically targeted fine structure dips for measurement of the MOC reflex. These investigators argue that a "bipolar" MOC effect, including both suppression and enhancement values, will be more robust than suppression alone. Müller *et al.* (2005) recorded DPOAEs at dip frequencies while systematically presenting

varied L_1/L_2 combinations in normal-hearing adults. With the presentation of CAS, they found an overall mean swing in DPOAE level from suppression to enhancement of 14 dB with small changes in primary tone level separation. Müller *et al.* (2005) suggested that this bipolar MOC effect was a tool well suited to measuring the strength of the MOC reflex in humans because it provided larger reflex values. Wagner *et al.* (2007) similarly recommended measuring MOC reflex only at dips in DPOAE fine structure because of these large bidirectional effects.

Results of the present study argue strongly against recording large bidirectional swings in DPOAE level and suggest that these indices do not accurately reflect MOC reflex strength. As primary tone level separation is altered and/or CAS is presented, the phase relationship between dual sources comprising the ear canal DPOAE shifts. The large, abrupt swings between suppression and enhancement most likely represent the changing phase relationship between these two components, not the strength of the MOC reflex. For this very reason, the present study chose to control for these shifts in component phase relationship by recording the MOC reflex only at fine structure peak frequencies. By choosing a maximum frequency, it was certain that both components were in a constructive relationship. Thus, any alteration in either component by CAS would produce a reduction in amplitude, as reliably shown for 97% of MOC reflex data points in this study. This stable reduction in amplitude of around 2 dB, though smaller than the large bidirectional effects, is reliable and more strongly linked to underlying MOC activation. The fact that we were able to eliminate DPOAE level enhancement almost completely by controlling the phase relationship between components during activation of the MOC reflex supports our contention that shifts in phase produce the large bipolar swings reported in literature. An additional advantage of measuring the MOC reflex at peak frequencies only is that DPOAE level is always significantly higher at maxima than at minima; thus, signal-to-noise ratio is better.

Finally, we have detected another effect of MOC activation on DPOAE fine structure that may serve as an additional index of MOC integrity. A mean frequency shift of 5–10 Hz was observed for some fine structure peaks with CAS. This effect has been reported in previous literature (Sun 2005; Sun, 2008b; Mauermann and Kollmeier, 2004) but not well defined or quantified. Future studies are warranted to further elucidate the mechanism and biological significance of this frequency shift.

V. CONCLUSIONS

The results of this study suggest that recording DPOAE contralateral suppression as an index of the MOC reflex is most effective at fine structure peak frequencies to exert some control over acoustic mixing of DP components. This mixing may interfere with accurate estimates of MOC reflex strength. When measured at peak frequencies only, episodes of DPOAE level enhancement were eliminated and 97% of all data points showed suppression. This is unlike past reports, that used the DPOAE-based CAS paradigm with no regard for fine structure and found DPOAE level enhancement in one-third of all observations. Findings in the present study suggest that when measured at DPOAE fine structure maxima, moderate levels of CAS reliably reduce the DPOAE by approximately 2 dB. We argue that this reduction in level more strongly reflects MOC activation and minimizes the effect of spurious phase shifts between components in the ear canal. Reducing and/or controlling for unwanted sources of variance moves us toward the noteworthy goal of developing a diagnostic assay of the human MOC reflex to assess integrity of the auditory system.

ACKNOWLEDGMENTS

This research was supported by the National Institutes of Health, NIDCD R01 DC003552, the A. Charles Holland Foundation, and the House Ear Institute. The authors would like to thank Dr. Sumitrajit Dhar for his input both during the analysis of these data and during the writing of this manuscript.

Abdala, C., Ma, E., and Sininger, Y. S. (**1999**). "Maturation of medial efferent system function in humans," J. Acoust. Soc. Am. **105**, 2392–2402.

Agrama, M. T., Waxman, G. M., Stagner, B. B., Martin, G. K., and Lonsbury-Martin, B. L. (**1998**). "Effects of efferent activation on distortion-product otoacoustic emissions in normal humans using ipsilateral acoustic stimulation," 21st Annual Midwinter Meeting of the Association for Research in Otolaryngology, Abstract 605.

Bassim, M. K., Miller, R. L., Smith, D. W., and Buss, E. (**2003**). "Rapid adaptation of the $2f_1$-f_2 DPOAE in humans: Binaural and contralateral stimulation effects ," Hear. Res. **182**, 140–152.

Collet, L., Kemp, D. T., Veuillet, E., Duclaux, R., Moulin, A., and Morgon, A. (**1990**). "Effect of contralateral auditory stimuli on active cochlear micro-mechanical properties in human subjects," Hear. Res. **43**, 251–261.

Dhar, S., and Abdala, C. (**2007**). "A comparative study of distortion-product-otoacoustic-emission fine structure in human newborns and adults with normal hearing," J. Acoust. Soc. Am. **122**, 2191–2202.

Dhar, S., Talmadge, C. L., Long, G. R., and Tubis, A. (**2002**). "Multiple internal reflections in the cochlea and their effect on DPOAE fine structure," J. Acoust. Soc. Am. **112**, 2882–2897.

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.

Gelfand, S. A. (**1984**). "The contralateral acoustic-reflex threshold," in *The Acoustic Reflex*, edited by S. Silman (Academic, Orlando, FL), pp. 137–186.

Guinan, J. J., Jr. (**2006**). "Olivocochlear efferents: Anatomy, physiology, function, and the measurement of efferent effects in humans," Ear Hear. **27**, 589–607.

Guinan, J. J., Jr., Backus, B. C., Lilaonitkul, W., and Aharonson, V. (**2003**). "Medial olivocochlear efferent reflex in humans: Otoacoustic emission (OAE) measurement issues and the advantages of stimulus frequency OAEs," J. Assoc. Res. Otolaryngol. **4**, 521–540.

He, N., and Schmiedt, R. A. (**1993**). "Fine structure of the $2f_1$-f_2 acoustic distortion product: Changes with primary levels ," 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 the objective audiometry as the result of fine structure," Eur. Arch. Otorhinolaryngol. **253**, 167–171.

Janssen, T., Gehr, D. D., and Kevanishvilli, Z. (**2003**). "Contralateral DPOAE suppression in humans at very low sound intensities," in *Biophysics of the Cochlea: From Molecules to Models*, edited by A. W. Gummer (World Scientific, London), pp. 498–505.

Kim, D. O. (**1980**). "Cochlear mechanics: Implications of electrophysiological and acoustical observations," Hear. Res. **2**, 297–317.

Kim, D. O., Dorn, P. A., Neely, S. T., and Gorga, M. P. (**2001**). "Adaptation of distortion product otoacoustic emissions in humans," J. Assoc. Res. Otolaryngol. **2**, 31–40.

Liberman, M. C., Puria, S., and Guinan, J. J., Jr. (**1996**). "The ipsilaterally

evoked olivocochlear reflex causes rapid adaptation of the $2f_1$-f_2 distortion product otoacoustic emission ," J. Acoust. Soc. Am. **99**, 3572–3584.
Long, G. R., Lee, J., and Talmadge, C. (**2004**). "Using sweeping tones to evaluate DPOAE fine structure," 27th Annual Midwinter Meeting of the Association for Research in Otolaryngology, Abstract 1076.
Long, G. R., Talmadge, C. L., and Lee, J. (**2008**). "Measuring distortion product otoacoustic emissions using continuously sweeping primaries," J. Acoust. Soc. Am. **124**, 1613–1625.
Maison, S. F., and Liberman, M. C. (**2000**). "Predicting vulnerability to acoustic injury with a noninvasive assay of olivocochlear reflex strength," J. Neurosci. **20**, 4701–4707.
Mauermann, M., and Kollmeier, B. (**2004**). "Distortion product otoacoustic emission (DPOAE) input/output functions and the influence of the second DPOAE source," J. Acoust. Soc. Am. **116**, 2199–2212.
Moulin, A., Collet, L., and Duclaux, R. (**1993**). "Contralateral auditory stimulation alters acoustic distortion products in humans," Hear. Res. **65**, 193–210.
Müller, J., Janssen, T., Heppelmann, G., and Wagner, W. (**2005**). "Relationship between fine-structure, contralateral suppression, and ipsilateral adaptation of distortion product otoacoustic emissions in humans," J. Acoust. Soc. Am. **118**, 3747–3756.
Puel, J. L., and Rebillard, G. (**1990**). "Effects of contralateral sound stimulation on the distortion product $2f_1$-f_2: Evidence that the medial efferent system is involved," J. Acoust. Soc. Am. **87**, 1630–1635.
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.
Shera, C. A. (**2003**). "Mammalian spontaneous otoacoustic emissions are amplitude-stabilized cochlear standing waves," J. Acoust. Soc. Am. **114**, 244–262.
Shera, C. A., and Guinan, J. J., Jr. (**1999**). "Evoked otoacoustic emissions arise by two fundamentally different mechanisms: A taxonomy for mammalian OAEs," J. Acoust. Soc. Am. **105**, 782–798.
Sun, X. M. (**2005**). "Distortion product otoacoustic emission (DPOAE) fine structure is responsible for the variability of DPOAE contralateral suppression effect," 25th Annual Midwinter Meeting of the Association for Research in Otolaryngology, Abstract 1327.
Sun, X. M. (**2008a**). "Contralateral suppression of distortion product otoacoustic emissions and the middle-ear muscle reflex in human ears," Hear. Res. **237**, 66–75.
Sun, X. M. (**2008b**). "Distortion product otoacoustic emission fine structure is responsible for variability of distortion product otoacoustic emission contralateral suppression," J. Acoust. Soc. Am. **123**, 4310–4320.
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.
Wagner, W., Heppelmann, G., Müller, J., Janssen, T., and Zenner, H. P. (**2007**). "Olivocochlear reflex effect on human distortion product otoacoustic emissions is largest at frequencies with distinct fine structure dips," Hear. Res. **223**, 83–92.
Williams, D. M., and Brown, A. M. (**1995**). "Contralateral and ipsilateral suppression of the $2f_1$-f_2 distortion product in human subjects ," J. Acoust. Soc. Am. **97**, 1130–1140.
Withnell, R. H., Shaffer, L. A., and Talmadge, C. L. (**2003**). "Generation of DPOAEs in the guinea pig," Hear. Res. **178**, 106–117.
Zhang, F., Boettcher, F. A., and Sun, X. M. (**2007**). "Contralateral suppression of distortion product otoacoustic emissions: Effect of the primary frequency in Dpgrams," Int. J. Audiol. **46**, 187–195.

Use of stimulus-frequency otoacoustic emissions to investigate efferent and cochlear contributions to temporal overshoot[a)]

Douglas H. Keefe,[b)] Kim S. Schairer,[c)] John C. Ellison,
Denis F. Fitzpatrick, and Walt Jesteadt
Boys Town National Research Hospital, 555 North 30th Street, Omaha, Nebraska 68131

(Received 29 July 2008; revised 9 December 2008; accepted 9 December 2008)

Behavioral threshold for a tone burst presented in a long-duration noise masker decreases as the onset of the tone burst is delayed relative to masker onset. The threshold difference between detection of early- and late-onset tone bursts is called overshoot. Although the underlying mechanisms are unclear, one hypothesis is that overshoot occurs due to efferent suppression of cochlear nonlinearity [von Klitzing, R., and Kohlrausch, A. (1994). J. Acoust. Soc. Am. **95**, 2192–2201]. This hypothesis was tested by using overshoot conditions to elicit stimulus-frequency otoacoustic emissions (SFOAEs), which provide a physiological measure of cochlear nonlinearity. SFOAE and behavioral thresholds were estimated using a modified maximum-likelihood yes-no procedure. The masker was a 400-ms "frozen" notched noise. The signal was a 20-ms, 4-kHz tone burst presented at 1 or 200 ms after the noise onset. Behavioral overshoot results replicated previous studies, but no overshoot was observed in SFOAE thresholds. This suggests that either efferent suppression of cochlear nonlinearity is not involved in overshoot, or a SFOAE threshold estimation procedure based on stimuli similar to those used to study behavioral overshoot is not sensitive enough to measure the effect. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3068443]

PACS number(s): 43.64.Jb, 43.64.Kc, 43.66.Ba, 43.66.Mk [BLM] Pages: 1595–1604

I. INTRODUCTION

The behavioral threshold of a brief tone burst or signal in the presence of a gated broadband or notched noise is higher when the burst is presented at the beginning of the noise than when the burst is delayed by approximately 200 ms within the duration of the gated noise (Zwicker, 1965; Elliott, 1965). The difference in thresholds between early and late bursts is called overshoot. In listeners with normal hearing, overshoot is around 10 dB in conditions with moderate narrow- or broadband noise levels, but this varies widely across individual listeners and stimulus conditions (e.g., Bacon, 1990; Bacon and Liu, 2000; Carlyon and White, 1992; Formby *et al.*, 2000; McFadden, 1989; Zwicker, 1965). The purpose of the current study is to examine a potential mechanism of the overshoot phenomenon. von Klitzing and Kohlrausch (1994) proposed that overshoot is the result of a shift in the basilar membrane (BM) input-output (I/O) function, and that the shift may be a result of efferent adaptation as suggested by Schmidt and Zwicker (1991). If this is the case, then the effects should be observable in stimulus-frequency otoacoustic emissions (SFOAEs), which are generated by the same mechanism as the cochlear nonlinearity. The hypothesis is that if overshoot is the result of efferent adaptation of cochlear nonlinearity, then comparable amounts of overshoot should be observed in behavioral and SFOAE thresholds.

In the past, overshoot has been explained as resulting from short-term adaptation in afferent auditory-nerve fibers tuned close to the signal frequency (Smith and Zwislocki, 1975). A difficulty with that model is that off-frequency tones (Bacon and Viemeister, 1985) and notched-noise maskers (Bacon and Smith, 1991; Strickland, 2004) can produce an overshoot effect. That is, overshoot occurs in the absence of excitation of fibers (by the masker) that are near the signal frequency. Overshoot effects are reduced in ears with impaired cochlear functioning. For example, overshoot is reduced in ears following intense sound exposure (Champlin and McFadden, 1989), following aspirin administration (McFadden and Champlin, 1990), and in ears with sensorineural hearing loss (Bacon and Takahashi, 1992; Strickland and Krishnan, 2005). These results are consistent with an explanation of overshoot based on efferent adaptations of cochlear mechanics: a reduction in cochlear nonlinearity would leave a smaller possible range of adaptation of cochlear function, thereby reducing or even eliminating overshoot.

Evidence for involvement of the efferent system is derived from a wide range of studies. Efferent adaptation is mediated by the medial olivary complex (MOC) in the auditory brainstem. A "fast" MOC adaptation occurs on time scales on the order of 100 ms in humans (Guinan, 2006). This is similar to the time course of behavioral overshoot results in which the magnitude of overshoot increases as the delay in signal onset relative to the masker onset increases toward 100 ms, and begins to asymptote around 200 ms (e.g., Elliot, 1965 and Zwicker, 1965). This suggests a possible relation of MOC functioning to overshoot. Such a relationship has been examined in studies performed on patients

[a)]Preliminary reports of this research were presented at meetings of the American Auditory Society (Keefe *et al.* 2006) and the Association for Research in Otolaryngology (Keefe *et al.* 2005, 2003).
[b)]Electronic mail: keefe@boystown.org
[c)]Present address: University of Wisconsin, 1975 Willow Drive, Madison, WI 53706.

who had undergone vestibular neurectomy, a procedure that severs vestibular and efferent nerves to alleviate severe vertigo. One study found differences within 1.1 dB in the overshoot measured in the surgery and non-surgery ears, thereby concluding that overshoot does not involve MOC efferent activity (Scharf *et al.*, 1997). Another study using different procedures found reduced overshoot in ears with vestibular neurectomy compared to non-surgery ears, thus supporting the role of MOC efferents as a contributor to overshoot (Zeng *et al.*, 2000). Interpretation of these data is complicated by the unknown effects of surgery on auditory function. Thus, there is no consensus of interpretation regarding available data from subjects who received surgery. Furthermore, due to their restriction to this class of patients, such studies cannot examine the role of the peripheral auditory system in overshoot in normal-hearing subjects.

Bacon and Liu (2000) tested listeners with normal hearing. Their results using both ipsilateral and contralateral precursors suggest either a possible role for peripheral adaptation in ipsilateral overshoot or a stronger overshoot effect in the ipsilateral ear due to the larger number of MOC efferent neurons responding best to ipsilateral stimulation, as was found by Liberman (1988) in cat ears. These results in normal-hearing listeners contradict theories of overshoot based solely on peripheral adaptation, but such adaptation may contribute to ipsilateral overshoot. Contralateral stimulation prior to noise onset reduced the overshoot effect in normal-hearing subjects but had no effect in hearing-impaired subjects (Turner and Doherty, 1997). These results support a role for the MOC system in overshoot, based on the assumption that MOC effects would be absent in hearing-impaired subjects.

Through experiments using cat in which the MOC fibers were cut, it is known that the MOC reflex enhances the compound action potential response to transient signals presented in a continuous background noise; this enhancement effect is termed antimasking (Kawase and Liberman, 1993). The MOC reflex also enhances the responses of single auditory-nerve fibers in cat for transients within a continuous background noise (Kawase *et al.*, 1993). Kawase and Liberman (1993) explained antimasking on the basis of MOC activation of outer hair cells (OHCs). MOC fibers terminate on the base of OHCs and modulate OHC function, and by extension, the cochlear nonlinearity. As described in reviews of efferent effects (Pujol, 1994; Guinan, 2006), the MOC modulates the OHC activity responsible for the amplification of low-level sounds. BM mechanics adapt in the presence of the noise over time due to MOC activation. This produces a shift to higher stimulus levels and a larger slope of the BM I/O function, i.e., the smaller, compressive slope is modified in the direction of a more nearly linear slope. The result is that the cochlear representation of noise is reduced and the internal signal-to-noise ratio (SNR) is increased for the late burst relative to the early burst. The detection threshold for the late burst is thereby lower. These effects are consistent with effects observed in animal studies in which the olivocochlear bundle was stimulated (Murugasu and Russell, 1996). Behavioral overshoot data support the antimasking mechanism (Strickland, 2001; Strickland, 2008), although it is not clear whether the improved representation of the tone burst in noise is due to adaptation of cochlear mechanics, neural processing, or both systems.

A shift in the BM I/O function due to efferent adaptation is the overshoot mechanism proposed by von Klitzing and Kohlrausch (1994), and is illustrated in Fig. 1. Each panel shows schematic BM I/O functions for the 200-ms delay condition (late burst), in which an efferent shift is present, versus the 1-ms delay condition (early burst), in which no efferent shift is present. Each I/O function is assumed to have linear growth at low input levels and compressive growth at higher input levels. The pair of triangles on each panel shows the input noise sound pressure level (SPL) in each condition. In this example, a signal is assumed present at a threshold based on an output level that is 8 dB higher than the noise level. The resulting signal levels at threshold are shown by circle symbols on each of the early-burst and late-burst I/O functions. The amount of overshoot reported in each panel is the input SPL difference of the threshold of the signal with 1-ms delay compared to the threshold of the signal with 200-ms delay.

SFOAEs elicited using sine tones in noise are a noninvasive correlate to behavioral tone-in-noise tests. If the above mechanism for overshoot is correct, the MOC effects on OHC functioning should be reflected in SFOAE thresholds in comparable stimulus conditions. Shifts in SFOAE level in the presence of contralateral (Souter, 1995) and ipsilateral noise (Guinan *et al.*, 2003; Goodman and Keefe, 2006; Guinan, 2006) have been used previously to study adaptation by the MOC efferent system.

In the present study, behavioral and SFOAE thresholds were measured in each subject using similar stimulus conditions delivered by the same insert earphones. OAE responses are often measured as I/O functions, which allow the estimation of an OAE threshold as that input level resulting in an OAE detected according to a chosen criterion level. A commonly used approach to defining the threshold is a minimum SNR. This definition has the disadvantage that noise fluctuations occurring near threshold can produce a nonmonotonic growth of response at stimulus levels close to threshold. One conclusion of a preliminary study (Keefe *et al.*, 2003) was that a better definition of SFOAE threshold was needed in order to accurately measure changes in threshold that might be associated with any overshoot effect. A novel aspect of the present study was the measurement of a SFOAE threshold using a sequential adaptive technique with properties that are well understood from psychophysical research. Both behavioral and SFOAE thresholds were measured at 4 kHz, which is a signal frequency often used in past overshoot studies, e.g., Turner and Doherty (1997), Bacon and Liu (2000), Smith *et al.* (2000), and Strickland (2001, 2004, 2008).

II. METHODS

A. Subjects

Subjects ($N=14$) were recruited from a young adult, normal-hearing population. The inclusion criteria for the experiments were that subjects have normal tympanograms and

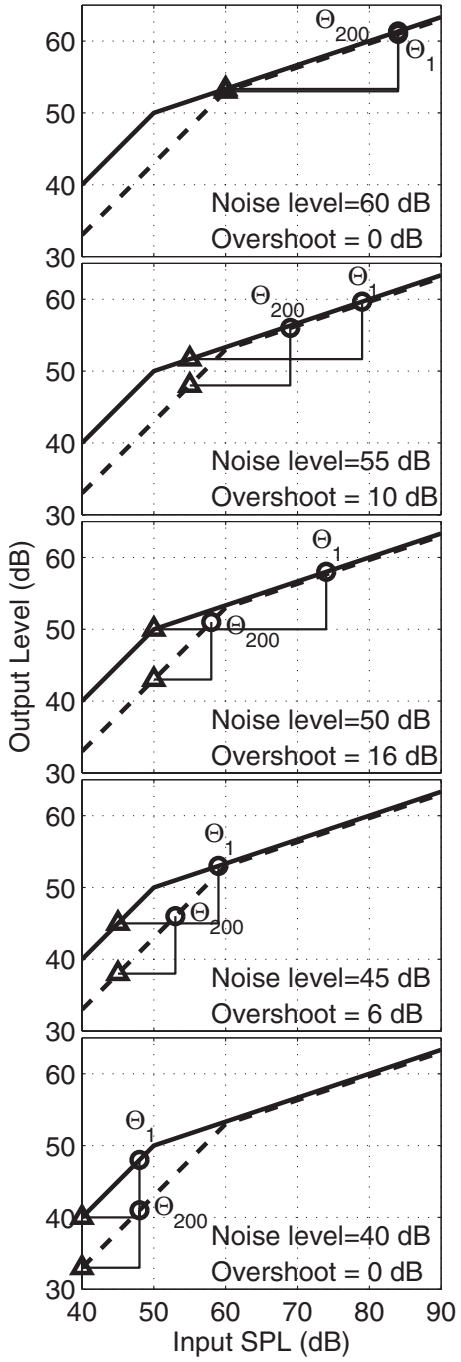


FIG. 1. This figure shows the threshold shifts that are predicted from the von Klitzing and Kohlrausch (1994) model. The modeled BM I/O functions (with arbitrary normalization) are plotted as a function of stimulus level at 1 ms after the noise onset (solid lines) and 200 ms after the noise onset (dashed line). The simulated noise sound pressure spectrum level is listed in each panel (as noise level) and represented on each plot by a triangle at this input level. The noise level decreases in 5-dB steps from 60 dB (top panel) to 40 dB (bottom panel). Each threshold is marked on each plot using a circle symbol and labeled with the appropriate threshold of the early (Θ_1) or late (Θ_{200}) burst. In this model, the SNR for threshold for the burst is assumed to be 8 dB above the noise level on the respective BM I/O function (i.e., 8 dB difference between triangles and circles on the *Y* axis, or output). For example, in the middle panel, the noise level is 50 dB, so that the output level at threshold is 8 dB higher, or 58 dB, for the late burst (Θ_{200}). The early-burst threshold Θ_1 attains this output level at an input SPL of 74 dB. The overshoot is the difference in input SPL of the early burst to that of the late burst, or 16 dB in this example.

normal air-conduction thresholds, i.e., less than or equal to 15-dB hearing level (HL) at all octave frequencies from 0.5 to 8 kHz. A 0.226-kHz tympanogram was classified within normal limits according to criteria described in Ellison and Keefe (2005). All testing was performed in one ear of each subject. Thresholds of a 4-kHz tone burst were measured in quiet and in noise presented over a range of levels. All behavioral thresholds for each subject were measured on the same day, but a different day from that on which SFOAE thresholds were measured. The more extensive SFOAE data collection was completed in seven 2-h test sessions, one test session for measuring the threshold of a tone-in-quiet condition and six sessions for the thresholds of tone-in-noise conditions. All 14 subjects completed the SFOAE testing whereas only 12 of these subjects participated in the behavioral testing. The other two subjects did not return for behavioral testing.

Research suggests the possibility of an attentional control of MOC functioning (Guinan, 2006), which might vary with subject state. The theory under evaluation was that MOC adaption of cochlear nonlinearity produced changes in SFOAEs, so that subject state (e.g., whether awake or asleep) might influence any adaptation of SFOAEs. To control for possible effects of subject state, subjects were kept awake during the SFOAE test sessions with the assistance of a DVD movie with sub-titles and no audible sound. This likely resulted in lower average levels of physiological noise in the SFOAE recordings, which might otherwise be elevated in a sleeping subject who begins to snore or in a bored subject who begins to fidget.

B. General methods

Subject responses were measured using custom software on a laboratory Windows computer with a high-quality sound card (24-bit resolution, 22.05-kHz sample rate, CardDeluxe). SFOAEs were measured for tone-in-quiet and tone-in-noise conditions using a commercial OAE ear-probe assembly comprised of a microphone and a pair of receivers (Etymotic ER-10C). The ear-probe was inserted into the ear canal using a soft foam tip provided by the manufacturer. The ear-probe system was modified by the manufacturer to provide 20 dB of additional gain in the output receivers. Behavioral thresholds were measured in response to stimuli presented through one receiver of the same ear-probe.

Behavioral and SFOAE measurements were performed using the same procedures and stimuli to the maximum extent possible. The signal was a 4-kHz, 20-ms (5-ms ramps) tone burst. The rationale for choosing this duration is described at the end of Sec. II. A 400-ms, fixed-length sample of noise was generated and used in all measurements as a frozen noise masker. The noise had a notch in the spectrum centered at 4 kHz, which coincided with the peak frequency in the tone-burst spectrum, and a notch width of 0.4 kHz, which corresponds to 0.14 octave. The lower and upper passbands each had a bandwidth of 1.6 kHz. The notch width was sufficiently broad that the main spectral lobe of the signal tone burst was contained completely within the notch. The early-burst condition presented the onset of the signal 1

ms after the onset of the noise, and the late-burst condition presented the onset of the signal 200 ms after the onset of the noise. The temporal fine structure of the frozen noise was identical for early and late bursts. Thresholds were measured at each of six noise masker levels, which varied over a 60-dB range in 12-dB increments. The noise sound pressure spectrum levels (reference bandwidth of 1 Hz) averaged across the upper and lower passbands were measured in an artificial ear simulator (IEC 711) to be −14, −2, 10, 22, 34, and 46 dB.

C. Adaptive behavioral threshold procedure using yes/no task

An adaptive, maximum-likelihood (ML) procedure was used to estimate thresholds in a yes-no task in the behavioral experiments. Green (1993) introduced a ML method for estimating thresholds in a yes-no task using a modified logistic function to represent the underlying psychometric function. Gu and Green (1994) improved the ability of the ML method to estimate false-alarm rates. Such ML procedures are accurate in behavioral threshold testing (Formby *et al.*, 1996; Leek *et al.*, 2000). We used a ML procedure similar to that of Gu and Green (1994) to measure thresholds with the following modification: to overcome any sensitivity to the results of the initial trials, the stimulus levels of the first four trials were set independent of the subject's responses to the trials. The stimulus dynamic range was split into four equal subranges. The stimulus level in each of the first four trials was selected by randomly choosing (without replacement) one of these subranges for the trial, and then randomly choosing a stimulus level within the sub-range. A threshold was then calculated using the ML estimate based on these initial four yes-no responses. The basic idea was that the stimulus levels of the fifth and subsequent trials would be chosen as in Green (1993) based on the current threshold estimate. This initialization procedure substantially reduced the sensitivity of the ML procedure to errors when testing subjects. Preliminary behavioral data were acquired with N as large as 30 trials and demonstrated that fewer trials gave results of sufficient accuracy. Therefore, $N=16$ trials were used in the main experiment. Of these 16 trials, 4 catch trials, i.e., 4 signal-absent trials, were randomly included after the first stimulus presentation. These catch trials identified whether the listener was inattentive, as would be evidenced by responding that a signal was present when no stimulus was presented, and were used to assess the false-alarm rate (Gu and Green, 1994). The presence of these catch trials on any of trials 2–16 interacted with the initialization procedure, which was based on the first four signal-present trials irrespective of the number of catch trials that might have been interpolated during the initialization. Thus, the stimulus levels of the fifth and subsequent signal-present trials were chosen using the ML procedure.

During the task, the subjects were instructed to ignore the masker and listen for the tone bursts. The observation interval was marked by a row of asterisks on the text panel of a custom-built response box. For each observation interval, they were instructed to push one button for "yes" and another button for "no" to indicate whether the tone burst was heard. Feedback included a message informing the subject of the beginning of the first trial of a run, the yes or no response entered on each trial, an alert to the next upcoming trial, and a message announcing the end of the threshold test. After each stimulus presentation, the automated procedure waited until the subject depressed the yes or no button. To measure a single behavioral threshold, the stimulus buffer was 1 s, with an approximate 1.3-s response time (that varied across subjects), with 16 trials per run and two runs per estimate for a total measurement time on the order of 1.2 min per condition. There was an additional brief waiting period for the subject between runs and an initial alert displayed at the beginning of each run. The final threshold was calculated as the mean of two successive runs as long as the standard deviation (SD) between runs was within 3 dB. If the SD was larger, then another run was performed.

Behavioral thresholds at 4 kHz were measured in 12 subjects, which included six left and six right ears, six males and six females, and a mean age of 21.2 years (SD of 1.5 years) in the −14, −2, and 22 dB noise level conditions. Behavioral thresholds were measured in 11 of these 12 subjects in the 10, 34, and 46 dB noise spectrum level conditions (the 12th subject was not tested in these conditions due to the subject's time constraints).

D. Adaptive SFOAE threshold procedure using yes/no task

1. Basic SFOAE measurement procedures

For the tone-in-quiet condition, the signal s_1 was the same 20-ms tone burst at 4 kHz as used in the behavioral task, and the suppressor stimulus s_2 was a 30-ms tone burst at a frequency 4% higher than that of the signal. The centers of each tone burst were time aligned. The suppressor level was fixed at 15 dB above the signal level. Both signal and suppressor levels were varied by equal increments within an adaptive threshold procedure. The pressure waveform p_1 was measured in response to stimulus s_1 delivered through receiver 1 of the probe, response p_2 was measured in response to stimulus s_2 delivered through receiver 2 of the probe, and response p_{12} was measured in response to the simultaneous presentation of s_1 through receiver 1 and s_2 through receiver 2. The SFOAE distortion response p_d was calculated as the nonlinear residual of $p_1+p_2-p_{12}$ (Keefe, 1998), as has been used in previous SFOAE studies (Schairer *et al.*, 2003; Schairer and Keefe, 2005; Schairer *et al.*, 2006). The distortion response is zero for a purely linear system and otherwise contains a residual that originates from the compressively nonlinear mechanics associated with OHC functioning. Recordings using the ear-probe inserted into an artificial ear (IEC 711, B&K model 4157) confirmed minimal distortion in the measurement system.

For the tone-in-noise conditions, the same tone-burst stimuli as described above were used, but the notched noise was added to the signal s_1 with no change in s_2. The early-burst condition presented the onset of the signal 1 ms after the onset of the noise (see left panels, Fig. 2), and the late-burst condition presented the onset of the signal 200 ms after the onset of the noise (see right panels, Fig. 2). The duration

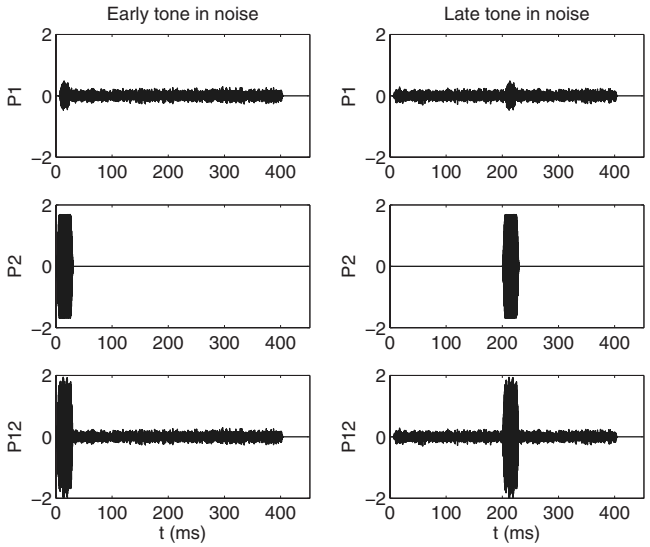


FIG. 2. The modeled stimulus waveform responses in the ear canal (in arbitrary units) are shown for the conditions used to estimate SFOAE thresholds. Modeled responses for SFOAEs elicited by tone bursts masked by notched noise in the early condition are shown in the left column, and responses in the late condition are shown in the right column. The top panels show the modeled signal+gated noise stimulus waveform responses p_1, the middle panels show the tonal suppressor responses p_2, and the bottom panels show the sum p_{12} of these responses.

of the gated noise was 400 ms, which was followed by a silent interval of 600 ms (total buffer duration of 1 s). The presentation of early and late tone bursts in a noise background (top row, Fig. 2) is analogous to the behavioral overshoot stimuli. The s_2 stimulus was identical to that used in quiet (middle row panels, Fig. 2). The third stimulus set consisted of the simultaneous presentation of s_1 and s_2 (bottom row panels, Fig. 2). The SFOAE distortion waveform p_d was calculated using the p_1, p_2, and p_{12} responses labeled in Fig. 2.

Using a technique similar to that described in Goodman and Keefe (2006), it was verified in preliminary experiments that there was no middle-ear muscle reflex shift evoked at even the highest-level noise condition. This technique compared the responses to a moderate-level low-frequency tone (0.25 kHz) that was presented in quiet and in the notched-noise masker. The lower passband in the notched-noise spectrum had negligible energy below 1.5 kHz, so that there was no spectral overlap between the noise and the low-frequency tone.

2. Measuring SFOAE thresholds adaptively using a yes/no task

The use of a yes-no task requires a decision that the signal, in this case the SFOAE, be classified as present (i.e., yes) or absent (i.e., no). To minimize total test time, the tone burst eliciting the SFOAE was presented on every trial. A 1024-sample Hanning window (with duration of 46 ms) was applied to a first time interval of the response that included the 20-ms tone burst followed by the interval in which the SFOAE would be present (including a few milliseconds of cochlear round-trip travel time). The mean SFOAE latency is in the range of 3–4.5 ms at 4 kHz (Shera and Guinan, 2003; Schairer *et al.*, 2006). As shown in simulated responses in Fig. 3, this first interval included all of the signal and suppressor tone energies for either the early or late tone burst. A second 1024-sample Hanning window was applied to the 1024 samples immediately following the first time interval. No signal or suppressor tone was present during this second interval, although the notched-noise signal was continuously present during both intervals. A discrete Fourier transform (DFT) of the windowed waveform was calculated for each of the intervals. A signal was classified as present if the energy level at 4 kHz in the first interval was at least 3 dB larger than the energy level at 4 kHz in the second interval. Classifier criteria of 0, 6, and 9 dB were also analyzed in preliminary experiments. The results appeared less consistent using the 0-dB criterion than with any other criterion. The 3-dB criterion was selected because results were reliable and thresholds were lower than with the 6- or 9-dB criteria.

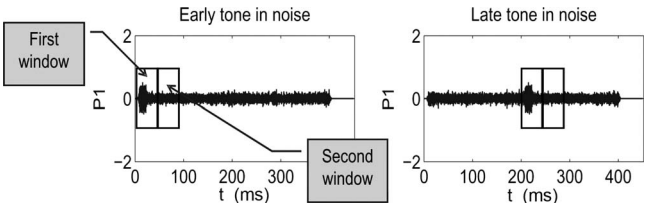


FIG. 3. The simulated signal response p_1 is plotted in arbitrary units as a function of time (in milliseconds) for early and late bursts in noise in the left and right panels, respectively. Each panel also shows the first and second windows, each 1024 samples (or 46 ms) in length, which were used for the signal-present and signal-absent detection conditions, respectively. Each window is drawn as a box.

The adaptive procedure to estimate SFOAE thresholds was generally similar to the behavioral procedure, except that no catch trials were included in the SFOAE procedure because it used an energy detection decision that was not subject to false alarms. On each trial of the adaptive procedure (i.e., a trial results in a yes or no decision), the stimulus set shown in Fig. 2 was presented multiple times to enable time averaging with real-time artifact rejection to exclude intermittent noise generated by the subject or other environmental source. In preliminary studies using two subjects, data were acquired using a number of repeated responses equal to 4, 8, 16, and 32. The response variability was analyzed in terms of the prevalence of far outliers in the sense defined by Tukey (1977). Such far outliers, when detected, were excluded from the time average. The artifact rejection method functioned better at excluding outliers with 16 or 32 responses than with 8 or fewer responses. The use of 32 responses did not reveal a clear benefit over the use of 16 responses. Thus, data in the main experiment were acquired using 16 responses per trial. After calculating the signal as the average of these 16 responses, the SFOAE threshold at each noise level or in quiet was the lowest signal level that produced a detectable SFOAE residual.

The final threshold was based on the median of three runs. The median was preferable to the mean because some runs were contaminated by excessively high SFOAE thresholds. A factor that possibly contributed to variability in thresholds is that some SFOAE I/O functions are nonmonotonic with increasing stimulus level (Schairer *et al.*, 2003;

Schairer and Keefe, 2005). Notches in the SFOAE output level may appear at a relatively high stimulus level, whereas an adaptive procedure to estimate threshold assumes a monotonic relationship. The role of this factor was not systematically studied because the entire I/O function was not recorded. Nevertheless, if present in any subject's run, this factor would play a role if the adaptive yes-no procedure of Green (1993) were used because of its sensitivity to the results of the first few trials. When the initial trial detected no SFOAE, the Green procedure set the stimulus to its maximum allowed level. This often resulted in preliminary experiments in SFOAE thresholds lying near the maximum of the range. As in the behavioral procedure, SFOAE thresholds were acquired in a sequential adaptive procedure in which the stimulus levels of the first four trials were set independently of the yes-no responses in the trials. Preliminary results showed that this initialization reduced the variability of SFOAE threshold estimates. This typically resulted in an initial threshold estimate within the low-to-moderate range of input levels on the SFOAE I/O function, so that there was reduced likelihood of an effect of a nonmonotonic notch in the I/O function.

The 20-ms duration of the 4-kHz tone burst in the SFOAE (and behavioral) measurements was selected based on a preliminary pilot study. Increasing the tone-burst duration would be expected to increase the SNR and thus decrease the SFOAE threshold. The tone-burst durations were varied between 10 ms, which is typical in behavioral overshoot studies, and 100 ms, which is on the order of MOC latency, and thus was an upper limit in the experimental design. If the stimulus levels were high enough, then it might produce system distortion that would contaminate the SFOAE recording in an ear. Therefore, it was potentially advantageous to use a longer-duration stimulus to evoke the SFOAE so as to measure its threshold over a lower range of stimulus levels. As expected, SFOAE thresholds were lower with increasing stimulus duration and increasing signal averaging. The signal averaging to extract the SFOAE was performed prior to any classification of the SFOAE as present or absent. Such signal averaging is not used in a behavioral experiment. The task in each of the behavioral and SFOAE experiments is to detect a signal in the presence of internal and external noise. The external noise in each experiment includes the noise masker and any environmental noise transmitted with the stimulus. The internal noise sources differ in the behavioral and SFOAE experiments: the behavioral internal noise source is attributed to the listener's internal state, while the SFOAE internal noise includes physiologic sources of noise and the noise at the output of the microphone preamplifier. The SFOAE internal noise is measured using the DFT at the frequency (4 kHz) of the signal tone burst. As described above (see Fig. 2), a signal duration of 20 ms with a suppressor duration of 30 ms was used in the main experiment. These durations were somewhat longer than tone bursts used in behavioral overshoot studies (e.g., Formby *et al.*, 2000; McFadden, 1989; Zwicker, 1965). However, the longer durations were deemed to be appropriate because they improved the SNR of the SFOAE measurements, yet were shorter than the onset latencies associated with fast MOC effects. Their use did not detract from the behavioral measurements.

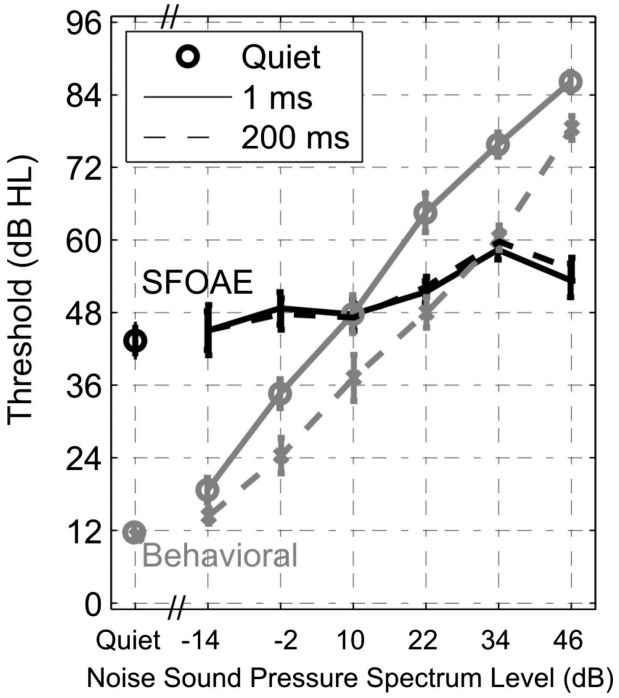


FIG. 4. Mean behavioral (gray lines) and SFOAE (black lines) thresholds in decibel HL as a function of the masker level, i.e., the noise sound pressure spectrum level (reference bandwidth of 1 Hz) for the tone burst presented in noise at 1 ms after the masker onset (solid lines) and 200 ms after the masker onset (dashed lines). The thresholds in quiet for each measure are plotted as detached symbols on the left side. Each error bar extends ±1 SE from its mean.

To measure a single SFOAE threshold at a fixed noise level or in quiet, three buffers (p_1, p_2, and p_{12}) comprising the OAE stimulus had a total duration of 3 s, with 16 averages per trial, 16 trials per run, and three runs per estimate for a total measurement time of 38.4 min. Note that only two runs per estimate were obtained in the behavioral paradigm because behavioral thresholds were less variable and the results more predictable than for the SFOAE paradigm. SFOAE thresholds at 4 kHz were measured in all 14 subjects, including six left and eight right ears, eight males and six females, and a mean age of 21.1 years (SD of 1.4 years).

III. RESULTS

A. Behavioral thresholds

The mean ±1 standard error of the mean (SE) of the behavioral thresholds for the tone burst in quiet and in notched noise at each noise spectrum level is plotted in Fig. 4 for the early-burst (labeled 1 ms, gray solid line) and late-burst (labeled 200 ms, gray dashed line) conditions. The two thresholds in quiet for the early and late bursts provided replications of the threshold-in-quiet measurement (12-dB HL). This threshold in quiet was based on SPL measured in a 2 cm^3 coupler in relation to its reference equivalent threshold sound pressure level (RETSPL) based on continuous tones. It lies above the nominal threshold of 0 dB HL due to the brief duration (20 ms) of the tone burst. The thresholds in

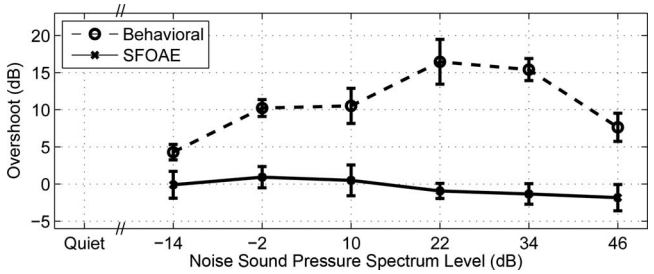


FIG. 5. The group mean ±1 SE of the behavioral overshoot (dashed line) and SFOAE overshoot (solid line) calculated from the group responses in Fig. 4 are plotted as a function of the masker level, i.e., the noise sound pressure spectrum level (reference bandwidth 1 Hz).

noise were elevated above the threshold in quiet and the late-burst thresholds increased linearly with increasing noise spectrum level (except for a slightly steeper rate of increase for the threshold at the highest noise spectrum level). The slope of the early-burst threshold function was slightly larger than that of the late-burst function at lower noise spectrum levels and slightly smaller at higher noise spectrum levels, resulting in an overall concave-downwards shape. The difference in the early-burst and late-burst thresholds was calculated for each subject as a function of noise spectrum level. In Fig. 5 (dashed line), the mean ±1 SE of this difference across subjects is plotted as the group-averaged overshoot. The mean overshoot had a maximum of 16 dB at an intermediate noise spectrum level of 22 dB. Mean overshoot decreased to 4 dB at the lowest noise sound pressure spectrum level (−14 dB) and to 7 dB at the highest noise level (46 dB).

B. SFOAE thresholds

The mean ±1 SE of the SFOAE thresholds evoked by the tone burst in quiet and in notched noise is plotted in Fig. 4 for the early-burst (labeled 1 ms, black solid line) and late-burst (labeled 200 ms, black dashed line) conditions. The mean threshold in quiet was 43-dB HL with a SE of approximately 3 dB. The SFOAE thresholds for tone-in-noise conditions were similar to the threshold for the tone in quiet at the lower noise spectrum levels (up to 10-dB SPL) and were more elevated at high noise spectrum levels (22 dB SPL and above). The thresholds in noise increased with noise spectrum level with a slope of 0.26 dB/dB for noise sound pressure spectrum levels in the range −14 to 34 dB. The threshold at the highest noise level (46 dB) was lower than that at the next highest level (34 dB). This SFOAE threshold slope of 0.26 dB/dB was much less than the behavioral-threshold slope of approximately 1 dB/dB.

There was little difference in the mean SFOAE thresholds of the early and late bursts in Fig. 4. This is quantified by the mean SFOAE overshoot in Fig. 5 (solid line), which represents the mean ±1 SE of the difference across subjects in the early-burst and late-burst thresholds. This mean was equal to 0 dB to within the measurement variability, i.e., the early and late SFOAE thresholds were equal.

As explained in Sec. II, there was concern in the adaptive SFOAE threshold procedure with the threshold variability between the three adaptive runs from which the threshold was estimated as the median. This variability was analyzed in terms of the SD across the three runs. This SD was calculated for the quiet and each of the noise conditions, and then averaged across subjects. This average SD was generally similar for the early- and late-bursts in noise conditions. The average SD for the tone-in-noise conditions did not appear to increase with increasing noise spectrum level. The grand average across all conditions and subjects of the SD across the three runs was 4.5 dB.

IV. DISCUSSION

The ability to accurately measure thresholds has been an important experimental test in psychoacoustics, with adaptive techniques to measure thresholds in widespread use. An adaptive procedure to measure the threshold of a SFOAE response was developed using an energy detection step embedded in a yes-no task. Each run in the adaptive procedure used maximum likelihood to set the stimulus levels on the 5th through the 16th trials and to calculate the threshold. The resulting SFOAE thresholds had sufficiently small standard errors to examine the predictions of a theory of behavioral overshoot.

The behavioral thresholds of the late tone burst increased linearly with increases in noise sound pressure spectrum level with the exception of the change in threshold from the 34- to 46-dB level conditions (Fig. 4). This linear relationship is consistent with the simultaneous masking of the tone burst by the notched noise, in which the notch width was relatively narrow (0.14 octave). The rate of change in behavioral thresholds of the early tone burst varied with noise spectrum level, that is, it was not linear as in the late tone-burst condition. This is expected because overshoot is larger in moderate masker level conditions than in lower and higher masker level conditions, and is attributed to nonlinear changes in the early tone-burst threshold relative to the linear changes in the late-burst threshold. The SFOAE thresholds increased more slowly with increasing noise spectrum level, with a slope of 0.26 dB/dB except at the highest noise spectrum level. The SFOAE thresholds would not be expected to increase linearly with increasing noise spectrum level, because the SFOAE was measured at 4 kHz, which is the center frequency within the notch of the notched-noise masker. The SFOAE threshold slope of 0.26 dB/dB appears consistent with a compressive growth of SFOAE suppression on the BM at higher notched-noise levels. This disparity in slopes in Fig. 4 resulted in more elevated SFOAE thresholds than behavioral thresholds in quiet and at low noise levels, and more elevated behavioral thresholds than SFOAE thresholds at high noise levels.

The absolute levels of the SFOAE thresholds were slightly elevated overall through the use of the 3-dB SNR criterion in the energy detection procedure used at the 4-kHz DFT bin (see Sec. II). SFOAE thresholds would have been reduced in quiet and for each noise condition if a 0-dB SNR criterion had been used or if additional averaging of responses had been performed. The fact that the SFOAE energy detector used the same SNR criterion in quiet and in

each noise condition means that differences in physiological noise would not be expected to account for the shallow slope of the SFOAE threshold function in Fig. 4.

The variability in mean thresholds across subjects was slightly larger for the SFOAE threshold data than the behavioral threshold data, as is evident in the size of the SEs in Figs. 4 and 5. This larger SFOAE variability must be interpreted in the context of the fact that these estimates were based on 16 averages per trial within the adaptive procedure. Thus, the adverse effects of noise were greater for SFOAE threshold detection than for behavioral threshold detection. SFOAE measurements included additional effects of noise and variability from reverse transmission, small biological signals, and any measurement equipment artifacts that are not a factor in behavioral thresholds.

The theory summarized in Sec. I explains the lower thresholds for the late burst in noise relative to the early burst in noise in terms of an adaptation mechanism associated with OHC and MOC efferent functioning, the effects of which are present for the late burst and absent for the early burst. Thus, noise excitation activates the MOC system, which affects OHC functioning to reduce the amount of compression associated with the cochlear-amplifier mechanism. As shown in Fig. 1, the predicted overshoot has a maximum at intermediate noise spectrum levels. This midlevel maximum in overshoot was reported in Bacon and Smith (1991) and in other studies cited in Sec. I, and, in particular, was observed in behavioral overshoot in the present study using notched noise (see Fig. 5).

OAEs are produced through the activity of OHCs evoked by a sound stimulus. A 4 kHz tone burst at moderate levels in quiet produces a transient SFOAE at 4-kHz in normal ears (Konrad-Martin and Keefe, 2003) and in some ears with mild hearing loss (Konrad-Martin and Keefe, 2005). It was hypothesized that if OHC functioning differs in the early- and late-bursts in noise conditions, then differences in evoked SFOAEs would be observed. Such differences were not observed. There was no mean SFOAE overshoot despite the presence of a behavioral overshoot as large as 16 dB (Fig. 5).

There are three possible explanations that might account for the absence of SFOAE overshoot. First, efferent adaptation of cochlear nonlinearity may not be the dominant underlying mechanism of overshoot. Second, although normal cochlear function is needed for sufficient afferent stimulation to elicit the MOC, the resulting MOC effects may be neural. That is, the effect might reflect MOC adaptation of neural elements rather than cochlear mechanisms. The detection of the late burst presented 200 ms after the noise onset is influenced by this adaptation, but the detection of the early burst is not. The adaptation may be regarded as lowering the internal gain of the noise, i.e., its internal level is reduced at 200 ms after the noise onset in comparison to immediately following the noise onset. This adaptation does not necessarily occur at the OHC level. When the late burst is presented at a given spectral level, its internal SNR in some neural representation(s) may be larger than that of the early burst. Because the burst is presented at a higher spectrum level than the noise spectrum level, and because the auditory system has a compressively nonlinear response with increasing stimulus level, the adaptation effect on the noise in the neural representation may be greater than that on the brief burst. This is generally similar to the antimasking effects observed in cat (Kawase and Liberman, 1993; Kawase *et al.*, 1993). However, antimasking effects in cat were explained as due to adaptation of OHC functioning, whereas the present results in humans could imply that the effects rely on factors other than OHC functioning. In Sec. I, reduction in overshoot in listeners with noise and aspirin exposure was interpreted as evidence for cochlear involvement. However, noise exposure and sensorineural hearing loss can also be associated with changes in neural functioning, and aspirin administration has systemic effects on neural functioning in addition to its effects on OHC function; salicylate effects are reviewed in Cazals (2000). Thus, the reduction in overshoot by any of these effects does not necessarily demonstrate that overshoot is produced by changes in OHC functioning alone, although it is consistent with this involvement by reducing the afferent auditory innervation received by MOC neurons.

Neural sources of excitatory and inhibitory feedback may contribute to the mechanisms underlying overshoot (McFadden, 1989). Neural circuits involving the MOC system are of particular interest. One possible circuit in the ventral cochlear nucleus (VCN) relies on cholinergic modulation of stellate cells (Fujino and Oertel, 2001). MOC efferent fibers projecting to OHCs also project through collateral branches to the cochlear nuclei, including to T stellate cells and D stellate cells in the VCN. T stellate cells are part of an excitatory auditory pathway from the VCN through the ventral nucleus of the trapezoid body (VNTB) to the contralateral inferior colliculus. The activity of T stellate cells is modulated by cholinergic inputs via feedback from the VNTB including MOC neurons. D stellate cells inhibit the activity of T stellate cells and are not modulated by cholinergic inputs. Stellate cells are tonotopically organized, with T stellate cells having a narrower excitatory tuning and D stellate cells having a broader inhibitory tuning. As further described by Fujino and Oertel (2001), this combination produces a lateral inhibition effect that may enhance the encoding of spectral peaks in noise. More research is needed to evaluate this and other possible neural sources of overshoot. Experiments using non-human mammals are also relevant, e.g., MOC lesions in macaques produced changes in behavioral thresholds for tones in continuous contralateral noise (Smith *et al.*, 2000).

A third explanation to account for the absence of SFOAE overshoot is that efferent adaptation of cochlear nonlinearity is an important underlying mechanism of behavioral overshoot, but the SFOAE threshold estimation procedure was not sensitive enough to detect SFOAE overshoot. It may be that a large behavioral overshoot effect is accompanied by a SFOAE overshoot effect that would require much more averaging to reveal. The variability in thresholds across subjects and the limits to threshold detection in the lower level masker conditions (Fig. 4) may obscure much smaller SFOAE effects. If this third explanation has merit, one implication is that the stimulus conditions that will be effective for studying the cochlear mechanisms underlying behavioral

overshoot may not resemble the stimulus conditions used in the present experiment. The latter were designed to be similar to the corresponding behavioral stimuli. The present results may not completely rule out the possibility that MOC efferent adaptation of cochlear nonlinearity is an important contributor to behavioral overshoot, but they constrain the possible set of models to those that can also account for the absence of a detectable MOC adaptation in SFOAE responses of the type reported in the present study. A difficulty with this explanation is that the SFOAE thresholds were approximately equal or lower than the behavioral thresholds for the tone-in-noise conditions at noise levels of 34 and 46 dB (see Fig. 4). Although behavioral overshoot was observed at these higher noise levels, no SFOAE overshoot was observed (see Fig. 5). This finding supports the view that MOC adaptation of OHC function cannot explain overshoot at these higher noise masker levels.

V. CONCLUSIONS

A sequential adaptive method to measure an OAE threshold based on a ML procedure was developed and applied to the measurement of the threshold of a SFOAE elicited by a tone burst in quiet and in noise. Such an adaptive method is a promising tool for efficient measurements of OAE thresholds. SFOAE thresholds were lower for the burst in quiet than for the burst in moderate-to-high noise levels. SFOAE thresholds did not differ for detection of early and late tone bursts to within the measurement variability. Thus, there was no overshoot in SFOAE threshold data despite confirming the presence of overshoot in behavioral thresholds in the same subject group. The results provide no support for a theory of behavioral overshoot based on temporal adaptation of cochlear OHC function. This negative result does not exclude the possibility that behavioral overshoot has its origins in efferent adaptation of nonlinear cochlear mechanics at lower and moderate noise levels, but it remains to be explained how such a mechanism would not lead to detectable overshoot in SFOAE responses at the higher noise levels. It is concluded that there are important neural contributors of overshoot.

ACKNOWLEDGMENTS

The first author benefited from discussions with Dr. Donata Oertel, who suggested the possible relevance of the functioning of stellate cells in the VCN to theories of behavioral overshoot. Two anonymous reviewers provided helpful critiques of a previous version of the manuscript. This research was supported by the NIH Grant Nos. DC003784 and DC006342, with core support provided by NIH Grant No. DC004662.

Bacon, S. P. (**1990**). "Effect of masker level on overshoot," J. Acoust. Soc. Am. **88**, 698–702.

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

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

Bacon, S. P., and Takahashi, G. A. (**1992**). "Overshoot in normal-hearing and hearing-impaired subjects," J. Acoust. Soc. Am. **91**, 2865–2871.

Bacon, S. P., and Viemeister, N. F. (**1985**). "The temporal course of simultaneous tone-on-tone masking," J. Acoust. Soc. Am. **78**, 1231–1235.

Carlyon, R. P., and White, L. J. (**1992**). "Effect of signal frequency and masker level on the frequency regions responsible for the overshoot effect," J. Acoust. Soc. Am. **91**, 1034–1041.

Cazals, Y. (**2000**). "Auditory sensori-neural alterations induced by salicylate," Prog. Neurobiol. **62**, 583–631.

Champlin, C. A., and McFadden, D. (**1989**). "Reductions in overshoot following intense sound exposures," J. Acoust. Soc. Am. **85**, 2005–2011.

Elliott, L. L. (**1965**). "Changes in the simultaneous masked threshold of brief tones," J. Acoust. Soc. Am. **38**, 738–746.

Ellison, J. C., and Keefe, D. H. (**2005**). "Audiometric predictions using SFOAE and middle-ear measurements," Ear Hear. **26**, 487–503.

Formby, C., Sherlock, L. P., and Green, D. M. (**1996**). "Evaluation of a maximum likelihood procedure for measuring pure-tone thresholds under computer control," J. Am. Acad. Audiol. **7**, 125–129.

Formby, C., Sherlock, L. P., and Ferguson, S. H. (**2000**). "Enhancement of the edges of temporal masking functions by complex patterns of overshoot and undershoot," J. Acoust. Soc. Am. **107**, 2169–2187.

Fujino, K., and Oertel, D. (**2001**). "Cholinergic modulation of stellate cells in the mammalian ventral cochlear nucleus," J. Neurosci. **21**, 7372–7383.

Goodman, S. S., and Keefe, D. H. (**2006**). "Simultaneous measurement of noise-activated middle-ear muscle reflex and stimulus frequency otoacoustic emissions," J. Assoc. Res. Otolaryngol. **7**, 125–139.

Green, D. (**1993**). "A maximum-likelihood method for estimating thresholds in a yes-no task," J. Acoust. Soc. Am. **93**, 2096–2105.

Gu, X., and Green, D. M. (**1994**). "Further studies of a maximum-likelihood yes-no procedure," J. Acoust. Soc. Am. **96**, 93–101.

Guinan, J. J., Jr. (**2006**). "Olivocochlear efferents: Anatomy, physiology, function, and the measurement of efferent effects in humans," Ear Hear. **27**, 589–607.

Guinan, J. J., Jr., Backus, B. C., Lilaonitkul, W., and Aharonson, V. (**2003**). "Medial olivocochlear efferent reflex in humans: Otoacoustic emission (OAE) measurement issues and the advantages of stimulus frequency OAEs," J. Assoc. Res. Otolaryngol. **4**, 521–40.

Kawase, T., and Liberman, M. C. (**1993**). "Antimasking effects of the olivocochlear reflex. I. Enhancement of compound action potentials to masked tones," J. Neurophysiol. **70**, 2519–2532.

Kawase, T., Delgutte, B., and Liberman, M. C. (**1993**). "Antimasking effects of the olivocochlear reflex. II. Enhancement of auditory-nerve response to masked tones," J. Neurophysiol. **70**, 2533–2549.

Keefe, D. H. (**1998**). "Double-evoked otoacoustic emissions. I. Measurement theory and nonlinear coherence," J. Acoust. Soc. Am. **103**, 3489–3498.

Keefe, D. H., Schairer, K. S., Ellison, J. C., Fitzpatrick, D. F., and Jesteadt, W. (**2005**). "Stimulus frequency otoacoustic emissions elicited by transient tones in noise," Assoc. Res. Otolaryngol. Abstr. **28**, 1269.

Keefe, D. H., Schairer, K. S., Ellison, J. C., Fitzpatrick, D. F., and Jesteadt, W. (**2006**). "Thresholds of behavioral and OAE responses to tones in noise," American Auditory Society Meeting, Scottsdale, AZ.

Keefe, D. H., Schairer, K. S., and Jesteadt, W. (**2003**). "Is there an OAE correlate to behavioral overshoot?" Assoc. Res. Otolaryngol. Abstr. **26**, 397.

Konrad-Martin, D., and Keefe, D. H. (**2003**). "Time-frequency analyses of transient-evoked stimulus-frequency and distortion-product otoacoustic emissions: Testing cochlear model predictions," J. Acoust. Soc. Am. **114**, 2021–2043.

Konrad-Martin, D., and Keefe, D. H. (**2005**). "Transient-evoked stimulus-frequency and distortion-product otoacoustic emissions in normal and impaired ears," J. Acoust. Soc. Am. **117**, 3799–3815.

Leek, M. R., Dubno, J. R., He, N., and Ahlstrom, J. B. (**2000**). "Experience with a yes-no single-interval maximum-likelihood procedure," J. Acoust. Soc. Am. **107**, 2674–2684.

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

McFadden, D. (**1989**). "Spectral differences in the ability of temporal gaps to reset the mechanisms underlying overshoot," J. Acoust. Soc. Am. **85**, 254–261.

McFadden, D., and Champlin, C. (**1990**). "Reductions in overshoot during aspirin use," J. Acoust. Soc. Am. **87**, 2634–2642.

Murugasu, E., and Russell, I. J. (**1996**). "The effect of efferent stimulation on basilar membrane displacement in the basal turn of the guinea pig cochlea," J. Neurosci. **16**, 325–332.

Pujol, R. (**1994**). "Lateral and medial efferents: A double neurochemical

mechanism to protect and regulate inner and outer hair cell function in the cochlea," Br. J. Audiol. **28**, 185–191.
Schairer, K. S., Ellison, J. C., Fitzpatrick, D., and Keefe, D. H. (**2006**). "Use of stimulus-frequency otoacoustic emission latency and level to investigate cochlear mechanics in human ears," J. Acoust. Soc. Am. **120**, 901–14.
Schairer, K. S., Fitzpatrick, D., and Keefe, D. H. (**2003**). "Input-output functions for stimulus-frequency otoacoustic emissions in normal-hearing adult ears," J. Acoust. Soc. Am. **114**, 944–966.
Schairer, K. S., and Keefe, D. H. (**2005**). "Simultaneous recording of stimulus-frequency and distortion-product otoacoustic emission input-output functions in adult ears," J. Acoust. Soc. Am. **117**, 818–832.
Scharf, B., Magnan, J., and Chays, A. (**1997**). "On the role of the olivocochlear bundle in hearing: 16 case studies," Hear. Res. **103**, 101–122.
Schmidt, S., and Zwicker, E. (**1991**). "The effect of masker spectral asymmetry on overshoot in simultaneous masking," J. Acoust. Soc. Am. **89**, 1324–1330.
Shera, C. A., and Guinan, J. J., Jr. (**2003**). "Stimulus-frequency-emission group delay: A test of coherent reflection filtering and a window on cochlear tuning," J. Acoust. Soc. Am. **113**, 2762–72.
Smith, D. W., Turner, D. A., and Henson, M. M. (**2000**). "Psychophysical correlates of contralateral efferent suppression. I. The role of the medial olivocochlear system in 'central masking' in nonhuman primates," J. Acoust. Soc. Am. **107**, 933–941.
Smith, R. L., and Zwislocki, J. J. (**1975**). "Short-term adaptation and incremental responses of single auditory-nerve fibers," Biol. Cybern. **17**, 169–182.
Souter, M. (**1995**). "Suppression of stimulus frequency otoacoustic emissions by contralateral noise," Hear. Res. **91**, 167–177.
Strickland, E. A. (**2001**). "The relationship between frequency selectivity and overshoot," J. Acoust. Soc. Am. **109**, 2062–2073.
Strickland, E. A. (**2004**). "The temporal effect with notched-noise maskers: Analysis in terms of input-output functions," J. Acoust. Soc. Am. **115**, 2234–2245.
Strickland, E. A. (**2008**). "The relationship between precursor level and the temporal effect," J. Acoust. Soc. Am. **123**, 946–954.
Strickland, E. A., and Krishnan, L. A. (**2005**). "The temporal effect in listeners with mild to moderate cochlear hearing impairment," J. Acoust. Soc. Am. **118**, 3211–3217.
Tukey, J. W. (**1977**). *Exploratory Data Analysis* (Addison-Wesley, Reading, MA).
Turner, C. W., and Doherty, K. A. (**1997**). "Temporal masking and the 'active process' in normal and hearing-impaired listeners," in *Modeling Sensorineural Hearing Loss*, edited by W. Jesteadt (Erlbaum, Hillsdale, NJ), pp. 387–396.
von Klitzing, R., and Kohlrausch, A. (**1994**). "Effect of masker level on overshoot in running- and frozen-noise maskers," J. Acoust. Soc. Am. **95**, 2192–2201.
Zeng, F. G., Martino, K. M., Linthicum, F. H., and Soli, S. D. (**2000**). "Auditory perception in vestibular neurectomy subjects," Hear. Res. **142**, 102–112.
Zwicker, E. (**1965**). "Temporal effects in simultaneous masking by white-noise bursts," J. Acoust. Soc. Am. **37**, 653–663.

An *in situ* calibration for hearing thresholds

Robert H. Withnell[a)]
Department of Speech and Hearing Sciences, Indiana University, Bloomington, Indiana 47405

Patricia S. Jeng
Mimosa Acoustics, Champaign, Illinois 61820

Kelly Waldvogel and Kari Morgenstein
Department of Speech and Hearing Sciences, Indiana University, Bloomington, Indiana 47405

Jont B. Allen
Department of Electrical and Computer Engineering, University of Illinois, Urbana, Illinois 61801

(Received 17 September 2008; revised 12 December 2008; accepted 14 December 2008)

Quantifying how the sound delivered to the ear canal relates to hearing threshold has historically relied on acoustic calibration in physical assemblies with an input impedance intended to match the human ear (e.g., a Zwislocki coupler). The variation in the input impedance of the human ear makes such a method of calibration questionable. It is preferable to calibrate the acoustic signal in each ear individually. By using a calibrated sound source and microphone, the acoustic input impedance of the ear can be determined, and the sound delivered to the ear calibrated in terms of either (i) the incident sound pressure wave or (ii) that portion of the incident sound pressure wave transmitted to the middle ear and cochlea. Hearing thresholds expressed in terms of these quantities are reported, these *in situ* calibrations not being confounded by ear canal standing waves. Either would serve as a suitable replacement for the current practice of hearing thresholds expressed in terms of sound pressure level calibrated in a 6cc or 2cc coupler.
 [DOI: 10.1121/1.3075551]

PACS number(s): 43.64.Yp [BLM] Pages: 1605–1611

I. INTRODUCTION

Pure tone audiometry has been the standard for measuring hearing sensitivity since soon after the development of the first commercial audiometer (Fowler and Wegel, 1922). Quantification of the signal delivered to the ear (i.e., calibration of the audiometer) entails measuring the sound signal generated by the headphone with a microphone physically coupled to the headphone through a volume intended to match the volume enclosed by the headphone on a human ear (Burkhard and Corliss, 1954). But the acoustic output of a headphone, "as a function of electrical signal, depends on the acoustic load with which it is terminated" (Burkhard and Corliss, 1954, p. 679). The acoustic input impedance of the human ear has been found to vary significantly (Voss and Allen, 1994), raising serious doubt about the validity of the standard method of calibrating an audiometer. This obviously has implications for the accuracy of hearing tests (e.g., Voss *et al.*, 2000). *In situ* measurement of acoustic signals in the ear, with a microphone placed in the ear canal, provides an alternative method of calibration but standing waves confound sound pressure measurements (Stinson *et al.*, 1982; Siegel, 1994). The use of a microphone in clinical settings for recording sound pressure has been predominantly used for recording otoacoustic emissions and monitoring stimulus signal levels, and for obtaining the input impedance of the ear (e.g., tympanometry and wideband power reflectance). Due to the effect of standing waves on the monitoring of signal levels for generating otoacoustic emissions at high frequencies (Siegel, 1994), Neely and Gorga (1998) calibrated their sound stimuli in terms of sound intensity rather than sound pressure. However, as observed by Farmer-Fedor and Rabbitt (2002), the total acoustic intensity also forms standing waves in the ear canal. Farmer-Fedor and Rabbitt (2002) suggested quantifying the incident or forward-going acoustic intensity in the ear canal for the calibration of acoustic signals in the ear canal, it being a "valid measure of the stimulus input to the ear over the entire frequency range" (p. 617), devoid of standing waves, and so "could be used to extend the range of audiometric tests to high frequencies (>8 kHz)" (p. 617). Consistent with this suggestion Hazlewood *et al.* (2007), in consideration of how the ear processes the power it receives, concluded that hearing thresholds must be expressed in terms of the forward-going sound pressure (or intensity) wave, i.e.,

$$P_i^{\mathrm{th}} = \frac{P_m^{\mathrm{th}}}{1 + R_m}, \tag{1}$$

where P_i^{th} is the forward-going or incident sound pressure wave at behavioral hearing threshold, P_m^{th} is the sound pressure at the measurement microphone at behavioral hearing threshold, and R_m is the complex reflectance of the ear at the measurement microphone in the ear canal. Pursuant to the clinical application of calibrating stimulus levels in terms of the forward-going sound intensity or pressure wave, Scheperle *et al.* (2008) investigated stimulus level variability in

a)Electronic mail: rwithnel@indiana.edu

distortion product otoacoustic emission measurement with the stimuli calibrated in terms of the forward-going sound wave, and found that stimulus level calibrated in terms of the forward-going sound pressure wave was less variable than total sound pressure as a function of probe insertion depth in the ear canal, consistent with the forward-going or incident sound pressure wave not being contaminated by standing waves.

Sound propagation in the human ear canal at low frequencies occurs predominantly as plane waves. Up to about 6 kHz, the ear canal can be modeled as a uniform cylinder (Stinson, 1985) and so can be represented by a one dimensional transmission line or waveguide, terminated by the impedance of the middle ear. At higher frequencies, multiple higher-order modes may be present due to the shape and varying cross-sectional area of the ear canal. Farmer-Fedor and Rabbitt (2002) explored the contribution of these higher-order modes by examining nonplanar sound waves in the ear canal by making sound pressure measurements at multiple locations in an open ear canal; they concluded that quantifying the nonplanar traveling wave, while more accurate, was technically challenging. In contrast, studies that assume that sound propagates along the ear canal as a plane wave (e.g., Rabinowitz, 1981; Keefe *et al.*, 1993; Voss and Allen, 1994) have provided for the development of clinical instruments that quantify the reflectance of the ear by measuring sound pressure in a closed ear canal with a probe assembly that houses both a sound source (a speaker) and a microphone (e.g., Feeney *et al.*, 2003; Allen *et al.*, 2005).

The use of a single microphone/earphone combination in a closed ear canal to quantify the input impedance/admittance (and therefore reflectance) of the ear for plane waves and subsequent calibration in terms of the forward-going sound intensity or pressure wave rather than total sound pressure requires the prior determination of the acoustical characteristics of the sound source, i.e., the acoustic impedance and pressure of the microphone/earphone combination. The acoustic impedance (Z_s) and sound pressure (P_s) of the sound source can be determined by connecting the sound source to various cavities with known acoustic impedance and measuring the sound pressure in these cavities for a constant electrical source signal (Rabinowitz, 1981; Allen, 1986; Keefe *et al.*, 1992, 1993). With P_s and Z_s determined, the sound source can be inserted in an ear canal and the load admittance or the acoustic input admittance of the ear (Y_m) determined by measuring the sound pressure in the ear canal, i.e.,

$$Y_m = \frac{U_s}{P_m} - Y_s, \tag{2}$$

where Y_m is the input admittance of the ear at the microphone location, U_s is the volume velocity of the sound from the source ($U_s = P_s/Z_s$), Y_s is the admittance of the source, and P_m is the sound pressure at the measurement microphone. A significant, if not the primary, determinant of the setting of hearing sensitivity is the acoustic input impedance/admittance of the ear. The (normalized) acoustic input admittance of the ear (Y_m/Y_0) can also be expressed in terms of the reflectance, i.e.,

$$\frac{Y_m}{Y_0} = \frac{1 - R_m}{1 + R_m}, \tag{3}$$

where $Y_0 = A/(\rho c)$, A is the cross-sectional area of the ear canal at the location of the microphone, ρ is the density of air in the ear canal, c is the wave velocity of sound in the ear canal, and R_m is the reflectance of the ear at the measurement microphone in the ear canal. This expression derives from the French physicist Fresnel's work (1823) on electromagnetic wave propagation in different media (Born and Wolf, 1999). For sound waves, it assumes a one dimensional waveguide with plane wave propagation. The reflectance of the ear, R_m, includes the acoustic delay in the ear canal.

Sound delivered to the ear should be quantified in terms of the forward-going sound intensity or sound pressure wave,[1] the forward-going sound wave not being contaminated by standing waves in the ear canal (Farmer-Fedor and Rabbitt, 2002). To obtain the forward-going or incident sound pressure wave, P_i, we note that the sound pressure at the microphone (P_m), is given by

$$P_m = P_i + P_r, \tag{4}$$

where P_r is the reflected or backward-going sound pressure wave, and that

$$R_m = \frac{P_r}{P_i}, \tag{5}$$

the reflectance at the measurement microphone being the ratio of the reflected planar sound pressure wave and the incident planar sound pressure wave. From Eqs. (4) and (5) we obtain Eq. (1). For the measurement of hearing thresholds, $|P_m^{\text{th}}|$ is the hearing threshold level in pascals at the measurement microphone, a value that can alter with the location of the microphone in the ear canal, particularly at high frequencies due to ear canal standing waves. The quantity $|1+R_m|$ also varies with position along the ear canal due to standing waves. The result is that $|P_m^{\text{th}}|$ and $|1+R_m|$ covary with position along the ear canal so that $|P_i^{\text{th}}|$ remains constant, i.e., $|P_i^{\text{th}}|$ does not vary as a function of position along the ear canal. Note that the phase of P_i^{th} does depend on the location of the microphone in the ear canal due to the acoustic delay of sound propagation along the ear canal adding to the phase.

The magnitude of the incident (forward-going) sound pressure wave quantifies *in situ* the sound incident on the eardrum. The sound incident on the eardrum is a suitable calibration reference for calculating hearing thresholds but it is not the sound transmitted to the middle ear. The impedance mismatch between the ear canal and the middle ear means that some of the incident sound pressure wave is reflected at the eardrum; we will quantify this reflection of sound by the reflection coefficient *at the eardrum* (not at the microphone), designated R_{tm}. The reflected wave at the eardrum is then given by $R_{tm}P_i$ and that portion of the incident wave transmitted to the middle ear is then given by $(1-R_{tm})P_i$. Therefore, the magnitude of the fraction of the incident wave transmitted to the middle ear at hearing threshold is given by

$$|P_i^{th}| = |(1 - R_{tm})P_i^{th}|. \quad (6)$$

Equation (6) makes no assumptions about how sound is transmitted through the middle ear (e.g., the question of transmission losses through the middle ear) being only an expression of that portion of the incident wave in pascals *transmitted to the* middle ear. The quantity R_{tm} is obtained from R_m by correcting for the acoustic delay between the microphone and the eardrum. It would be nice if we could further quantify the reflectance at the stapes footplate but we do not know the acoustic delay through the middle ear.

Using a calibrated sound source and microphone, we can obtain the input admittance (Y_m) and reflectance (R_m) of the ear from the measurement of sound pressure in the ear canal. Behavioral hearing thresholds can then be expressed in terms of the incident planar sound pressure wave, a value uncontaminated by ear canal standing waves. Here we examine hearing thresholds in terms of (i) $|P_i|$, the magnitude of the incident or forward-going sound pressure wave, and (ii) $|P_t|$, the magnitude of the portion of the incident sound pressure wave transmitted to the middle ear. We compare these hearing thresholds with (i) hearing thresholds obtained in terms of the total sound pressure measured in the ear canal and (ii) hearing thresholds obtained in terms of the sound pressure measured in a Zwislocki (DB100) coupler.

II. METHOD

Thirteen females, aged 20–30, with no neuro-otological history, served as subjects for this study. This study was completed with the approval of the human ethics committee, Indiana University, Bloomington.

Signal generation and data acquisition was computer controlled using a Mimosa HearID system with version R4 software module with a type II PCMCIA soundcard, coupled to an Etymotic Research 10CP probe assembly, the microphone signal amplified 40 dB and digitized at a rate of 48 kHz. Microphone sensitivity was 50 mV/Pa; sound pressure measurements were corrected in software for the frequency response of the microphone. Fourier analysis was performed with a 2048 point fast Fourier transform, data analysis restricted to 256 points and an upper frequency limit of 6 kHz. The eartip was sized to each ear with eartip size providing the dimension for cross-sectional area, A, for the calculation of the characteristic impedance (for further discussion on ear canal area estimation, see Voss and Allen, 1994, and Keefe and Abdala, 2007). The Thevenin equivalent acoustic impedance and sound pressure of the probe assembly was determined using four cavities of known acoustic impedance and solving four simultaneous equations with two unknowns, Z_s and P_s, the source impedance and sound pressure, visco-thermal effects being accounted for in determining cavity lengths (Allen, 1986; Voss and Allen, 1994; Keefe, 1984). Cavity calibration to obtain Z_s and P_s was performed prior to each day of data acquisition. The probe assembly was inserted in the ear canal with the goal of the distal end of the eartip being flush with the entrance to the ear canal. The ear canal sound pressure frequency response was obtained from sound pressure measurements in the ear canal of one ear of each subject using a sweep frequency or chirp stimulus and the load admittance (Y_m) calculated by solving Eq. (2). Hearing thresholds in dB sound pressure level (SPL) at the measurement microphone for pure tones were obtained using a method of limits with a 1 dB step-size and six reversals. Hearing thresholds were measured within the frequency range 250–6000 Hz.

Equivalent threshold sound pressure levels (ETSPLs) were calculated from the voltage delivered to the ER10CP probe at each frequency at behavioral threshold multiplied by the sound pressure per volt measured at the behavioral test frequencies in a Zwislocki (DB100) coupler with a condenser microphone.

Data analysis was performed in MATLAB.

III. RESULTS

A. $|1+R_m|$, $|1-R_m|$, and the phase of R_m

Figure 1 shows $|1-R_m|$ and $|1+R_m|$, and the phase of R_m, as a function of frequency for eight subjects. $(1-R_m)$ and $(1+R_m)$ are the numerator and denominator terms of Eq. (3) for the normalized input admittance of the ear (Y_m/Y_0). In a hard-walled cylinder terminated at a right angle by a hard-walled boundary, the magnitude of the input admittance [Eq. (3)] is a maximum when $|1+R(f)|$ is a minimum, when the phase of R is π and the magnitude of R is 1. *A phase of π for the reflectance defines the standing wave frequency*, the standing wave frequency denoted in the phase plots of Figs. 1(a)–1(d) by " SWF." The input admittance of the ear is dominated by the ear canal, the ear canal being reasonably represented by a uniform cylinder up to 6 kHz (Stinson, 1985), and so the standing wave frequency is defined by the angle of π.

In Fig. 1, $|1+R_m|$ has a spectrum consistent with a standing wave at the microphone in the ear canal. In panels (a)–(d), the characteristic notch in the magnitude spectrum is evident, the notch frequency being associated with the standing wave frequency. The frequency corresponding to the reflectance phase of −0.5 cycles, the standing wave frequency, is not exactly the same as the frequency where $|1+R_m|$ is a minimum due to $|R_m|$ varying with frequency (when $|R_m|$ is independent of frequency, the two are the same). In panels (e)–(h), $|1+R_m|$ decreases as a function of frequency, no notch being evident because the standing wave frequency is above the upper frequency limit (6 kHz). The standing wave frequency may be predicted from the straight line fit to the reflectance phase. The term $(1+R_m)$ is also the denominator term in Eq. (1), the magnitude of which should have the same frequency-dependence as the magnitude of the sound pressure measured at the microphone due to standing waves in the ear canal. The numerator term for admittance magnitude, $|1-R_m|$, has a general tendency to have an opposite frequency-dependence to $|1+R_m|$. For a hard-walled cavity, $|1-R_m|$ will have a maximum value when the phase of R is π and the magnitude of R is 1, the same condition for the minimum in $|1+R_m|$. In the ear, as for the minimum of $|1+R_m|$, the frequency where $|1-R_m|$ is a maximum may not coincide with the standing wave frequency due to $|R_m|$ varying with frequency.

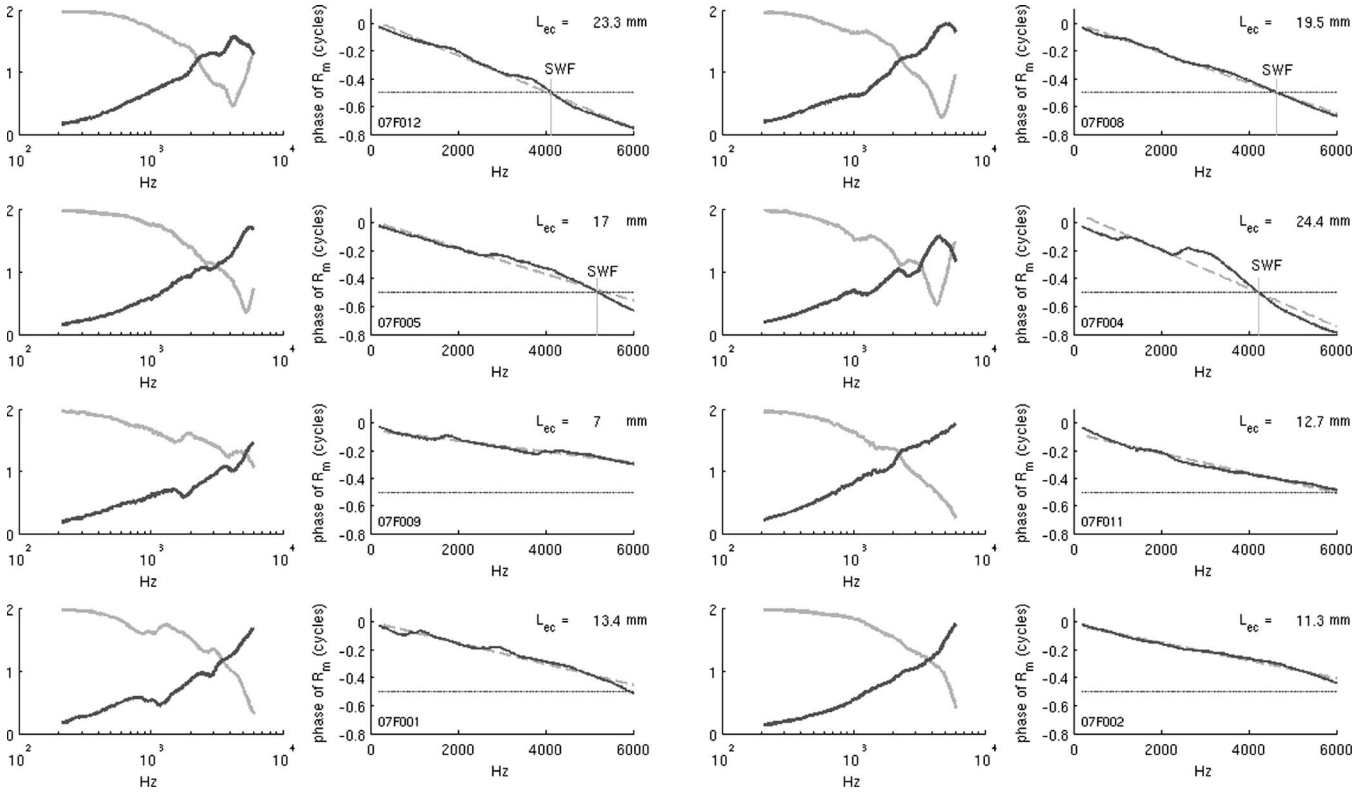


FIG. 1. $|1-R_m|$ and $|1+R_m|$, and the phase of R_m, as a function of frequency for eight subjects. $|1+R_m|$ has a spectrum consistent with a standing wave at the microphone in the ear canal; in four cases, a notch in the magnitude spectrum is evident. The reflectance phase corresponding to the magnitude data identifies the standing wave frequency by the intersection of a straight line at a phase of −0.5 cycles and the reflectance phase. In four cases, no standing wave frequency is identified, the standing wave frequency being above 6 kHz. The phase of the reflectance in each case has been fitted with a straight line. The standing wave frequency may be predicted from this straight line fit to the reflectance phase.

The phase of the reflectance is presumably dominated by ear canal acoustic delay. In Fig. 1, the phase of the reflectance in each case has been fitted with a line of best fit. The slope of this line of best fit provides an estimate of the ear canal acoustic delay from which the length of the ear canal between the eardrum and the microphone can be derived. It is notable that for seven of the eight subjects (excluding 07F004), the phase of the reflectance is well described by the line of best fit. This argues for sound to predominantly be reflected from a single site of reflection (the eardrum) with a pure delay, consistent with planar sound propagation in the ear canal. The contribution of the tympanic membrane/middle ear to the reflectance phase appears to be small, a not surprising finding, particularly for that part of the frequency range above 1 kHz where the input impedance of the middle ear in humans is predominantly resistive.[2] A poor fit of a straight line to the reflectance phase data requires multiple sites of reflection.

From the slope of the linear regression of reflectance phase versus frequency, the distance from the microphone to the eardrum was calculated for each subject, this estimate (in millimeters) given in Fig. 1 in the phase data panels.

B. Quantifying the sound delivered to the ear

Figures 2 and 3 show hearing threshold levels in terms of (i) the SPL measured at the microphone [SPL $=20\log(|P_m^{\text{th}}|/P_{\text{ref}})$ (dB)], where $P_{\text{ref}}=0.000\,02$ Pa, (ii) the incident sound pressure wave [FPL$=20\log(|P_t^{\text{th}}|/P_{\text{ref}})$ (dB)], (iii) the portion of the incident wave transmitted to the middle ear [TPL$=20\log(|P_t^{\text{th}}|/P_{\text{ref}})$ (dB)], and (iv) ETSPL (dB), as a function of frequency for the eight subjects of Fig. 1. Figure 2 shows the four subjects whose ear canal standing wave frequency is below 6 kHz. The SPL at the measurement microphone is larger than the incident SPL below 3 kHz. This is to be expected at low frequencies where, as $|R_m|\to 1$ and the phase of $R_m\to 0$, Eq. (5) reduces to $P_i \simeq P_m/2$ and so the SPL measured at the microphone should tend to 6 dB larger than the incident SPL. As frequency increases, any discrepancy between SPL and FPL will depend on the value of $|R_m|$ and the difference in phase between the incident and reflected waves. At 4 kHz, SPL is less than FPL, this being the frequency region where the ear canal standing wave will have the most effect on the measurement of sound pressure at the probe microphone. The standing wave frequency for the four cases shown in Fig. 2, based on the slope of the phase of the reflectance, is 4.1, 4.6, 5.2, and 4.2 kHz, respectively. Note that the magnitude of the difference between the SPL measured at the probe microphone and the SPL in terms of the incident pressure wave near the standing wave frequency is dependent on the reflectance magnitude near this frequency. Hearing thresholds, in terms of that portion of the incident sound pressure wave transmitted to the middle ear, show no particular frequency-dependence across the four subjects; these thresholds better quantify the signal the cochlea is receiving but are qualified by the accuracy of estimating the length of the ear canal. Hearing thresholds in

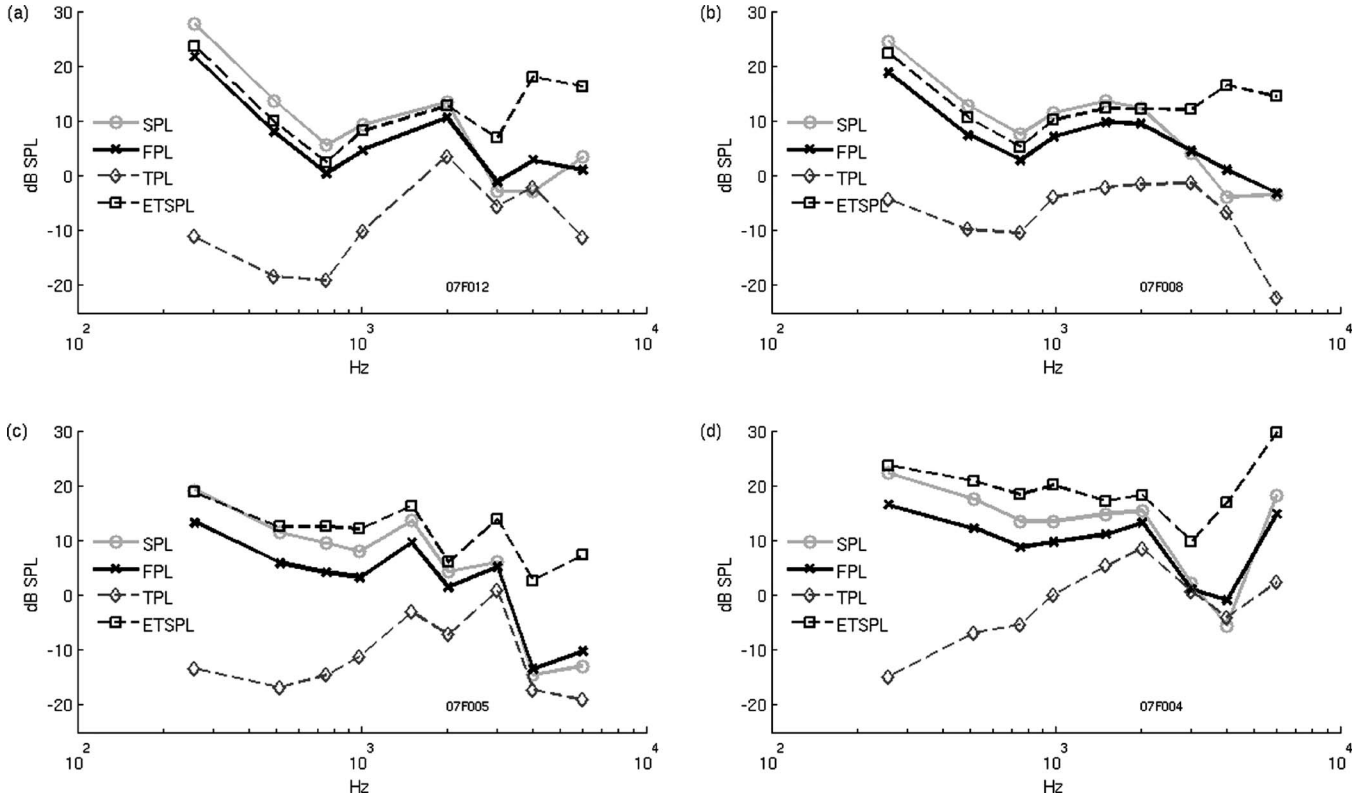


FIG. 2. Hearing threshold levels in terms of (i) the sound pressure measured at the microphone [SPL=20 log($|P_m^{th}|/P_{ref}$) (dB)], (ii) the incident sound pressure wave [FPL=20 log($|P_t^{th}|/P_{ref}$) (dB)], (iii) the fraction of the incident wave transmitted to the middle ear [TPL=20 log($|P_t^{th}|/P_{ref}$) (dB)], and (iv) ETSPL, as a function of frequency for the four subjects whose ear canal standing wave frequency was *below* 6 kHz. P_{ref}=0.000 02 Pa.

terms of ETSPL, based on calibration of the ER10CP ouput in a DB100 coupler measured with a coupler microphone, show reasonable agreement with the SPL configuration up to 2 kHz. At low frequencies, where the ear canal and DB100 effectively reduce acoustically to simple volumes, SPL and ETSPL should be similar in value if the ear canal acoustic volume is similar to the DB100 acoustic volume. Above 2 kHz, ETSPL configuration departs significantly from the configurations for SPL and FPL, although SPL is confounded by standing waves. The discrepancy between ETSPL and FPL demonstrates the coupler-based calibration error in quantifying SPLs at the eardrum.

Figure 3 shows hearing thresholds for the four subjects whose ear canal standing wave frequency at the measurement microphone was above 6 kHz. As for Fig. 2, hearing thresholds for SPL are greater than FPL up to 3 kHz. If the ear canal standing wave frequency is well above 6 kHz, then any observed convergence between the values of SPL and FPL would be mostly attributable to the value of $|R_m| \rightarrow 0$. This is the case for panel (a) where SPL remains larger than FPL up to 6 kHz but converges to FPL as frequency increases. Figure 1 for subject 07F009, corresponding to panel (a) of Fig. 3, shows 07F009 to be the ear with the shortest length from eardrum to microphone and so has the highest ear canal standing wave frequency; based on the length estimate of 7 mm, the standing wave frequency is 12.7 kHz. In panels (b)–(d), ears where the eardrum to microphone acoustic length corresponds to standing wave frequencies in the range 6.6–7.9 kHz, FPL exceeds SPL at 6 kHz [and at 4 kHz in panel (b)]. As for Fig. 2, SPL for these subjects is confounded by standing waves at one or more frequencies. Hearing thresholds in terms of that portion of the incident sound pressure wave transmitted to the middle ear (TPL) show no particular frequency-dependence across the four subjects, as observed in Fig. 2. Hearing thresholds in terms of ETSPL, similar to Fig. 2, show divergence in configuration relative to SPL and FPL above 2 kHz, although least so in panel (a) where the standing wave frequency is well above 6 kHz.

It is to be expected that the transmission line properties of the ear canal will preclude either ETSPL or SPL being an accurate estimate of the sound signal at the eardrum. Hearing thresholds expressed in terms of the incident or transmitted sound pressure are not confounded by ear canal standing waves. Either would serve as a suitable replacement for the current practice of hearing thresholds expressed in terms of hearing level or SPL calibrated in a 6cc or 2cc coupler.

IV. DISCUSSION

The acoustic input admittance of the ear can be quantified by using a sound delivery and measurement system with the Thevenin equivalent acoustic parameters, P_s and Z_s, determined using a cavity calibration procedure (e.g., Allen, 1986). The sound signal delivered to the ear can then be quantified in terms of the forward-going sound pressure (or forward-going sound intensity) wave, providing a calibrated signal that is not affected by standing waves in the ear canal (Farmer-Fedor and Rabbitt, 2002). This calibration is valid up to 6 kHz, the plane wave assumption up to 6 kHz being

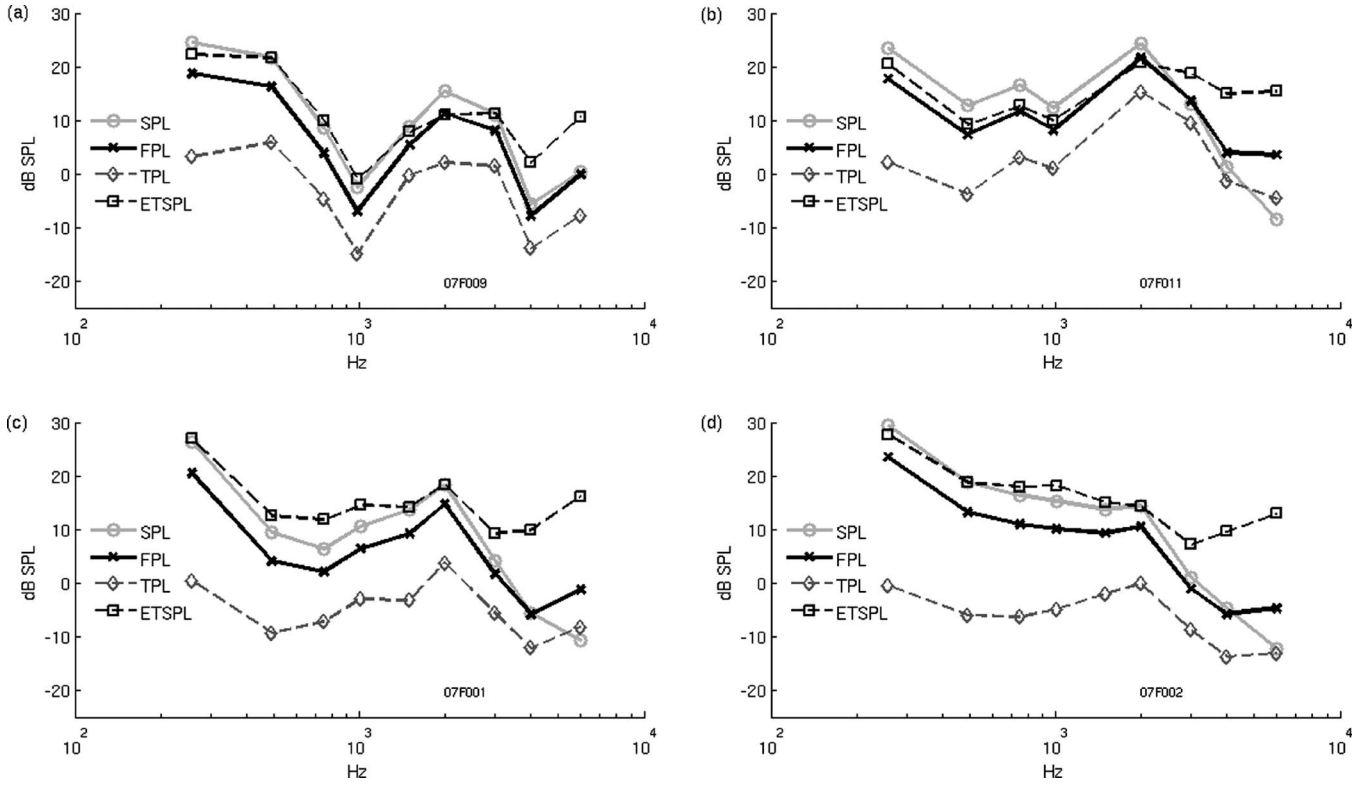


FIG. 3. Hearing threshold levels in terms of (i) the sound pressure measured at the microphone [SPL=20 log($|P_m^{th}|/P_{ref}$) (dB)], (ii) the incident sound pressure wave [FPL=20 log($|P_i^{th}|/P_{ref}$) (dB)], (iii) the fraction of the incident wave transmitted to the middle ear [TPL=20 log($|P_t^{th}|/P_{ref}$) (dB)], and (iv) ETSPL, as a function of frequency for the four subjects whose ear canal standing wave frequency was *above* 6 kHz.

supported by reflectance phase data that are well described by a pure delay and a single site of reflection at the eardrum. Expressing the sound delivered to the ear in terms of $|P_i|$ is preferable to the current convention of not measuring the acoustic signal *in situ* and estimating the sound pressure from measurements in a calibration cavity or artificial ear. Neither the sound pressure measured at the microphone nor sound pressure calibrations in a coupler accurately quantify the sound pressure at the eardrum. The availability of commercial probe assemblies that house both a speaker and a microphone (e.g., ER.10C) and the capacity to quantify the forward-going sound pressure wave argues for pure tone audiometry to be revised so that the acoustic signal is calibrated in each and every ear canal, thereby removing any contamination by standing waves.

The value of R used in this study to calculate hearing thresholds in terms of the incident and transmitted sound pressure was that obtained to a 60 dB (peak) SPL chirp. It is expected that the cochlear input impedance at hearing threshold stimulus levels contributes to R (Allen, 2001), perhaps generating hearing threshold microstructure (Elliot, 1958). At high stimulus levels, cochlear input impedance has been found to be predominantly resistive over much of the frequency range examined in this study (Aibara *et al.*, 2001). Obtaining R to stimulus levels near hearing threshold may furnish a more accurate estimate of the incident sound pressure wave at hearing threshold. However, the increase in signal averaging time to obtain such a value of R (and the use of a pure tone stimulus rather than a chirp to exclude nonlinear stimulus interaction in the cochlea) may render threshold reflectance measurements clinically impractical.

The portion of the incident sound pressure wave transmitted to the middle ear is an attempt to estimate the signal the cochlea receives from the sound pressure incident on the eardrum. The impedance mismatch between the ear canal and the middle ear determines the percentage of the sound incident on the eardrum that is reflected. However, the necessity of calculating the acoustic length from the microphone to the eardrum only applies to calculating hearing thresholds in terms of the transmitted sound pressure. Knowledge of acoustic length is not required for calibration in terms of the forward-going sound pressure wave. The uncertainty in calculating acoustic length argues for expressing hearing thresholds in terms of the incident or forward-going sound pressure as the preferred calibration.

[1]For a plane sound wave, sound intensity and sound pressure are related by the equation $I=P^2/Z_0$ (Lighthill, 1978).
[2]It is likely that the phase oscillates about some mean value over the frequency range that the input impedance is predominantly resistive (O'Connor and Puria, 2008, Parent and Allen, 2007).

Aibara, R., Welsh, J. T., Puria, S., and Goode, R. L. (**2001**). "Human middle-ear sound transfer function and cochlear input impedance," Hear. Res. **152**(1–2), 100–109.
Allen, J. B. (**1986**). "Measurement of eardrum acoustic impedance," in *Peripheral Auditory Mechanisms*, edited by J. L. Hall, J. B. Allen, A. Hubbard, S. T. Neely, and A. Tubis (Springer-Verlag, New York), pp. 44–51.
Allen, J. B. (**2001**). "Nonlinear cochlear signal processing," in *Physiology of the Ear*, Second edition, edited by A. F. Jahn and J. Santos-Sacchi (Singular Thomson Learning, San Diego, CA), pp. 393–442

Allen, J. B., Jeng, P. S., and Levitt, H. (**2005**). "Evaluation of human middle ear function via an acoustic power assessment," J. Rehabil. Res. Dev. **42**, 63–78.

Born, M., and Wolf, E. (**1999**). *Principles of Optics*. (Cambridge University Press, New York).

Burkhard, M. D., and Corliss, E. L. (**1954**). "The response of earphones in ears and couplers," J. Acoust. Soc. Am. **26**, 679–685.

Elliot, E. (**1958**). "A ripple effect in the audiogram," Nature (London) **181**, 1076.

Farmer-Fedor, B. L., and Rabbitt, R. D. (**2002**). "Acoustic intensity, impedance and reflection coefficient in the human ear canal," J. Acoust. Soc. Am. **112**, 600–620.

Feeney, M. P., Grant, I. L., and Marryott, L. P. (**2003**). "Wideband energy reflectance measurements in adults with middle-ear disorders," J. Speech Lang. Hear. Res. **46**, 901–911.

Fowler, E. P., and Wegel, R. L. (**1922**). "Audiometric methods and their applications," Trans. Am. Laryngol. Rhinol. Otol. Soc., 98–132.

Hazlewood, C., Jeng, P., Withnell, R. H., and Allen, J. B. (**2007**). "How does the power delivered to the ear relate to hearing?" American Auditory Society Meeting, Scottsdale, AZ.

Keefe, D. H. (**1984**). "Acoustical wave propagation in cylindrical ducts: Transmission line parameter approximations for isothermal and nonisothermal boundary conditions," J. Acoust. Soc. Am. **75**, 58–62.

Keefe, D. H., and Abdala, C. (**2007**). "Theory of forward and reverse middle-ear transmission applied to otoacoustic emissions in infant and adult ears," J. Acoust. Soc. Am. **121**, 978–993.

Keefe, D. H., Bulen, J. C., Arehart, K. H., and Burns, E. M. (**1993**). "Ear-canal impedance and reflection coefficient in human infants and adults," J. Acoust. Soc. Am. **94**, 2617–2638.

Keefe, D. H., Ling, R., and Bulen, J. C. (**1992**). "Method to measure acoustic impedance and reflection coefficient," J. Acoust. Soc. Am. **91**, 470–485.

Lighthill, J. (**1978**). *Waves in Fluids*. (Cambridge University Press, Cambridge, UK).

Neely, S. T., and Gorga, M. P. (**1998**). "Comparison between intensity and pressure as measures of sound level in the ear canal," J. Acoust. Soc. Am. **104**, 2925–2934.

O'Connor, K. N., and Puria, S. (**2008**). "Middle-ear circuit model parameters based on a population of human ears," J. Acoust. Soc. Am. **123**, 197–211.

Rabinowitz, W. M. (**1981**). "Measurement of the acoustic input immittance of the human ear," J. Acoust. Soc. Am. **70**, 1025–1035.

Parent, P., and Allen, J. B. (**2007**). "Wave model of the cat tympanic membrane," J. Acoust. Soc. Am. **122**, 918–931.

Scheperle, R. A., Neely, S. T., Kopun, J. G., and Gorga, M. P. (**2008**). "Influence of in situ, sound-level calibration on distortion product otoacoustic emission variability,"J. Acoust. Soc. Am. **124**, 288–300.

Siegel, J. H. (**1994**). "Ear-canal standing waves and high-frequency sound calibration using otoacoustic emission probes," J. Acoust. Soc. Am. **95**, 2589–2597.

Stinson, M. R. (**1985**). "The spatial distribution of sound pressure within scaled replicas of the human ear canal," J. Acoust. Soc. Am. **78**, 1596–1602.

Stinson, M. R., Shaw, E. A., and Lawton, B. W. (**1982**). "Estimation of acoustical energy reflectance at the eardrum from measurements of pressure distribution in the human ear canal," J. Acoust. Soc. Am. **72**, 766–773.

Voss, S. E., and Allen, J. B. (**1994**). "Measurement of acoustic impedance and reflectance in the human ear canal," J. Acoust. Soc. Am. **95**, 372–384.

Voss, S. E., Rosowski, J. J., Merchant, S. N., Thornton, A. R., Shera, C. A., and Peake, W. T. (**2000**). "Middle ear pathology can affect the ear-canal sound pressure generated by audiologic earphones," Ear Hear. **21**, 265–274.

Spectral integration under conditions of comodulation masking release

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 25 March 2008; revised 14 August 2008; accepted 17 December 2008)

Detection of a pure tone signal in a narrowband noise masker can be improved by the introduction of coherently amplitude modulated masker bands in neighboring frequency regions, an effect called comodulation masking release (CMR). Experiment 1 tested the hypothesis that detection of a spectrally complex signal in a comodulated masker critically depends on the signal/masker interaction, with best sensitivity in conditions where the signal introduces across-frequency stimulus envelope differences. Consistent with this hypothesis, thresholds for a multi-frequency signal differed by approximately 10 dB depending on the relative patterns of signal/masker interaction across frequency. In comodulated maskers, there was no improvement in threshold relative to the single-frequency signal threshold even in cases where the multi-frequency signal introduced across-frequency envelope differences. Experiment 2 tested conditions that have previously been associated with large spectral integration in comodulated but not random maskers. Results depended on the masker configuration used as the reference condition, with comparable integration for random and comodulated noise in some cases. The results suggest that CMR obtained with a pure tone signal can differ greatly from that obtained with a complex signal, and that spectral integration is inversely related to the amount of CMR under some conditions.
 [DOI: 10.1121/1.3075579]

PACS number(s): 43.66.Dc, 43.66.Ba [MW] Pages: 1612–1621

I. INTRODUCTION

Comodulation masking release (CMR) is the detection advantage associated with coherence in the patterns of masker amplitude modulation across frequency. This effect is typically demonstrated with a single pure tone signal, presented either in maskers of increasing bandwidth or in one of a family of narrowband noise maskers. While some of the masking release demonstrated with these stimuli may be due to within-channel cues (Schooneveldt and Moore, 1987; Berg, 1996; Verhey *et al.*, 1999), such as envelope beats, it is often argued that CMR in the strictest definition is due to across-channel comparisons (Schooneveldt and Moore, 1987; Carlyon *et al.*, 1989). Psychoacoustic models of across-channel cues underlying "true" across-channel CMR tend to be based on the change in envelope statistics across frequency (Buus, 1985; Cohen and Schubert, 1987; Piechowiak *et al.*, 2007), such as a reduction in envelope correlation or "listening in the dips." Naturally occurring sounds tend to be coherently modulated across frequency (Hall *et al.*, 1984; Nelken *et al.*, 1999), so the CMR paradigm could reflect a general auditory adaptation to the processing of natural sounds.

While most studies in the CMR literature have used spectrally simple signals, such as a pure tone presented for several hundred milliseconds, naturally occurring signals are usually spectrally complex. Several studies have demonstrated CMR for spectrally complex signals, such as speech (Grose and Hall, 1992; Festen, 1993; Kwon and Turner, 2001) or tonal complexes (Hall *et al.*, 1988; Grose *et al.*, 2005). Hall *et al.* (1988), for example, showed that the presence of a signal-free flanking band was not a prerequisite for obtaining masking release. In that study there were three maskers, all 30-Hz wide and centered on the fourth to sixth harmonics of 100, and the signal to be detected was a pure tone at one or more of those frequencies; all signals were presented at equal amplitude and with random starting phase. Thresholds in that study were lower for comodulated as compared to random maskers even when there was a signal present at all three masker frequencies.

Spectral integration for detection is defined as the advantage conferred by presenting multiple signal components in different frequency regions. Buus *et al.* (1986) showed that detection of relatively long pure tone signals in wideband noise improved by $10\log(\sqrt{n})$ for tones presented to spectrally independent auditory channels (referred to hereafter as the $\sqrt{n}$ rule). This model assumes that d' increases linearly with signal intensity. The $\sqrt{n}$ rule is also consistent with detection of pure tone signals in narrowbands of Gaussian noise and with detection of intensity increments in bands of noise (Grose and Hall, 1997). The form of spectral integration for tones presented in coherently amplitude modulated masker bands is less clear. One reason to expect less than a $\sqrt{n}$ reduction in threshold is that coherent masker modulation reduces the statistical independence of information distributed across frequency; another reason is that the inclusion of multiple signals is likely to change the across-channel cues available for detection, such that detectability of each tone individually could be reduced in the context of multiple tones. Some psychoacoustic studies report that integration in a comodulated masker is comparable to that observed in random noise (Grose *et al.*, 2005), while others

report less integration (Hall *et al.*, 1988; van den Brink *et al.*, 1992) or substantially more integration in coherently modulated as compared to random bands (Bacon *et al.*, 2002). The differences in integration across studies are related in part to the choice of single-component signal condition used as baseline. For example, Grose *et al.* (2005) defined integration relative to threshold for a single pure tone in the presence of four comodulated noise bands, whereas Bacon *et al.* (2002) defined it relative to pure tone detection threshold in a single band of noise. The implication of this difference will be revisited in the results and discussion of experiment 2.

Data on detection of multi-frequency signals have been interpreted as discriminating between potential cues in CMR. For example, Hall *et al.* (1988) argued that the CMR obtained with multi-frequency signals is difficult to reconcile with models based on envelope decorrelation or cued listening. van den Brink *et al.* (1992) asserted that CMR models based on either dip listening or envelope decorrelation predict no CMR for a multi-component signal in cases where the interactions between signals and masker produce compound output envelopes that are identical across frequency. If this is true for the auditory system, then thresholds for multi-component signals should be quite poor under conditions where the signal/masker interactions are identical across frequency, a result which would be associated with reduced estimates of spectral integration. The first experiment tested this hypothesis by controlling the regularity of signal/masker interactions via manipulation of signal amplitude and phase, with the prediction that masking release would be absent for conditions in which the signal/masker interactions produce identical envelope cues across frequency. The second experiment assessed spectral integration in a family of Gaussian noise bands, a stimulus for which Bacon *et al.* (2002) reported elevated levels of spectral integration when the masker bands were coherently sinusoidally amplitude modulated (SAM) as compared to random noise conditions.

II. EXPERIMENT 1

The manipulation of signal/masker interaction in this experiment can be understood in terms of the beating that occurs between pairs of tones. The inherent modulation of a narrowband noise is due to beating between components making up that band and is dependent on the phase and relative frequency of those components. Transposing those components uniformly up or down in frequency produces a comodulated band, whereas adjusting the phase or amplitude of one or more components can modify the pattern of inherent envelope fluctuation. Similarly, adding a pure tone signal to a narrowband masker can change the beating pattern of the summed stimulus. In the case of a multi-component signal, if each signal tone sums with the corresponding masker band in the same way—with the same amplitude and phase, and at the same relative frequency within the masker complex—then the associated envelope change is identical across frequency, whereas deviating from this configuration (e.g., randomizing signal phase) introduces potential envelope differences across frequency. The perceptual consequences of these differences for detecting a multi-component signal in comodulated maskers are the focus of the present experiment.

If the signal detection benefit associated with coherent masker modulation is based in part on the differences in stimulus envelope across frequency, then multi-signal configurations that reduce or eliminate those differences should likewise reduce the detection benefit. In the experiment that follows, the signal/masker phase and amplitude relationships across frequency are manipulated. It is hypothesized that signal parameters minimizing across-frequency envelope differences will likewise reduce any detection benefit associated with masker modulation coherence, with greater benefit under conditions of greater across-frequency envelope differences associated with addition of the signal. This effect will be quantified in terms of the threshold difference for detection of a tone as compared to a tonal complex, a quantity defined as spectral integration.

A. Methods

1. Observers

Observers were six adults, from 23 to 51 years old (mean of 34 years). All had thresholds of 20 dB HL or less at octave frequencies of 250–8000 Hz (ANSI, 1996), and none reported a history of chronic ear disease. All observers were practiced in psychoacoustical tasks at the outset of the experiment, having participated in at least one prior experiment.

2. Stimuli

Maskers were 15-Hz wide bands of Gaussian noise presented at 50 dB SPL per band, and there were either one or five bands. Band center frequencies were separated by a factor of ~1.9 and included 276, 525, 1000, 1904, and 3624 Hz. There were three types of maskers—on-signal (one single band for each of five frequencies), all-coherent, and all-random. The maskers were generated in the frequency domain at the outset of each threshold estimation track. A band of noise was generated based on random Gaussian draws defining the real and imaginary components contained within the masker passband. In the *all-coherent* conditions the same set of random draws was used to generate all five masker bands. In the *all-random* conditions each band was generated based on independent random draws. The stimulus generation array was 2^{17} points in length; when converted to the time domain and played at a sampling rate of 12 207 Hz, this produced a 10.7-s stimulus that repeated seamlessly and played continuously over the course of a threshold estimation track.

The signal was a pure tone or a set of five pure tones, with frequencies corresponding exactly to the center frequencies of the masker bands. There were five single-frequency signal conditions, with a single pure tone serving as the signal, and either a single masker band (on-signal) or a family of five masker bands present (all-coherent or all-random). For the all-coherent and all-random masker types there were four additional multi-frequency signal conditions, defined in terms of the relative level of tones (*equal* or *normalized*) and

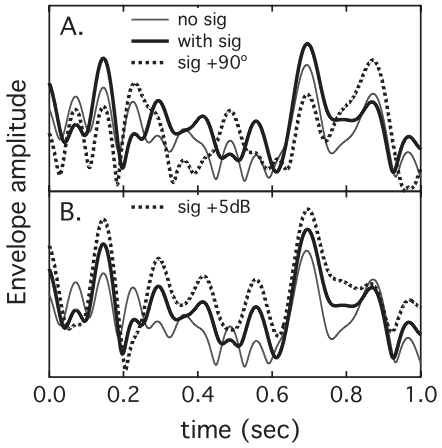


FIG. 1. Hilbert envelopes of example stimuli are plotted as a function of time. The masker, indicated with thin gray lines in both panels, was a 15-Hz wide band of noise scaled to 50 dB SPL. The thick black lines in both panels show the envelope of that same masker sample summed with a 45-dB SPL pure tone signal at the masker center frequency. The dotted lines in each panel indicate the different envelope effects obtained by incrementing the signal starting phase by 90° [panel (A)] or incrementing the signal amplitude by 5 dB [panel (B)].

the starting phase of the signal tones (*fixed-*ϕ or *random-*ϕ). The normalized level tones were individually adjusted in amplitude based on detection threshold in the associated single-frequency conditions, with either all-coherent or all-random maskers. As such, each tone was presented at equal level in dB SL, estimated independently for each observer and for each masker type. The starting phases of signal tones in the multi-signal conditions were either based on a single random draw or five independent draws from a uniform distribution ($0-2\pi$). This difference has consequences for the all-coherent condition. In the all-coherent random-ϕ condition each signal tone had a different effect on the envelope of the band to which it was added. In contrast, signals in the fixed-ϕ condition had a similar effect on envelopes across frequency in the normalized level condition and identical effects in the equal level condition. Signals were generated in the frequency domain using similar methods as those used to generate the maskers in order to maintain precise control of the signal/masker relationship.

Figure 1 shows envelopes of example stimuli, illustrating the effects of signal phase and level manipulations. The Hilbert envelope associated with a 15-Hz wide band of noise is plotted as a function of time, shown with the thin gray lines in both panels. This 50-dB SPL masker sample was summed with a 45-dB pure tone signal at the masker center frequency, and the envelope of the result is plotted with the thick black line in each panel. The dotted lines in each panel indicate the different envelope effects obtained by incrementing the signal starting phase by 90° [panel (A)] or the signal amplitude by 5 dB [panel (B)]. On average adjusting starting phase of the signal does not affect the overall stimulus level, but it does have a marked effect on the envelope pattern; for a signal at −5 dB signal-to-noise ratio (SNR), randomizing signal starting phase reduces the envelope correlation across bands to a median of approximately $r=0.75$. Fixing starting phase and comparing signals at −5 and 0 dB SNRs, the overall level is incremented by a median of 1.8 dB and the envelope correlation across bands is relatively unchanged, with a median of $r=0.96$. In both cases, mismatches in signal parameters lead to differences in the envelope across bands. A signal level of 45-dB SPL was chosen for illustration purposes because observer thresholds in the single-signal, all-coherent masker condition were on the order of 45 dB. Likewise, a level mismatch of 5 dB was chosen because that was the median range of single-signal, all-coherent masker thresholds across frequency for an individual observer; as such, signal level adjustments in the normalized condition were on the order of 5 dB.

The arrays defining masker and signal stimuli were loaded into an RPvds circuit (TDT), and signal gating was applied with 50-ms raised-cosine ramps implemented in software. All signal tones present in a given interval were ramped on and off synchronously, with a total duration of 400 ms including ramps.

3. Procedures

Stimuli were presented in a three-alternative forced-choice paradigm, with the signal equally likely to be present in each interval. Those intervals were marked visually and separated by 250 ms. Feedback was provided after each observer response. Signal level was adaptively varied following a three-down, one-up rule estimating 79% correct (Levitt, 1971). Level was initially adjusted in steps of 4 dB, and reduced to 2 dB after the second track reversal. The track continued for a total of eight reversals, and the associated threshold estimate was computed as the average signal level at the last six track reversals.

Single-frequency thresholds were collected first. For each observer the five signal frequencies were assigned a random order, and thresholds were collected in blocks by signal condition. Three threshold estimates were collected in each condition, with a fourth estimate collected if the first three spanned a range of 3 dB or more; all estimates were averaged to generate a final threshold estimate. After completing the single-frequency conditions the multi-frequency conditions were likewise run in random order. In the normalized multi-signal conditions thresholds are reported in decibels relative to the lowest-level tone (i.e., the level of the tone at the frequency associated with the lowest single-frequency threshold).

B. Results

Results from all six observers were similar, so the mean threshold across observers is reported. Figure 2 shows thresholds as a function of signal condition, with error bars indicating ±1 standard deviation. Symbols reflect masker types, as indicated in the key above the figure. The left panel shows results for the single-frequency signal conditions, and those to the right show results for multi-signal conditions.

1. Single-frequency signals

The single-frequency signal conditions will be considered first. Thresholds in the on-signal and all-random masker conditions appear quite similar and consistent across frequency, with mean thresholds ranging from 52.1 to 53 dB

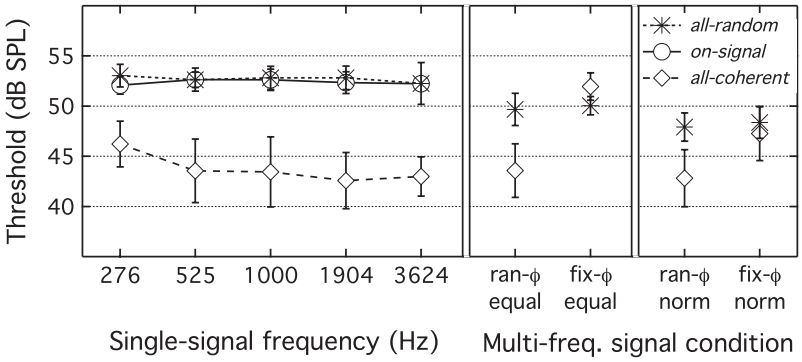


FIG. 2. Mean thresholds are plotted in dB SPL as a function of condition, with error bars showing ±1 standard deviation. Symbols reflect the masker type, either all-random (stars), on-signal (circles), or all-coherent (diamonds). The left panel shows single-frequency signal conditions, with frequency in hertz indicated on the abscissa. The middle panel shows multi-frequency signal conditions in which all signal tones were of equal amplitude. The panel on the right shows multi-frequency signal conditions in which the signal tones were presented at equal dB SL, determined separately for each observer and masker.

across conditions. In contrast, thresholds in the all-coherent masker are on average 8.9 dB lower than those in the other two conditions. This CMR appears relatively constant across frequency, with the exception of the lowest frequency, where the all-coherent threshold is elevated by approximately 3 dB.

These observations were confirmed with a repeated-measures analysis of variance (ANOVA), with three levels of MASKER (on-signal, all-random, and all-coherent) and five levels of FREQ (276, 525, 1000, 1904, and 3624 Hz). There was a main effect of MASKER ($F_{2,10}=90.33$, $p<0.0001$), a main effect of FREQ ($F_{4,20}=3.11$, $p<0.05$), and a significant interaction ($F_{8,40}=3.09$, $p<0.01$). Preplanned comparisons indicated that the on-signal and all-random masker conditions did not differ ($p=0.45$), but both were significantly different from the all-coherent masker conditions ($p<0.0001$). Computing CMR as the difference between mean threshold in the all-random and all-coherent masker conditions, masking release ranged from 6.5 to 10.1 dB, with the smallest CMR occurring for a signal added to the lowest frequency band. A paired t-test indicated that the masking release differed significantly between the 276- and 525-Hz signal frequencies ($t_5=5.09$, $p<0.005$ two-tailed). No other paired comparison for adjacent bands approached significance ($p>=0.25$).

2. Multi-frequency signals

Data for the multi-frequency signals appear in the right two panels of Fig. 2. Thresholds in the all-random masker, multi-signal conditions spanned 47.9–50 dB SPL for the four signal conditions. Relative to the best single-frequency condition for each observer, this represents a threshold reduction of 1.2–3.3 dB.

In contrast, thresholds in the multi-frequency signal, all-coherent masker conditions spanned a range of about 9 dB across the four signal conditions (42.8–52.0 dB). A mean threshold of 43.6 dB was obtained in the random-ϕ, equal signal condition, where each signal tone had equal amplitude and had a random starting phase. Thresholds in this condition were not significantly different from those of the single-frequency condition with the lowest threshold (1904 Hz; $t_5=0.79$, $p=0.46$). Similarly, thresholds in the random-ϕ, normalized signal condition were not significantly different from those for the single-frequency, 1904-Hz signal ($t_5=0.24$, $p=0.82$), with an average of 42.8 dB. Whereas thresholds for the multi-signal, random-ϕ signal conditions were not different from those in the comparable single-frequency conditions, thresholds in both of the fixed-ϕ signal, all-coherent masker conditions were significantly elevated relative to the comparable single-frequency signal conditions; these trends held for both the equal and normalized level conditions. Relative to the 1904-Hz single-frequency condition, thresholds in the fixed-ϕ, equal signal condition were elevated by 9.4 dB ($t_5=10.62$, $p<0.0001$) and those in the fixed-ϕ, normalized signal condition were elevated by 4.7 dB ($t_5=4.99$, $p<0.005$).

For both the equal and normalized signal conditions, thresholds with random-ϕ signal were lower for the all-coherent than the all-random maskers ($p<0.05$), consistent with a significant CMR. No such CMR was observed for the fixed-ϕ signal conditions.

C. Discussion

Results of the single-frequency conditions are similar to those reported previously in literature. Thresholds in the on-signal masker conditions were quite consistent as a function of frequency, with a mean of 52–53 dB SPL. This result is in line with those of Bos and de Boer (1966) under comparable conditions. Thresholds in the all-random and on-signal masker conditions were statistically indistinguishable, suggesting that the bands were sufficiently separated in frequency to preclude energetic masking effects with random flanking bands. Mean thresholds in the comodulated masker conditions were more variable as a function of signal frequency, spanning about 3 dB. The poorest thresholds were obtained at the 256-Hz signal frequency. An analogous frequency effect was reported by Hall *et al.* (1988), where the smallest CMR was obtained for a signal added to the lowest masker band.

The most interesting aspect of the present results is the spectral integration observed in multi-frequency signal conditions. Table I reports the mean improvement in threshold for a multi-component signal relative to the lowest of the five associated single-frequency thresholds as measured in the associated all-coherent or all-random masker condition and computed separately for each of the six observers. The standard error of the mean (sem) is noted below each estimate of integration. For a five-component signal, the $\sqrt{n}$ rule predicts a 3.5 dB threshold improvement. For the all-random masker, spectral integration in both equal signal conditions is more than two sems below that prediction, but integration is within one sem of that target for both conditions utilizing the normalized amplitude signal. This result is consistent with the

TABLE I. Estimates of spectral integration were computed as the lowest single-frequency signal threshold minus the associated multi-frequency signal threshold. The sem is indicated in parentheses. Assuming d' is proportional to signal intensity, integration of five independent cues is predicted to be 3.5 dB.

	Signal condition (phase/amplitude)			
Masker condition	Random-ϕ, equal	Fixed-ϕ, equal	Random-ϕ, normalized	Fixed-ϕ, normalized
All-random	1.55 (0.51)	1.19 (0.60)	3.30 (0.62)	2.87 (0.49)
All-coherent	−2.07 (0.89)	−10.45 (0.83)	−1.31 (0.69)	−5.76 (0.82)

interpretation that individual differences in thresholds across frequency in the all-random masker conditions were reliable, such that across-frequency adjustments characterizing the normalized signal condition achieved equal signal audibility across frequency.

In contrast to spectral integration computed for the all-random masker conditions, integration in the all-coherent masker conditions was uniformly *negative*. That is, best performance in a single-frequency condition was superior to that for a multi-component signal. This effect was most striking for the fixed-ϕ, equal signal condition, where thresholds rose to 10.45 dB with inclusion of all five signal tones. Thresholds in this condition were on average 51.9 dB, comparable to those in the single-frequency, all-random and the on-signal masker conditions. This threshold was also slightly poorer than the 50.0-dB threshold in the all-random masker, fixed-ϕ, equal signal condition ($t_5=3.19$, $p<0.05$). These comparisons suggest that fixing signal phase and level across frequency eliminates CMR and may elevate threshold above that obtained with a multi-frequency signal presented in random noise. One reason why thresholds for a fixed-ϕ, equal signal might be poorer in the all-coherent as compared to all-random masker conditions has to do with the redundancy of information across frequency in the all-coherent masker. If integration in the all-random masker is based in part on the benefits associated with having independent samples of signal-plus-masker available across frequency, then the fact that envelope patterns are identical across frequency in the all-coherent masker condition would reduce the available cues and increase threshold.

While integration is negative for all of the all-coherent masker, multi-frequency signal conditions, the magnitude of that effect is reduced for signal conditions associated with different signal/masker interactions across frequency. Performance was worst for the fixed-ϕ, equal signal condition, but normalizing signal tone level relative to single-frequency threshold improved performance by 4.7 dB, and randomizing signal tone starting phase improved thresholds by 8.4 dB, with a combined effect of 9.1 dB. These results are consistent with the hypothesis that spectral integration in the presence of coherently amplitude modulated maskers depends strongly on the signal/masker interaction and the resulting pattern of temporal envelopes across frequency.

One unpredicted result of the present study is the finding of negative integration for the random-ϕ, normalized signal condition with the all-coherent masker type. Mean thresholds in this condition rose by an average of 1.31 dB relative to thresholds in the best single-frequency condition. While this estimate of spectral integration is not significantly different from zero ($t_5=1.91$, $p=0.11$), the 95% confidence interval extends only up to 0.45 dB, well shy of the 3.5 dB predicted from a $\sqrt{n}$ rule. This outcome was not predicted at the outset of the experiment and suggests that integration in the presence of coherently modulated maskers may be less than that in random noise even in conditions where the signal/masker interaction is non-uniform across frequency. This finding is not without precedent. In one set of conditions, Hall *et al.* (1988) measured detection thresholds in a family of three continuous 30-Hz wide bands of noise, each with the same pattern of inherent modulation. That study reported results in terms of CMR, computed as the threshold in the multi-signal condition, with random signal starting phase, minus the mean single-frequency thresholds obtained with a single (on-signal) masker band; these values can be used to compute thresholds and CMR (Table I; Hall *et al.*, 1988). Thresholds for a single-frequency signal were lower in the presence of three coherently modulated masker bands relative to the on-signal threshold, an effect of 5.1 dB at 400 Hz, 9.9 dB at 500 Hz, and 8 dB at 600 Hz. Threshold for the three-tone signal, with tones at equal amplitude, was also reduced relative to the on-signal threshold, an effect of 6.3 dB. That is, the tone at 500 Hz was 3.6 dB more intense at threshold in the multi-signal condition as compared to the single-signal condition. When analyzed like the present data, this would be characterized as a −3.6 dB spectral integration.

In summary, the results of experiment 1 show that detection of a spectrally complex signal in a set of coherently modulated maskers is highly sensitive to the across-frequency envelope differences resulting from addition of a signal. There was no evidence of spectral integration in the all-coherent masker type even under conditions of robust across frequency cues. These results are also consistent with the hypothesis of van den Brink *et al.* (1992) that there should be no CMR for a multi-component signal in cases where the interactions between signals and masker produce compound output envelopes that are identical across frequency.

III. EXPERIMENT 2

The finding of reduced spectral integration in the context of coherently modulated masker bands stands in stark contrast to the conclusions of Bacon *et al.* (2002), where it was argued that spectral integration in the presence of coherently modulated bands can be substantially larger than that observed with random noise bands. In that study maskers were 100-Hz wide bands of Gaussian noise or noise that was sinusoidally modulated at a rate of 8 Hz. In one set of conditions thresholds were measured for a tonal signal at 500, 1000, or 2000 Hz, or a combination of all three frequencies. In each case there was a masker centered on each signal tone, but no "signal-free" maskers. Integration in Gaussian noise

or incoherently modulated noise was close to 2.4 dB, as expected by the $\sqrt{n}$ rule. In coherently modulated noise integration was on the order of 5.5 dB, a result which could not be explained in terms of psychometric function slope, but was interpreted instead as evidence that spectral integration and CMR effects are additive.

In contrast to the paradigm of experiment 1 in the present study, Bacon *et al.* (2002) did not measure single-frequency thresholds in the three-masker complex; integration was computed instead based on single-frequency thresholds measured in the presence of a single masker band. The purpose of experiment 2 was therefore to replicate and extend the findings of Bacon *et al.* (2002) to include thresholds for individual signal tones in the three-masker complex. It was hypothesized that integration would be not be "greater than expected" when computed relative to single-frequency thresholds measured in a multi-masker complex. Another motivation for experiment 2 was to determine whether integration with coherent masker modulation is greater for SAM noise bands than for conditions where inherent masker modulation determines envelope coherence, as in experiment 1. Results of experiment 2 were therefore expected to provide insight into the stimulus features that drive the amount of spectral integration for comodulated maskers.

A. Methods

1. Observers

Observers were eight adults, from 21 to 53 years old (mean of 32 years). All had thresholds of 20 dB HL or less at octave frequencies of 250–8000 Hz (ANSI, 1996), and none reported a history of chronic ear disease. All observers were practiced in psychoacoustical tasks at the outset of the experiment, having participated in at least one prior experiment. One observer had previously participated in experiment 1.

2. Stimuli

Maskers were 100-Hz wide bands of noise, presented at 55 dB SPL per band. There was either a single band (*on-signal*) or a family of three bands (*complex*), with band center frequencies of 500, 1000, and 2000 Hz. Masker bands were either Gaussian noise or noise that was SAM at 8 Hz. Each band of noise was generated in the frequency domain based on Gaussian random draws defining the real and imaginary components contained within the masker passband, with an array size of 2^{17} points. When converted to the time domain and played at a sampling rate of 12 207 Hz, this stimulus was 10.7 s in duration.

The signal was a pure tone or a set of three pure tones, with frequencies corresponding to the center frequencies of the masker bands. Single-frequency signal thresholds were measured in two conditions: once in the presence of an on-signal masker alone and once in the presence of the complex masker including all three masker bands. Signal level in the multi-frequency signal conditions was defined in terms of these single-frequency thresholds, comparable to the normalized signal conditions of experiment 1. The starting phase of the signal tones was always coherent across frequency; because the maskers were based on independent Gaussian noise samples across frequency, the starting phase of the signal was assumed to be of no special significance.

As in experiment 1, the arrays describing masker and signal stimuli were loaded into an RPvds circuit (TDT) and stimulus gating was applied with software ramps. The masker and signal tone(s) present in a given interval were ramped on and off synchronously with 50-ms raised-cosine ramps and a 300-ms steady state. Masker amplitude modulation, when present, was synchronized to the listening interval, such that the modulation in each 400-ms listening interval began in sine phase.

3. Procedures

As in experiment 1, stimuli were presented in a three-alternative forced-choice paradigm, with the signal equally likely to be present in each interval. Listening intervals were marked visually and separated by 350 ms. Feedback was provided after each observer response. Signal level was adaptively varied following a three-down, one-up rule estimating 79% correct (Levitt, 1971). Level was initially adjusted in steps of 4 dB, reduced to 2 dB after the second track reversal. The track continued for a total of eight reversals, and threshold was estimated as the average signal level at the last six track reversals.

Single-frequency thresholds were collected first and in random order blocked by frequency. After completing the single-frequency conditions the multi-frequency conditions were likewise run in random order. In multi-frequency conditions signal level at threshold is reported in decibels relative to the highest-level tone (i.e., the level of the tone at the frequency associated with the highest single-frequency threshold). This convention is different from that adopted in experiment 1, where the lowest-level tone was the reference; the highest-level reference was used here to facilitate comparison with the data of Bacon *et al.* (2002). At completion of the experiment thresholds were examined for stability. Data were replaced if thresholds across the three or four estimates spanned a range of 8 dB or more. In cases where single-frequency thresholds were replaced the associated multi-frequency conditions were likewise replaced using the new estimates of threshold to normalize signal tone level.

B. Results

Mean thresholds for each observer are plotted in Fig. 3 as a function of signal condition, with symbols reflecting the masker condition. Triangles show thresholds in Gaussian noise conditions, and squares show those in SAM noise; filled symbols correspond to thresholds obtained in an on-signal masker alone and open symbols correspond to those obtained in a complex of three masker bands. Observer number, which appears in the top right of each panel, was assigned based on rank order of thresholds for the multi-frequency signal, with signal level normalization based on complex masker data; this condition may be viewed as a rough indicator of sensitivity to multi-frequency signals. Despite the individual differences evident in the figure, thresh-

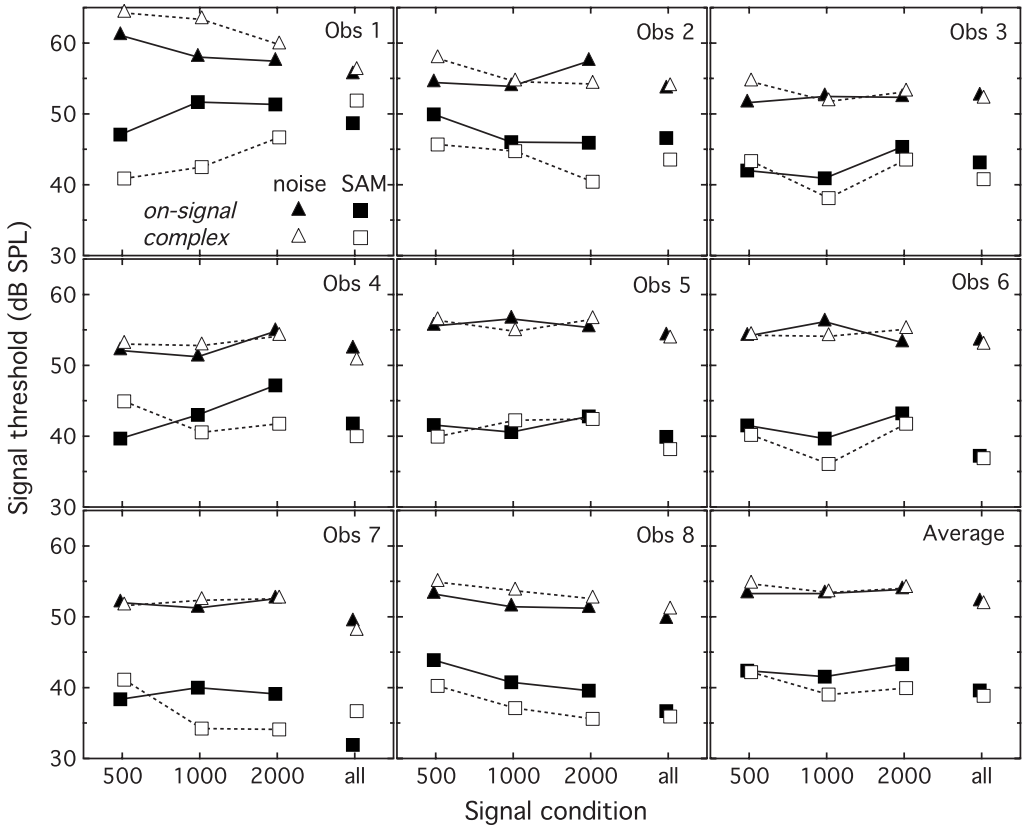


FIG. 3. Mean thresholds for individual observers are shown in each panel, as well as the mean across Obs 2–8. Thresholds are plotted in dB SPL as a function of signal condition, with thresholds in the multi-frequency signal conditions plotted relative to the most intense of the three tones normalized to single-frequency signal thresholds. Symbols indicate the masker condition, either noise (triangle) or SAM noise (square). Filled symbols indicate threshold for a pure tone in a single masker band or a complex signal where relative tone levels are normalized based on results obtained with on-signal maskers. Open symbols indicate pure tone thresholds obtained in the presence of three masker bands, as well as the associated complex signal thresholds.

old estimates were relatively stable within observer (with a median standard deviation of 1.3 dB), and several trends in the data are evident.

Data were replaced due to excessive variability in seven instances. In the original data of Obs 1, thresholds for the complex signal in SAM noise were quite variable, with two estimates near 40 dB SPL and two near 50 dB SPL. When these data were replaced all estimates were near 50 dB SPL. This threshold was 10 dB greater than the mean across observers and 8.3 dB greater than the next poorest threshold, suggesting that this observer's multi-frequency signal thresholds should be viewed caution. For that reason all statistical tests reported below were performed omitting data from Obs 1. While this omission affected the level of significance reported for each test, repeating these statistical tests with those data included did not change the conclusions reached below.

1. Single-frequency signals

The single-frequency signal data will be considered first. As indicated by the connected triangles in Fig. 3, thresholds in the random noise masker tended to be similar for the on-signal and complex maskers, with mean thresholds spanning from 53.3 to 54.6 dB across frequency and noise masker condition. A repeated-measures ANOVA was performed with three levels of FREQ (500 Hz, 1 kHz, and 2 kHz) and two levels of MASKER (on-signal and complex). There was no effect of MASKER ($F_{1,6}=5.35$; $p=0.06$), no effect of FREQ ($F_{2,12}=1.27$; $p=0.32$), and no interaction ($F_{2,12}=1.00$; $p=0.40$). Though the effect of masker did not reach significance, there was a trend ($p<0.1$) for higher thresholds in the complex than on-signal noise masker, consistent with the possibility of more masking for a tone presented in a three-masker complex.

Thresholds in the on-signal SAM noise masker conditions (filled squares) appear to be unaffected or inconsistently affected by signal frequency, with mean thresholds of 41.6 to 43.3 dB. There was some evidence of an improvement with increasing frequency for the complex SAM noise masker (open squares) in some observer's data (e.g., Obs 4, 7, and 8), but there were also striking counterexamples to this trend (Obs 1). A repeated-measures ANOVA was performed with three levels of FREQ (500 Hz, 1 kHz, and 2 kHz) and two levels of MASKER (on-signal and complex). There was a main effect of MASKER ($F_{1,6}=14.60$; $p<0.01$), no effect of FREQ ($F_{2,12}=2.15$; $p=0.16$), and no interaction ($F_{2,12}=2.33$; $p=0.14$). The main effect of MASKER reflects the fact that thresholds are on average 2.0 dB lower in the complex than on-signal SAM noise masker. Interpretation of these results is tempered by marked individual differences.

TABLE II. Estimates of spectral integration were computed as the maximum single-frequency signal threshold minus the associated multi-frequency signal threshold. The sem is indicated in parentheses. Assuming d' is proportional to signal intensity, integration of three independent cues is predicted to be 2.4 dB.

	Signal condition	
Relative levels based on	On-signal	Complex
Noise	2.59 (0.50)	3.32 (0.33)
SAM noise	5.01 (0.86)	3.95 (0.41)

2. Multi-frequency signals

Attention now turns to the multi-frequency signal conditions indicated at the right-hand side of Fig. 3, denoted "all" on the abscissa. In these conditions there were three masker bands present, each with a signal tone, and symbol shading indicates the single-signal conditions used to normalize the relative levels of the three signal tones, either the on-signal masker (filled symbols) or complex masker (open symbols) conditions. As previously, symbol shape reflects masker type, either noise (triangles) or SAM noise (squares). In contrast to experiment 1, thresholds in these conditions are plotted relative to the level of the tone associated with the poorest (highest) single-frequency threshold. In general multi-signal thresholds in the complex masker conditions fall 2 dB or more below the highest associated signal-signal thresholds, with only two exceptions. For Obs 1 the multi-signal threshold in the complex SAM condition was 5.2 dB higher than the highest associated single-signal condition. For Obs 3, the multi-signal threshold in the on-signal noise condition was 0.11 dB higher than the highest associated single-signal condition.

Table II shows the mean spectral integration across individual observers. When signal tone level was normalized based on thresholds from the on-signal, single-frequency conditions, the estimates of spectral integration differed for noise and SAM noise conditions. There was 2.4 dB more integration in the SAM noise, a difference that was statistically significant ($t_6=3.42$, $p<0.01$ one-tailed). Thresholds were similar for noise and SAM noise when thresholds were normalized using thresholds from the complex, single-frequency conditions ($t_6=1.18$, $p=0.28$ two-tailed). One unexpected finding is that estimates of spectral integration based on single-frequency thresholds in the complex masker are significantly greater than the expected 2.4 dB ($p<0.05$) for both the noise and SAM noise conditions.

C. Discussion

The stimuli used here were roughly based on those used by Bacon *et al.* (2002). That study used a slightly higher presentation level (57 compared to 55 dB per band) and longer duration (500 ms compared to 400 ms), and masker bands were "frozen" noise generated as the sum of equal amplitude tones with 2 Hz spacing rather than Gaussian noise. Despite these differences, results are consistent across analogous conditions of the two studies. In the study of Bacon *et al.* (2002) integration was computed as the difference in threshold for the multi-frequency signal in a complex masker relative to single-frequency signals in the associated on-signal maskers; that is, there were no signal-free bands in the single-frequency conditions. For the masker composed of bands at 500, 1000, and 2000 Hz mean integration was approximately 2.4 dB for the noise masker and 5.5 dB for the SAM noise masker. Analogous estimates from the present study were 2.6 and 5.0 dB, replicating the original finding of a significant difference. Estimates of integration based instead on single-frequency thresholds in a complex masker, the method used in experiment 1, resulted in comparable estimates of integration for noise and SAM noise. This finding suggests that the large value of integration reported by Bacon *et al.* (2002) for the SAM noise as compared to the Gaussian noise conditions can be attributed to the choice of reference condition: Estimates based on the on-signal band alone reference result in greater estimates of integration in comodulated bands than those based on the complex masker as the reference condition, likely due to the masking release associated with inclusion of flanking masker bands.

It is unclear how to account for the integration of approximately 3.5 dB in both noise and SAM noise computed relative to the complex masker reference, a value which is 1.2 dB greater than the expected 2.4-dB effect size. Grose and Hall (1997) reported that spectral integration for a family of pure tone signals in random narrowband noise followed the $\sqrt{n}$ rule. In that experiment, however, there was a masker at each signal frequency and no signal-free maskers. Hall *et al.* (1988) measured thresholds for one or more signal tones presented in a family of three continuous narrowband random noise maskers and reported less than expected integration (1.1 dB as compared to 2.4 dB). Using gated masker presentation, Grose *et al.* (2005) reported approximately 3 dB of integration for both random and coherently modulated noise, the integration expected for the four-frequency signal used in that study. This range of results suggests that integration for bandpass noise maskers could depend on the stimulus details, such as gated versus continuous presentation. The finding of comparable integration in noise and SAM noise in the present paradigm, however, suggests that these effects exist independent of masking release based on coherent modulation.

One factor that could affect estimates of spectral integration is the degree to which channel independence can be assumed. In the present paradigm thresholds for a 500-Hz signal measured in the random noise conditions were on average 1.3 dB greater in the complex than in the on-signal masker, suggesting that some of the improvement observed could be due to a release from across-channel masking (Moore *et al.*, 1990a).

IV. GENERAL DISCUSSION AND CONCLUSIONS

Experiment 1 provided support for the hypothesis that spectral integration in the presence of a coherently amplitude modulated masker depends on the signal/masker interaction differing across frequency. In cases where that interaction

was consistent across frequency there was no evidence of integration, and, in fact, thresholds were elevated by more than 10 dB relative to threshold for a single-signal tone. Thresholds improved under signal conditions associated with different patterns of signal/masker interaction. The best thresholds for a multi-frequency signal failed to show a positive spectral integration, however. One possible explanation for the lack of spectral integration in the context of coherently modulated masker bands has to do with the factors limiting performance. Langhans and Kohlrausch (1992) argued that detection of a brief tone in a frozen noise is limited by internal rather than external noise; this hypothesis was supported by the finding of better performance in diotic than monotic listening conditions, a result that would be expected if internal noise is independent across ears. Langhans and Kohlrausch (1992) noted that a similar diotic advantage is obtained with running comodulated noise (Cohen and Schubert, 1987; Schooneveldt and Moore, 1989), leading to the hypothesis that the auditory system can make use of the coherent envelope across frequency to reduce the effective variability of the external noise, leaving internal noise as the limit to performance. If accurate information about masker fluctuation allows observers to work at the limits of internal noise, then any benefits associated with spectral integration might be offset by corruption of the signal-free masker template.

Experiment 2 tested the hypothesis that the increased spectral integration for tones presented in amplitude modulated noise reported by Bacon *et al.* (2002) can be reconciled with previous literature by redefining integration referenced to the complex masker thresholds. Data were mostly consistent with that hypothesis. Defining integration relative to on-signal masker thresholds produced estimates of integration consistent with those reported by Bacon *et al.* (2002), whereas estimates based on the complex masker produced similar estimates for noise and SAM noise.

Results of experiment 1 provide no evidence of spectral integration in comodulated noise bands, but experiment 2 showed comparable integration for noise and SAM noise maskers. The differences in outcome between the experiments may well arise because of stimulus differences. A supplemental experiment described in the Appendix considered and rejected the possibility that differences in psychometric function slope could be responsible for these effects. Another possibility that we will consider briefly is that performance in the comodulated conditions was limited by internal noise for the stimuli used in experiment 1 but not for the stimuli used in experiment 2. In experiment 1 envelope coherence was manipulated via inherent modulation of each narrowband of noise; as a result, masker envelopes were identical across frequency prior to transduction by the auditory system. In experiment 2 the maskers were bands of independent Gaussian noise that had been sinusoidally amplitude modulated; as a result of random inherent modulation of these bands, envelopes across frequency were not perfectly correlated. If accurate representation of the masker alone envelope is necessary to effectively remove the masking associated with stimulus variability, then results of experiment 2 could be more strongly influenced by external noise than those of experiment 1. This reasoning is consistent with estimates of masking release. In experiment 1 inclusion of coherently modulated flanking masker bands improved threshold over that in the on-signal masker condition by 6.5–10.1 dB (mean of 8.9 dB). In experiment 2 thresholds in the three-band SAM noise improved by 0.2–3.4 dB (mean 2.0 dB) relative to the on-signal SAM masker threshold, suggesting that flanking masker bands were less beneficial to pure tone signal detection in the second experiment.

This interpretation suggests that spectral integration may be inversely related to the magnitude of CMR. The two studies on spectral integration in CMR previously reported from our laboratory are generally consistent with this idea. Comparing detection threshold for a pure tone signal in a single 30-Hz wide on-signal masker versus threshold in a three-band masker, Hall *et al.* (1988) reported a mean CMR of 9.9 dB: As in experiment 1, there was no evidence of spectral integration for a multi-frequency signal. Computing CMR in a similar way, the results of Grose *et al.* (2005) are consistent with a 3.8 dB CMR; as in experiment 2, there was robust spectral integration. Another difference between paradigms showing robust integration versus little benefit of additional signal components is continuous masker presentation in the former and gated masker presentations in the latter paradigms. Since CMR tends to be greater for continuous than gated stimuli (Fantini *et al.*, 1993; Hatch *et al.*, 1995), it could be difficult to tease apart effects associated with gating as compared to those related to the magnitude of CMR.

The most novel aspect of the present results is the demonstration that spectral integration under conditions associated with CMR depends critically on the details of signal/masker interaction. It is likely that integration may also depend on baseline performance with a single signal, with greater integration under conditions of poorer performance.

ACKNOWLEDGMENTS

This work was supported by NIH NIDCD Grant No. RO1-DC00739780. This manuscript benefited from comments of Joseph Hall and two anonymous reviewers.

APPENDIX

Four observers from experiment 2 subsequently participated in a supplemental experiment designed to determine whether psychometric function slope could be responsible for the differences in spectral integration observed for coherently amplitude modulated maskers in experiments 1 and 2. There were four stimulus conditions in total, each with a single pure tone signal at 1 kHz. Two conditions used stimuli identical to those described above for experiment 1; in these conditions the masker was a family of five 15-Hz wide masker bands, with modulation patterns being either all-coherent or all-random. The remaining two conditions used stimuli described above for experiment 2; in these conditions the masker was a set of three 100-Hz wide bands of noise, with either 8-Hz SAM or no modulation. In all cases the masker played continuously, and the signal was presented in one of three listening intervals.

TABLE III. Estimates of psychometric function slope (k) for a subset of stimulus conditions from experiments 1 and 2.

	Exp. 1		Exp. 2	
	All-coherent	All-random	Complex SAM	Complex noise
Obs 1	2.70	2.92	2.52	2.06
Obs 2	2.30	2.24	2.11	2.76
Obs 3	2.33	2.55	2.05	2.97
Obs 4	3.38	2.86	2.13	2.65
Mean	2.68	2.64	2.20	2.61

Psychometric functions were estimated in two stages of testing. In the first stage, a tracking procedure was used to estimate the 71% correct using a two-down, one-up tracking rule. Four estimates based on four reversals each were collected for each observer. The mean (m) and standard deviation (sd) of these four estimates were used to select five signal levels for each observer: m-2sd, m-sd, m, m+sd, and m+2sd. Percent correct was then estimated for these five signal levels. Data were collected in ten blocks, each with eight repetitions of each signal level presented in random order.

Resulting estimates of percent correct were fitted with a logit function of the form

$$p(x) = \frac{1}{n} + \left(1 - \frac{1}{n}\right)\frac{1}{1 + e^{-(x-\mu)/k}}, \tag{A1}$$

where n is the number of listening intervals (in this case 3), x is the signal level in decibels, μ is the mean of the function, and k is the slope. These fits were quite accurate, with a median of 96.6% of the variance accounted for. Slopes for all observers appear in Table III.

In general, integration is inversely related to the steepness of the psychometric function, with little integration in cases of steep psychometric functions. For the logit fitted here, steep functions are reflected by small values of k. Slopes are comparable for random noise for stimuli from the two experiments, with mean values near 2.6 in both cases. There are substantial individual differences, however, with estimates spanning 2.1–2.9. For comodulated maskers the mean values of k are larger (i.e., slopes are shallower) for stimuli from experiment 1 as compared to those from experiment 2. This trend is opposite from the predicted slope difference based on the integration reported above. This finding supports the conclusion that the negative values of spectral integration observed in experiment 1 may be affected by a change in cue quality with inclusion of multiple signal tones, such as a reduction in across-channel masking.

The finding of comparable or shallower slopes for comodulated as compared to random masker conditions is in contrast to the reports of Moore *et al.* (1990b). That study estimated function slope in a family of multiplied noise bands and reported steeper psychometric functions for coherently modulated than random bands.

ANSI (**1996**). *ANSI S3–1996, American National Standards Specification for Audiometers* (American National Standards Institute, New York).

Bacon, S. R., Grimault, N., and Lee, J. (**2002**). "Spectral integration in bands of modulated or unmodulated noise," J. Acoust. Soc. Am. **112**, 219–226.

Berg, B. G. (**1996**). "On the relation between comodulation masking release and temporal modulation transfer functions," J. Acoust. Soc. Am. **100**, 1013–1023.

Bos, C. E., and de Boer, E. (**1966**). "Masking and discrimination," J. Acoust. Soc. Am. **39**, 708–715.

Buus, S. (**1985**). "Release from masking caused by envelope fluctuations," J. Acoust. Soc. Am. **78**, 1958–1965.

Buus, S., Schorer, E., Florentine, M., and Zwicker, E. (**1986**). "Decision rules in detection of simple and complex tones," J. Acoust. Soc. Am. **80**, 1646–1657.

Carlyon, R. P., Buus, S., and Florentine, M. (**1989**). "Comodulation masking release for three types of modulator as a function of modulation rate," Hear. Res. **42**, 37–45.

Cohen, M. F., and Schubert, E. D. (**1987**). "Influence of place synchrony on detection of a sinusoid," J. Acoust. Soc. Am. **81**, 452–458.

Fantini, D. A., Moore, B. C. J., and Schooneveldt, G. P. (**1993**). "Comodulation masking release as a function of type of signal, gated or continuous masking, monaural or dichotic presentation of flanking bands, and center frequency," J. Acoust. Soc. Am. **93**, 2106–2115.

Festen, J. M. (**1993**). "Contributions of comodulation masking release and temporal resolution to the speech-reception threshold masked by an interfering voice," J. Acoust. Soc. Am. **94**, 1295–1300.

Grose, J. H., and Hall, J. W. (**1992**). "Comodulation masking release for speech stimuli," J. Acoust. Soc. Am. **91**, 1042–1050.

Grose, J. H., and Hall, J. W. (**1997**). "Multiband detection of energy fluctuations," J. Acoust. Soc. Am. **102**, 1088–1096.

Grose, J. H., Hall, J. W., Buss, E., and Hatch, D. R. (**2005**). "Detection of spectrally complex signals in comodulated maskers: Effect of temporal fringe," J. Acoust. Soc. Am. **118**, 3774–3782.

Hall, J. W., Grose, J. H., and Haggard, M. P. (**1988**). "Comodulation masking release for multicomponent signals," J. Acoust. Soc. Am. **83**, 677–686.

Hall, J. W., Haggard, M. P., and Fernandes, M. A. (**1984**). "Detection in noise by spectro-temporal pattern analysis," J. Acoust. Soc. Am. **76**, 50–56.

Hatch, D. R., Arne, B. C., and Hall, J. W. (**1995**). "Comodulation masking release (CMR): Effects of gating as a function of number of flanking bands and masker bandwidth," J. Acoust. Soc. Am. **97**, 3768–3774.

Kwon, B. J., and Turner, C. W. (**2001**). "Consonant identification under maskers with sinusoidal modulation: Masking release or modulation interference?," J. Acoust. Soc. Am. **110**, 1130–1140.

Langhans, A., and Kohlrausch, A. (**1992**). "Differences in auditory performance between monaural and dichotic conditions. I: Masked thresholds in frozen noise," J. Acoust. Soc. Am. **91**, 3456–3470.

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

Moore, B. C. J., Glasberg, B. R., and Schooneveldt, G. P. (**1990a**). "Across-channel masking and comodulation masking release," J. Acoust. Soc. Am. **87**, 1683–1694.

Moore, B. C. J., Hall, J. W., Grose, J. H., and Schooneveldt, G. P. (**1990b**). "Some factors affecting the magnitude of comodulation masking release," J. Acoust. Soc. Am. **88**, 1694–1702.

Nelken, I., Rotman, Y., and Bar Yosef, O. (**1999**). "Responses of auditory-cortex neurons to structural features of natural sounds," Nature (London) **397**, 154–157.

Piechowiak, T., Ewert, S. D., and Dau, T. (**2007**). "Modeling comodulation masking release using an equalization-cancellation mechanism," J. Acoust. Soc. Am. **121**, 2111–2126.

Schooneveldt, G. P., and Moore, B. C. (**1987**). "Comodulation masking release (CMR): Effects of signal frequency, flanking-band frequency, masker bandwidth, flanking-band level, and monotic versus dichotic presentation of the flanking band," J. Acoust. Soc. Am. **82**, 1944–1956.

Schooneveldt, G. P., and Moore, B. C. J. (**1989**). "Comodulation masking release for various monaural and binaural combinations of the signal, on-frequency, and flanking bands," J. Acoust. Soc. Am. **85**, 262–272.

van den Brink, W. A., Houtgast, T., and Smoorenburg, G. F. (**1992**). "Signal detection in temporally modulated and spectrally shaped maskers," J. Acoust. Soc. Am. **91**, 267–278.

Verhey, J. L., Dau, T., and Kollmeier, B. (**1999**). "Within-channel cues in comodulation masking release (CMR): Experiments and model predictions using a modulation-filterbank model," J. Acoust. Soc. Am. **106**, 2733–2745.

Lateralization of stimuli with independent fine-structure and envelope-based temporal disparities

Mathias Dietz,[a] Stephan D. Ewert, and Volker Hohmann
Medizinische Physik, Universität Oldenburg, 26111 Oldenburg, Germany

(Received 10 March 2008; revised 7 January 2009; accepted 7 January 2009)

Psychoacoustic experiments were conducted to investigate the role and interaction of fine-structure and envelope-based interaural temporal disparities. A computational model for the lateralization of binaural stimuli, motivated by recent physiological findings, is suggested and evaluated against the psychoacoustic data. The model is based on the independent extraction of the interaural phase difference (IPD) from the stimulus fine-structure and envelope. Sinusoidally amplitude-modulated 1-kHz tones were used in the experiments. The lateralization from either carrier (fine-structure) or modulator (envelope) IPD was matched with an interaural level difference, revealing a nearly linear dependence for both IPD types up to 135°, independent of the modulation frequency. However, if a carrier IPD was traded with an opposed modulator IPD to produce a centered sound image, a carrier IPD of 45° required the largest opposed modulator IPD. The data could be modeled assuming a population of binaural neurons with a physiological distribution of the best IPDs clustered around 45°–50°. The model was also used to predict the perceived lateralization of previously published data. Subject-dependent differences in the perceptual salience of fine-structure and envelope cues, also reported previously, could be modeled by individual weighting coefficients for the two cues.
 [DOI: 10.1121/1.3076045]

PACS number(s): 43.66.Pn, 43.66.Nm, 43.66.Ba, 43.64.Bt [RLF] Pages: 1622–1635

I. INTRODUCTION

Headphone experiments allow for an independent manipulation of interaural time differences (ITDs) and interaural level differences (ILDs) in a binaural stimulus. These ITD/ILD-manipulated stimuli typically result in a lateralization of the sound along an intracranial axis between both ears. The influence of an ILD is such that the intracranial image of a sound source is shifted toward the ear with the higher level (e.g., Halverson, 1922). For ILDs in excess of about 20 dB, the stimulus is perceived as shifted all the way toward one ear (e.g., Lindemann, 1986). Lateralization of stimuli with ILDs is possible over the whole audible frequency range.

Lateralization of sounds with ITDs, however, depends on the frequency composition of the stimulus. ITDs cause a lateralization of pure tones only in the frequency region below about 1.5 kHz (e.g., Kuhn, 1977), which may be related to the progressive loss of phase-locking in the auditory nerve (AN) as frequency is increased (Palmer and Russell, 1986; Weiss and Rose, 1988). It is generally assumed that fine-structure ITDs are not accessible if the envelope of the signal after peripheral filtering is essentially flat and the phase of the signal, i.e., its fine-structure, is not coded on the AN (e.g., Bernstein and Trahiotis, 1996). In addition to interaural temporal disparities in the fine-structure, the auditory system is able to exploit interaural temporal disparities in the stimulus envelope (e.g., Henning, 1974; Bernstein and Trahiotis, 1985b, 1994; van de Par and Kohlrausch, 1997; Buell *et al.*, 2008). For instance, high-frequency bandpass noise that is presented with an ITD is perceived at a lateral position even if the lowest-frequency component is well above 1.5 kHz (Bernstein and Trahiotis, 1994). Sinusoidally amplitude-modulated (SAM) tones have also been employed in the critical frequency region of 500 and 1000 Hz where fine-structure cues still play a role and envelope cues remain accessible (Bernstein and Trahiotis, 1985b). Either the entire waveform or only the stimulus envelope was delayed. The modulation frequency was chosen to be 50 Hz for the 500-Hz carrier and to be either 50 or 100 Hz for the 1-kHz carrier. The envelope shift alone revealed that the binaural system exploits temporal disparities in the envelope also in these low-frequency stimuli. Large differences between the results for the waveform shift and the envelope shift alone indicated a strong influence of the temporal disparities in the carrier. Furthermore, large individual differences between subjects were found.

In experiments 1 and 2 of the current paper, the individual contributions of carrier and modulator interaural phase differences[1] (IPDs) to the lateralization of SAM tones were investigated further by manipulating each of the two variables separately while holding the other fixed at 0°. Experiment 3 introduced opposing IPDs for carrier and modulator in order to determine their relative strengths, comparable to the classical time-intensity trading studies (Young and Carhart, 1974; Trahiotis and Kappauf, 1978; Ruotolo *et al.*, 1979). Additional studies investigated the effect of modulator shape (experiment 4) and signal intensity (experiment 5).

A further goal of the current study was to develop a model of binaural lateralization that can account for the experimental data. The first conceptual model for processing interaural temporal disparities is the so-called "Jeffress

[a]Electronic mail: mathias.dietz@uni-oldenburg.de

model" or "delay-line model" (Jeffress, 1948). The model consists of a sequence of coincidence detectors, which receive their input along two opposed chains of delay elements. In such a configuration, the ITD can be determined by the position along the delay-line at which the internal delay elements compensate for the external ITD in the stimulus and the coincidence neurons at this position find the best match. Even though delay-line models are able to account for a variety of interaural timing-related experimental data, the direct implementation of a delay-line model (e.g., Sayers and Cherry, 1957) fails for some more complex stimuli, which carry information in both fine-structure and envelope, particularly for SAM tones (Stern and Colburn, 1978). In order to account for more complex stimuli, the delay-line model has been extended in different ways (e.g., Colburn, 1977; Lindemann, 1986; Zerbs, 2000; Breebaart *et al.*, 2001a, 2001b, 2001c; Faller and Merimaa, 2004). Two influential extensions for modeling lateralization are the position-variable model (Stern and Colburn, 1978; Stern and Shear, 1996) and the weighted-image model (Stern *et al.*, 1988). The position-variable model determines the center of gravity over the complete delay-line with an optimized weighting as a function of delay. Due to the exponential decrease in the weighting for long delays, the delay-line can be restricted to maximum delays of 2–3 ms. The weighted-image model determines several maxima of the cross-correlation function over a range of auditory filters and applies higher weights to those maxima which align at a constant ITD in all auditory filters ("straightness") and those with a generally smaller ITD ("centrality"). Finally, the weighted sum over all maxima determines the lateralization. In order to derive this weighted sum, delay-lines need to be several times longer than the cycle duration of the stimulus.

A fundamental question about these models is whether or not the assumptions about the neuronal population as a function of "best delay" are in line with the populations found in physiological studies in mammals (e.g., Crow *et al.*, 1978, Fitzpatrick *et al.*, 1997; McAlpine *et al.*, 2001; Brand *et al.*, 2002; Hancock and Delgutte, 2004; Marquardt and McAlpine, 2007). The term best delay refers to the ITD at which a neuron has its highest response rate. Physiological findings suggest that in mammals, the best delay of a neuron is almost always within a half-cycle with respect to the frequency, where the neuron shows the highest rate response. In terms of IPD, this boundary is referred to as the π-limit (Marquardt and McAlpine, 2007). However, such a π-limit is not in line with the delay-lines of 2–3 ms in the models mentioned above. By definition, a π-limited system can only detect one maximum of the cross-correlation function. If the ITD is larger than a half-cycle of the respective center frequency, the detected maximum is not identical with the input ITD.

A crucial stimulus for analyzing the physiological existence of the π-limit is a noise stimulus centered at 500 Hz with a bandwidth of 400 Hz and an ITD of 1500 μs (Trahiotis and Stern, 1989). In this stimulus, the waveform is shifted by $3\pi/2$ with respect to the center frequency which is equal to $-\pi/2$ (with the minus indicating a shift to the side of the lag). Since only $-\pi/2$ would be in the range of the π-limit, $[-\pi;\pi]$, the highest response should switch from one hemisphere to the other when the ITD is increased from 500 to 1500 μs. This hypothesis was confirmed by Thompson *et al.* (2006) in a functional magnetic resonance imaging study of the human inferior colliculus (IC). They showed that the ipsilateral activity in response to the leading ear dominated in the 1500-μs condition, in contrast to a 500-μs delay where the maximum activity of most neurons was on the contralateral side. Nevertheless, the 1500-μs stimulus was still lateralized toward the "correct" leading side by all subjects (Trahiotis and Stern, 1989). Thompson *et al.* (2006) showed that the higher neural activity on the ipsilateral side is in line with π-limited delay-lines but not with delay-lines longer than 1.5 ms. They concluded that this limitation would invalidate the correlation-based models since their predictions cannot hold up with the reduced length of delay-lines. The ambiguity of $+3\pi/2$ and $-\pi/2$ would force the π-limited delay-line models to predict a reversed lateralization toward the lagging ear. In psychoacoustic experiments, however, this reversed lateralization is only observed for pure tones and narrow-band noise with a bandwidth up to about 200 Hz (Trahiotis and Stern, 1989). Stimuli with higher bandwidths are "correctly" lateralized to the side of the lead, even though π-limited delay-line models would still predict the lateralization to the side of the lag. The reason why delay-line models with a π-limit cannot explain the correct lateralization is that they can no longer exploit the information in the stimulus envelope. Since envelope frequencies are much lower than the center frequency of the respective auditory filter, peaks of the envelope correlations would lie far outside of the π-limit. However, as mentioned earlier, interaural envelope cues can be assessed by the auditory system, e.g., for lateralization at high-frequencies (Bernstein and Trahiotis, 1994) or for binaural masking level differences with transposed stimuli (van de Par and Kohlrausch, 1997; Bernstein and Trahiotis, 2007).

A modeling approach that seems suitable for assessing fine-structure and envelope information, including the assumption of a π-limit, is the recent IPD model of Dietz *et al.* (2008). In this model, fine-structure cues and envelope cues are separated into different channels. This separation may also help in overcoming the constraints that existing hard-wired models have with subject and level dependence. However, the IPD model has only been applied so far to analyze the temporal resolution of the binaural system for detecting broadband binaural beats (Siveke *et al.*, 2007, 2008), binaural masking level differences of pure tones (Hirsh, 1948), and transposed stimuli (van de Par and Kohlrausch, 1997). For these tasks, a quantitative estimation of the perceived lateralization was unimportant and was therefore not implemented in the model.

In Sec. II, an implementation of a π-limited lateralization model is developed. Thereafter psychoacoustic experiments are presented where temporal disparities of the fine-structure and the envelope cues were varied independently. In Sec. IV, the experimental results are compared to predictions of the lateralization model.

II. MODEL STRUCTURE

The current lateralization model is based on the IPD model as suggested in Dietz *et al.* (2008). In this section a more elaborate lateralization stage is introduced after a brief description of the monaural preprocessing stages.

- The middle-ear transfer characteristic was approximated by a 500-Hz to 2-kHz first-order bandpass filter according to Puria *et al.* (1997).
- Auditory bandpass filtering on the basilar membrane was modeled with a linear, fourth-order all-pole gammatone filterbank (Patterson *et al.*, 1987; Hohmann, 2002). Cochlear compression was modeled by an instantaneous compression with a power of 0.4 (e.g., Ruggero and Rich, 1991; Oxenham and Moore, 1994; Ewert and Dau, 2000) after bandpass filtering.
- The mechano-electrical transduction process in the inner hair cells was accounted for by half-wave rectification with a successive 770-Hz fifth-order lowpass filter as used in Breebaart *et al.* (2001a).
- Uncorrelated Gaussian noise was added to all auditory bands after hair cell transformation, in order to establish a finite hearing threshold and to mimic the loss of fine-structure phase-locking for frequencies above the cutoff frequency of the filter. The noise had the same rms-value as a 0 dB sound pressure level (SPL), 1-kHz pure-tone after half-wave rectification and prior to lowpass filtering. For the current study, however, the effect of the noise is negligible since all stimuli were presented well above threshold.

In Fig. 1, the preprocessing is depicted as step (A) "monaural preprocessing." Binaural processing starts with a separation of the monaurally preprocessed signals by a "fine-structure" and a "modulation" bandpass filter (step B). The fine-structure filter has the same center frequency as the respective peripheral auditory filter. The center frequency of the modulation (or envelope) channel was set to the frequency of the strongest envelope fluctuations. For the SAM tones used here, this is the modulation frequency of the amplitude modulation (25, 50, or 100 Hz). The use of one adjustable filter is a simplification of a modulation filterbank (Dau *et al.*, 1997; Ewert and Dau, 2000). Both bandpass filters were realized as complex-valued all-pole gammatone filters (see Hohmann, 2002 for implementation details). The order of the filters was set to 2 and their equivalent rectangular bandwidth to half of the respective center frequency (i.e., $Q=2$).

Each filter is described by the two parameters *cf* and *type*. *cf* is the center frequency of its respective peripheral auditory filter and type is either fine-structure or modulation. Therefore, the filter outputs of step (B) can be described as functions $g_{\text{left}}^{cf,type}(t)$ and $g_{\text{right}}^{cf,type}(t)$. The IPD can now be derived from the difference of the arguments $\phi_{\text{left}}^{cf,type}(t)$ and $\phi_{\text{right}}^{cf,type}(t)$ of two corresponding filters, as shown in Fig. 1 (step C). A temporal smoothing of the IPD was not assumed, following the findings of Dietz *et al.* (2008) and Siveke *et al.* (2008). Furthermore, only stationary IPDs were considered in the current study and thus model predictions do not depend on the temporal resolution of the internal IPD representation of the model.

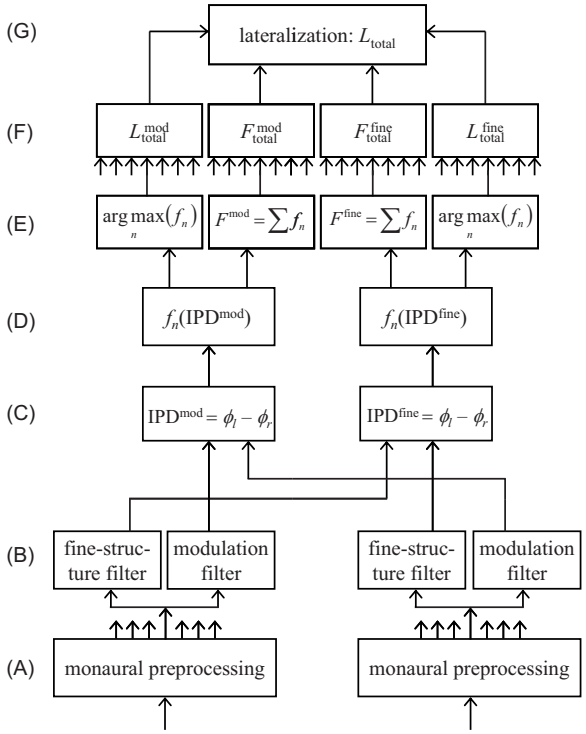


FIG. 1. Sketch of the binaural processing stages of the lateralization model. After peripheral preprocessing (output of step A), the signals are spectrally limited by two bandpass filters (step B). One filter extracts the dominant low-frequency modulation and the other extracts the fine-structure so that the respective phases can be determined. The IPDs are calculated (C) in order to determine the neuronal response (D). In the two outer blocks of step E, the best IPDs of the neurons with the highest response rates are determined. Furthermore, the total response is determined in the two inner blocks. Note that the two hemispheres are counteracting since the response of neurons with negative best IPDs is subtracted. Finally, the model integrates over all peripheral filter bands (F) and determines the total lateralization by a weighted spatial interpolation between the individual lateralizations $L_{\text{total}}^{\text{fine}}$ and $L_{\text{total}}^{\text{mod}}$ (G). The weighting coefficients are proportional to the two total response rates $F_{\text{total}}^{\text{fine}}$ and $F_{\text{total}}^{\text{mod}}$.

The crucial point for the lateralization is the response function of primary binaural neurons and how their response is coded in later stages of the auditory processing. For a realistic representation of lateralization, a population of binaural neurons is assumed with a characteristic tuning of their rate response to a certain best IPD (step D). The rate response function of the neurons neglects spontaneous activity and is expressed by a half-cycle of a cosine squared function:

$$f_n(\text{IPD}) = \begin{cases} \cos^2(q(\text{IPD-BestIPD}_n)) \text{ for } |\text{IPD-BestIPD}_n| < \pi/(2q) \\ 0 \text{ for } |\text{IPD-BestIPD}_n| \geq \pi/(2q), \end{cases}$$

with BestIPD_n representing the IPD at which neuron n has the highest response rate. The parameter q determines the sharpness of the tuning around its best IPD.

The distribution of the best IPDs in terms of their frequency of occurrence in the neuron population was modeled according to data of McAlpine *et al.* (2001) as it is presented in Marquardt and McAlpine (2007). They recorded neurons

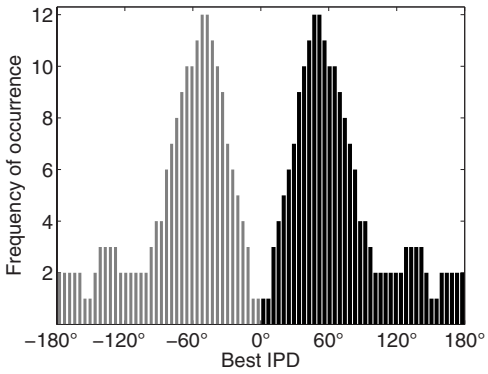


FIG. 2. Distribution of the best IPDs. The data are taken from the IC of guinea pigs (Marquardt and McAlpine, 2007). The height of each bar indicates how many neurons had their highest response (best IPD) within the respective interval. The width of each interval was set to 4.5°. An IPD in the range [0° 180°] leads to dominant response in the right IC (black bars) and an IPD [−180° 0°] to response in the left IC (gray bars).

in the IC of guinea pigs contralateral to the leading side of the wideband noise stimulus. The best-IPD distribution assumed in the present study is shown in Fig. 2. It was estimated from Fig. 1(c) in Marquardt and McAlpine (2007) by the following procedure: only BestIPD values between 0 and π (180°) were considered. The best IPDs were manually grouped in 40 bins with a width of 4.5°. 191 neurons were identified in the interval [0° 180°] and used for this study. The 36 neurons found outside this interval were discarded. Seven from 234 neurons could not be identified, due to a strong overlap of the respective data points in their plot. The highest density was found in the two neighboring bins [45° 49.5°] and [49.5° 54°] with 12 neurons each. For IPDs between $-\pi$ and 0, the distribution was mirrored and can be assigned to the IC neurons ipsilateral to the leading side of the stimulus. In terms of the model an IPD in the interval [0° 180°] means right side leading and left IC response (black bars in Fig. 2). An IPD in the interval [−180° 0°] means left side leading and right IC response (gray bars). It was further assumed that the IPDs in the envelope are processed in the same way as fine-structure IPDs, even though the discharge patterns of the respective neurons have some differences (e.g., Dreyer and Delgutte, 2006).

The response functions serve for two processing steps. First, the response functions are integrated in a way that expresses the response difference in the two hemispheres. Response from neurons of the right IC is subtracted from response of the left IC:

$$F^{cf,type} = \sum_{n|\mathrm{BestIPD}>0} f_n^{cf,type} - \sum_{n|\mathrm{BestIPD}<0} f_n^{cf,type}.$$

These quantities are displayed in Fig. 1 (middle blocks in step E).

Second, a place coding is employed in which the resulting lateralization estimate $L^{cf,type}$ for each specific band is determined by the best IPD of the neuron k with the highest response rate (outer blocks in step E):

$$k = (\arg\max_n(f_n^{cf,type})),$$

$$L^{cf,type} = \mathrm{BestIPD}_k.$$

Integration across different auditory bands is assumed as

$$L_{\mathrm{total}}^{type} = \frac{\sum_{cf} L^{cf,type} I^{cf,type}}{\sum_{cf} I^{\mathrm{cf},type}},$$

where $I^{cf,type}$ is the intensity of the respective fine-structure or envelope channel given in dB (step F). This intensity weighted integration assumes that channels with more energy contribute greater weight to the total lateralization than channels with less energy. The general motivation of such a weighting seems intuitively plausible (further motivation is provided in Sec. V). However, the specific implementation proportional to I in dB is just an assumption. Due to the middle-ear filter, the dB-SPL input is roughly transformed to dB-hearing level. For high-frequency fine-structure bands, it is also damped by the hair-cell lowpass filter; thus $I^{cf,type}$ represents the phase-locked intensity that is useful for binaural interaction. It was assumed that negative dB-SPL values[2] of $I^{cf,type}$ are below a "binaural perception threshold" and are therefore set to zero.

The across-frequency integration reduces the estimates from each filter pair to two estimates of lateralization $L_{\mathrm{total}}^{\mathrm{fine}}$ and $L_{\mathrm{total}}^{\mathrm{mod}}$ for the fine-structure and envelope, respectively. In the final step the overall lateralization L_{total} is determined by finding a weighted mean of $L_{\mathrm{total}}^{\mathrm{fine}}$ and $L_{\mathrm{total}}^{\mathrm{mod}}$ based on the response $F^{cf,type}$ in each band (step F in Fig. 1). For such a combination no neurophysiological evidence were available. The model assumes that the response in each band is integrated over frequencies (inner blocks in step F):

$$F_{\mathrm{total}}^{type} = p^{type} \sum_{cf} F^{cf,type} I^{cf,type}.$$

The parameter p^{type} represents the two scalar values p^{mod} and p^{fine} that can be adjusted for each subject. These parameters allow the model to account for the large individual differences found in experiments of SAM lateralization (Bernstein and Trahiotis, 1985a, 1985b) and in narrow-band noise lateralization (Trahiotis and Bernstein, 1986). The relative strength of $F_{\mathrm{total}}^{\mathrm{mod}}$ and $F_{\mathrm{total}}^{\mathrm{fine}}$ determines the weighting of $L_{\mathrm{total}}^{\mathrm{fine}}$ and $L_{\mathrm{total}}^{\mathrm{mod}}$. These two lateralizations are the left and the right extremes of L_{total}. Now a lateralization coefficient r, which is limited to a fixed interval, is created and the left and the right extremes are assigned to the lower and the upper limit of the interval:

$$r = \frac{F_{\mathrm{total}}^{\mathrm{fine}} + F_{\mathrm{total}}^{\mathrm{mod}}}{|F_{\mathrm{total}}^{\mathrm{fine}}| + |F_{\mathrm{total}}^{\mathrm{mod}}|}.$$

The denominator normalizes the difference between left and right variables to the interval [−1 1]. A value of $r=-1$ is assigned to the left of the two lateralizations:

$$L_{\mathrm{total}}(r=-1) =: \min\{L_{\mathrm{total}}^{\mathrm{fine}}, L_{\mathrm{total}}^{\mathrm{mod}}\}.$$

Accordingly $r=+1$ is assigned to the lateralization cue which is more on the right:

$$L_{\mathrm{total}}(r=+1) =: \max\{L_{\mathrm{total}}^{\mathrm{fine}}, L_{\mathrm{total}}^{\mathrm{mod}}\}.$$

Another important property of the coefficient occurs at $r=0$. In this case, the two opposing cues are traded and the "center of gravity" of the stimulus is predicted as being perceived from the midline:

$$L_{\text{total}}(r=0) =: 0.$$

Linear interpolations were assumed between these three defined values.

In Sec. III, psychoacoustic experiments are introduced, which serve for an evaluation of the model. In Sec. IV, the data of the experiments are presented and compared with predictions of the lateralization model.

III. METHODS

A. Subjects

Four normal-hearing listeners aged between 26 and 34 years participated in the experiments. All subjects had prior experience with binaural psychoacoustic measurements. Brief training was given to the participants until they were familiar with the stimuli and the task. Subject ID received a compensation for taking part in the experiment on an hourly basis.

B. Apparatus and stimuli

The subjects were seated in a double-walled, sound-attenuating booth and listened via Sennheiser HD 580 headphones. Signal generation and presentation during the experiments were computer controlled using the AFC software package for MATLAB, developed at the University of Oldenburg. The stimuli were digitally generated at a sampling rate of 48 kHz. The transfer function of the headphones was measured in an artificial ear (B&K 4153) and digitally equalized in order to obtain a flat (±1.5 dB) amplitude response between 0.1 and 20 kHz. The 500-ms stimuli were gated simultaneously with 20-ms raised-cosine ramps and were presented at a level of 65 dB SPL (if not otherwise stated) and with pause intervals of 300 ms.

SAM pure-tone stimuli were employed with a carrier frequency $f_c=1$ kHz:

$$s_l(t) = \sin(2\pi f_c t)(1 + m\,\sin(2\pi f_m t)),$$

$$s_r(t) = \sin(2\pi f_c t + \phi_c)(1 + m\,\sin(2\pi f_m t + \phi_m)),$$

where $s_l(t)$ and $s_r(t)$ represent the stimuli in the left and right ears, respectively. The phase lead of the right channel with respect to the left channel is denoted as ϕ_c for the carrier and as ϕ_m for the modulator. The modulation depth was always $m=1$. The stimulus generation method allows the interaural phase of the modulator [Fig. 3(a)] and the carrier [Fig. 3(b)] to be controlled independently.

In experiment 1, carrier phase shifts of 0°, 45°, 90°, and 135° were employed. Phase shifts were restricted to 135° since some subjects reported two intracranial positions for phase shifts $\phi_c > 135°$, one corresponding to ϕ_c and the other to $2\pi - \phi_c$ (e.g., Sayers, 1964). Modulation frequencies of 25, 50, and 100 Hz and an unmodulated condition ($f_m=0$ Hz) were employed. The phase difference of the amplitude modulation was kept constant at $\phi_m=0°$.

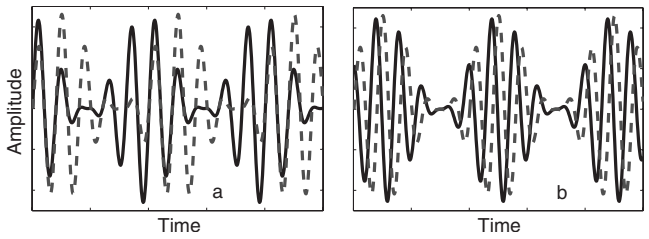


FIG. 3. Two dichotic SAM tones. Left channels in dashed gray, right channel in black. (a) Modulator shift alone. The carriers are still synchronized. (b) Carrier shift alone. The two channels have the same envelope.

In experiment 2, only the modulator phase was shifted by either 0°, 45°, 90°, or 135° and the carrier phase difference was set to zero. Modulation frequencies were again 25, 50, and 100 Hz.

In the third experiment, carrier phase differences were employed as in experiment 1 (0°, 45°, 90°, or 135°). Additionally, a variable modulator-phase difference was simultaneously applied to the stimulus. Experiment 4 was a repetition of experiment 3 with a two-tone complex (McFadden and Pasanen, 1976) instead of a SAM tone. The left channel of the stimulus was the sum of two pure tones, spectrally separated by the beat (or modulation) frequency. In the right channel, the same two tones were presented with modified phases in order to produce the desired interaural differences for the carrier and the envelope phase. The only difference from the stimulus of McFadden and Pasanen (1976) was, again, the possibility of controlling fine-structure and beat shift independently. In contrast to the "soft" sinusoidal envelope waveform of the SAM tones, the envelope waveform of the two-tone complex has steeper slopes in the envelope and shorter "modulation troughs." In experiment 5, a simple time delay of 750 μs was applied to the entire waveform of the SAM tones ($f_c=1$ kHz, $f_m=100$ Hz). Therefore, it was possible to investigate the relative role of fine-structure and envelope cues in a more natural matching experiment at different sound levels. These stimuli can be interpreted as the 1-kHz SAM counterparts of the 1500-μs delayed noise at 500 Hz from Trahiotis and Stern (1989), which was mentioned in the Introduction. The 750-μs delay translates to a $3\pi/2$ IPD for the carrier that can be interpreted as a $\pi/2$ IPD toward the side of the lag. The IPD of the amplitude modulation is only 27° and offers a cue toward the side of the lead. Since the sensitivity of modulation detection depends on level (e.g., Kohlrausch *et al.*, 2000), observing lateralization as a function of level should reveal the relative importance of both cues and help in testing modeling approaches. The stimuli were therefore presented at five different sound-pressure levels: 35, 45, 55, 65, and 75 dB SPL.

C. Procedure

For all experiments, a two-interval, two-alternative forced-choice paradigm was used. By pressing one of two left-right aligned buttons, the listener had to respond to the question: "Was the second stimulus left or right of the first stimulus?" In experiments 1, 2, and 5, the interaural phase shifts of either carrier (1) or modulator (2) were presented in the "target" interval according to the description in Sec.

III B. The perceived position was matched by an adaptively varied ILD ($=I_{\text{right}}/I_{\text{left}}$) in the reference interval. The target stimulus and the reference stimulus were randomly assigned to the two presentation intervals. When the listener's response indicated that the position of the reference was perceived to the left of the target, the ILD was increased for the next presentation and vice-versa. At the beginning of each experimental run, the ILD was randomly set to a value in the range ± 4 dB. The step size by which the ILD was initially changed was also 4 dB. After two reversals of the dependent variable, the step size was reduced to 2 dB and after two further reversals to 1 dB. The mean value of six reversals collected at the minimum step size was used as the resulting estimate. If the standard deviation of the estimate was larger than 2 dB, the run was discarded.

In experiments 3 and 4, the fixed carrier phase shift ϕ_c was combined with a modulator-phase shift ϕ_m as the dependent variable. The modulator IPD was traded against ϕ_c by adjusting it to the opposing direction until the resulting stimulus was perceived from the midline. Both presentation intervals contained the same stimulus with flipped left and right channels in random order. The measurement paradigm was the same as in experiments 1 and 2, except that the step sizes for the modulator shift were $4\pi/50$, $2\pi/50$, and $\pi/50$. Trading runs with standard deviations larger than $\pi/18$ (10°) were discarded.

In all experiments the conditions were grouped in blocks. Each block had a constant modulation frequency and contained the four different IPDs in randomized order. The four blocks (experiment 1) or three blocks (experiments 2 and 3) were randomly ordered. In total this made 16 conditions for experiment 1 and 12 conditions for experiments 2 and 3. After the last block, four repetitions of all blocks were measured in the same order. Experiment 4 was performed in the same way as experiment 3 with a modulation frequency of 100 Hz only, resulting in four conditions. Experiment 5 was conducted without randomization. The five conditions resulting from the different presentation levels were measured in the order of ascending level. Five repetitions of the block were measured. For each experiment data were collected on at least two different days. At least four runs had to be valid (standard deviation smaller than 2 dB or 10°) in each condition. In the few cases in which a subject had only three valid runs, an additional run was measured at the end of the experiment. The final result for each condition was the median of all valid runs.

IV. EXPERIMENTAL RESULTS AND MODEL PREDICTIONS

In the following, the experimental results and predictions of the model described in Sec. II are presented. For the model predictions, the individual coefficients p^{type} were chosen in units of "dB/deg." With this assumption, the model outcome is a predicted dB value according to an ILD that produces the same lateralization. The assumption implies a linear dependence between IPD and perceived lateral position in terms of an equivalent ILD as an approximation. In the matching experiments 1, 2, and 5, the target stimuli were fed into the model and the output values are directly comparable to the reference ILD in the psychoacoustic data. The output variation due to internal noise was less than 0.2 dB in most cases and was therefore ignored in the plots. The predictions for the trading experiments 3 and 4 were found by an incremental search of a sign change in the output variable ($r \approx 0$), with increments of the modulator IPD of either 1° or 2°. For all simulations, auditory filters in the frequency range 761–1296 Hz were used. The frequency spacing between adjacent filters of the filterbank was 0.2 times the equivalent rectangular bandwidth. This setting results in ten bands below 1 kHz, one band centered at 1 kHz and ten bands above 1 kHz. Any test with smaller spacing leads to the same results.

The parameter for the sharpness of the neuronal tuning was set to $q=2$. This was chosen as the broadest possible tuning of the response functions that still led to meaningful predictions of the data. It is also possible to choose $q>2$ which leads to sharper tuning of the IPD-sensitive neurons.

A. Experiment 1: Matching of carrier IPD

Figure 4 shows psychoacoustic data (left panel) and model predictions (right panel). The perceived lateral position of the sound was matched by an ILD for the different interaural carrier phases indicated on the x-axis. The experiment was conducted without amplitude modulation (∗) and with SAM at 25, 50, and 100 Hz (indicated by □, ◇, and ○, respectively). The interaural phase of the amplitude modulation was kept constant at $\phi_m=0$. In the model predictions (right panel), the AM had no effect; thus only a single set of data was plotted (indicated by the "+" symbols). In the left panel, each subpanel shows the median data for one subject. The error bars indicate the quartile boundaries. An increasing carrier IPD required an increasing ILD to be matched. For all subjects except ID, there is a more or less linear relationship between IPD and matched ILD with some flattening of the curve for increasing IPDs. Subject ID required a similar ILD to match the three carrier IPDs of 45°, 90°, and 135°. The matched ILD range varies across subjects. The results of subjects HK and MD differ by a factor of 2.5. The modulation frequency had no systematic effect on the results.

In case of the model predictions (right panel), four individually-adjusted versions of the model are shown that account for the main features of the data. The subject-dependent constant p^{fine} was individually adjusted to give the closest fit to the data. It was set to 12.2°dB/100° for subject HK and to 10.0, 7.5, and 4.8 dB/100° for subjects ID, SE, and MD, respectively. The increase in the matched ILD is mostly linear as a result of the simplified linear transformation between angle (IPD) and ILD. A slight deviation from linearity was observed since the internal noise causes statistical fluctuations in the IPD. For IPDs of 135°, the fluctuations cause a few instantaneous IPD values >180° that are interpreted as opposing cues and reduce the mean values. Saturation effects as in subject ID for all conditions and in subject SE for the 100-Hz condition cannot be modeled with such a linear approach. The model output does not depend on

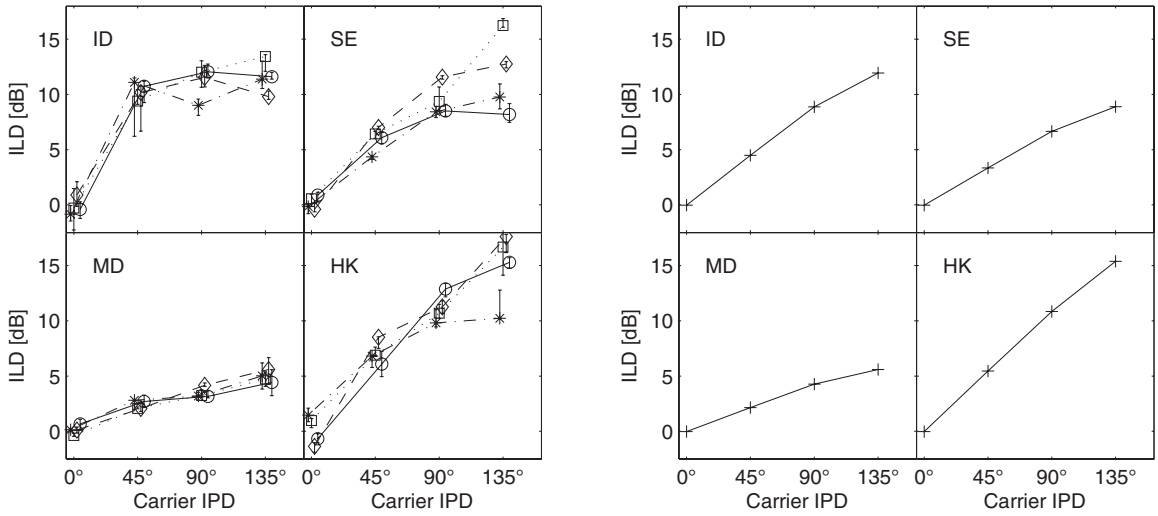


FIG. 4. Left panel: Psychoacoustic results of experiment 1. An IPD in the carrier of a 1-kHz SAM tone (indicated on the x-axis) was matched with an ILD of a 1-kHz SAM tone at different modulation frequencies. The symbols indicate the modulation frequencies of ∗: 0 Hz (pure tone), □: 25 Hz, ◇: 50 Hz, and ○: 100 Hz. Median values are shown and the error bars indicate the quartile boundaries. Right panel: Model predictions for experiment 1. The subject-dependent variable p^{fine} was fitted to the experimental data. No modulation frequency dependence was observed and only a single data set is plotted for all modulation frequency conditions, indicated by the "+" symbols.

the modulation frequency or on p^{mod} since the mean value of the modulator IPD is zero in all frequency bands.

B. Experiment 2: Matching of modulator IPD

Figure 5 shows psychoacoustic data and model predictions in the same format as in Fig. 4. In the left panel, the matched ILDs are shown as a function of modulator IPD. Some saturation was observed for subjects MD and SE at the highest modulator IPD of 135°. Subject ID showed a reasonably linear relationship between matched ILD and IPD, while HK showed the most pronounced flattening of the matched ILD curve with increasing modulator IPD. Again, there was no systematic effect of modulation frequency on the results when the data are plotted on an IPD-axis. If plotted over the modulator ITD, the gradient of the 100-Hz condition would be four times the gradient of the 25-Hz condition. Across-subject variability was smaller than in experiment 1. Model predictions are shown in the right panel of Fig. 5. The model predicts the main observations in the data. A common, subject-independent constant $p^{\text{mod}} = 7.2$ dB/100° was assumed. Nevertheless, the model predictions show minor dependence on subject and on the modulation frequency, most obvious in the 100-Hz condition (● symbols). The reason for this effect is a modulator-induced carrier phase shift in the off-frequency bands. Further analysis is provided in Sec. V. The difference between subjects is therefore caused by the different values of p^{fine} but the calibration of p^{mod} has no influence on the predictions of experiment 1.

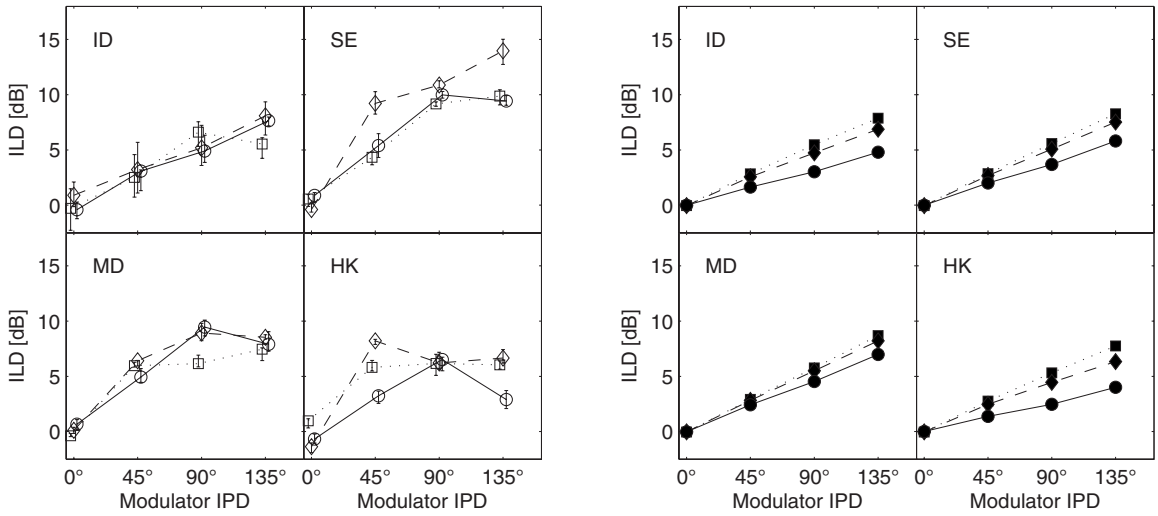


FIG. 5. Left panel: Psychoacoustic results of experiment 2. Matched ILD is shown as a function of the IPD in the modulator of a 1-kHz SAM tone. Different modulation frequencies of □: 25 Hz, ◇: 50 Hz, and ○: 100 Hz were used. Median values are shown and the error bars indicate the quartile boundaries. Right panel: Model predictions. A subject-independent variable $p^{\text{mod}} = 7.2$ dB/100° was fitted to the experimental data and was used in all subpanels. Note that the ITDs introduced on the modulator are in the range of 1–15 ms and therefore generally much larger than interaural differences occurring in free field listening.

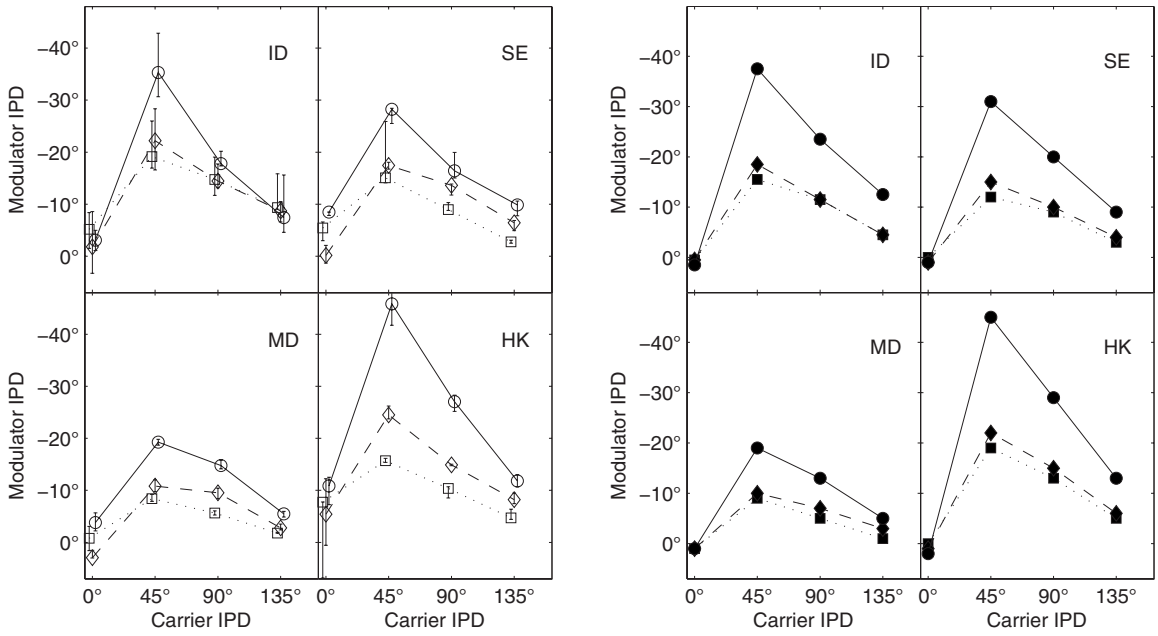


FIG. 6. Left panel: Psychoacoustic results of experiment 3. The modulator IPD required for trading an opposed IPD in the fine-structure is shown as a function of the carrier IPD. The stimulus was a 1-kHz SAM tone presented at modulation frequencies of 25 Hz (□), 50 Hz (◇), and 100 Hz (○). Median values are shown and the error bars indicate the quartile boundaries. Right panel: Trading functions predicted by the model. The parameters p^{fine} and p^{mod} were kept unchanged from experiments 1 and 2.

C. Experiment 3: Trading of carrier and modulator IPD

Figure 6 shows the results of the trading experiment with simultaneously applied opposing interaural carrier- and modulator-phase shifts. Again, the left panel shows the experimental data while the model predictions are shown in the right panel. A peaked pattern of the trading function is observed for all subjects. A carrier IPD of 45° always required the highest modulator IPD to be traded. The resulting modulator IPDs show significant deviations across subjects. Similar to experiment 1, the largest differences are observed between subjects MD and HK who required $\phi_m=-19.3°$ and $\phi_m=-45.8°$, respectively, to trade $\phi_c=45°$ in the condition with $f_m=100$ Hz. Unlike in experiments 1 and 2, a significant dependence of the data on the modulation frequency is observed. Larger modulation frequencies require higher opposing modulator IPDs for a midline percept. In order to define the peak position more precisely, subject MD performed an additional measurement with carrier IPDs of 30° and 60° for $f_m=100$ Hz. The median values were $\phi_m=-14.1°$ for the trading of a 30° carrier IPD and $\phi_m=-17.2°$ for the 60° carrier compared to $\phi_m=-19.3°$ for the 45° carrier IPD in Fig. 6. Therefore, at least for this subject, the peak position seems to be very close to the 45° carrier IPD.

The model predictions in the right panel of Fig. 6 are in good agreement with the data. The peak at 45° in the simulated data is related to the high values of $F_{\text{total}}^{\text{fine}}$ at this IPD which, in turn, is caused by the increased density of neurons with a best delay of 45° (Fig. 2). The high values of $F_{\text{total}}^{\text{fine}}$ lead to a strong "reliability" of the fine-structure cue in the model. Thus the 45° carrier IPD is harder to compensate for by an opposing envelope IPD than the 90° and 135° carrier IPDs, even though 90° and 135° produced a larger lateralization in experiment 1. After the p^{type} values had been adjusted to match the data in experiments 1 and 2, the model was kept unchanged for the predictions shown in Fig. 6. Taken together, the model accounts for the main effects in the data, i.e., the peak position, the subject dependence, and the effect of the modulation frequency.

D. Experiment 4: Trading of carrier and modulator IPD with a two-tone complex

Figure 7 shows psychoacoustic data (left panel) and model predictions (right panel) for trading the envelope of a two-tone complex against a carrier IPD, comparable to experiment 3. In contrast to experiment 3 where a sinusoidal envelope waveform was used, the envelope waveform of the two-tone complex is a full-wave rectified sinusoid with the same repetition period. In psychoacoustics, the faster attack

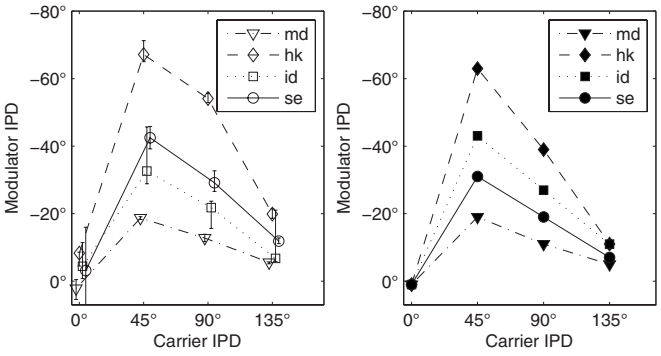


FIG. 7. Trading of a two-tone complex with a beat frequency of 100 Hz using the same procedure as in experiment 3 (Fig. 6). In the left panel, the psychoacoustic trading functions of all subjects are plotted. Median values for each subject are indicated by the different symbols and the error bars indicate the quartile boundaries. The predicted trading functions are shown in the right panel.

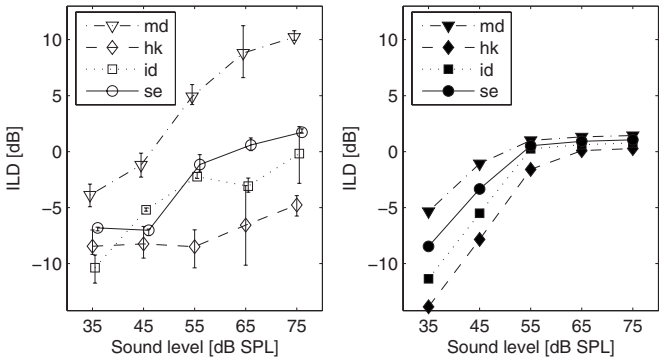


FIG. 8. Left panel: Psychoacoustic results of experiment 5. Matching the lateralization of a 1-kHz SAM tone with a modulation frequency of 100 Hz as a function of presentation level. The whole waveform was delayed by 750 μs. Median values are plotted and the error bars indicate the quartile boundaries. Right panel: Model predictions of the same experiment.

of the envelope waveform might mediate a different perceptual salience of the envelope cues (Bernstein and Trahiotis, 2007).

As in Fig. 6, a peaked pattern is obvious in the data. In the model, the different envelope shape leads to a different ratio of fine-structure versus envelope intensity and thus to a slight increase in the necessary modulator IPD compared to the 100-Hz condition of experiment 3. A similar increase was found in the psychoacoustic data. However, in contrast to experiment 3, the values of subject SE are higher than the values of subject ID. Such an inversion cannot be modeled with fixed p-values for these two subjects.

E. Experiment 5: Level dependence of lateralization

The level dependence on the lateralization of SAM tones (f_m=100 Hz, f_c=1 kHz) was investigated with a simple time delay of 750 μs.

Monaural modulation detection improves with stimulus level (Kohlrausch *et al.*, 2000; Millman and Bacon, 2008). If the monaurally detected amplitude modulations form the input to a binaural processor of envelope disparities, it is likely that the envelope cues are also more salient for higher-level stimuli. In the current model this results in a level-dependent lateralization for binaural stimuli with nonzero carrier and modulator IPD. The experimental results are shown in the left panel of Fig. 8 for the four subjects. All subjects showed a level dependence of the matched ILD. With increasing level, the ILD required to match the fixed temporal disparities changes in the positive direction. Again, a profound across-subject variation was observed. For the lowest level, all subjects matched the perceived lateral position of the stimulus by an ILD toward the side of lag (negative ILD values). For increasing levels, the ILDs increased toward zero for ID and SE, indicating a more centered perception at the midline. For subject MD, the matched ILD switched completely from negative to positive values, indicating that the stimulus was perceived to the side of lead for high levels. For subject HK, all matched ILDs were negative, equivalent to a perception toward the side of lag. However, this subject had some differences depending on the day of the measurement indicated by the generally larger error bars.

In the model predictions (right panel), the matched ILD changes from large negative values (equivalent to strong lateralization to the side of lag) to slightly positive ILDs (small lateralization to side of lead) as the stimulus level increases. For subject HK, the simulation predicts almost 0-dB matched ILD (centered midline perception) for the highest-level conditions. The pronounced flip to positive ILDs, as observed in the data for subject MD, cannot be explained by the model.

F. Accounting for published data

The current model was used to predict the perceived lateralization (characterized as matched ILD) of delayed noise, comparable to the stimuli described in the Introduction (Trahiotis and Stern, 1989). Since both frequency region and subjects differ from the model calibration, the parameter p^{fine} was used to fit the data and p^{mod} was kept unchanged. Trahiotis and Stern (1989) reported that all of the subjects lateralized the stimuli with 50- and 100-Hz bandwidths toward the side of the lag. The 200-Hz stimulus was lateralized differently, depending on the subject, and the stimulus with a bandwidth of 400 Hz was lateralized toward the side of the lead by all subjects. These results could be modeled with p^{fine}=1 dB/100°. Stimuli with bandwidths below 200 Hz were predicted to be lateralized toward the side of the lag and above 200 Hz toward the side of the lead.

Another general test for the model is the lateralization of very long ITDs (e.g., Blodgett *et al.*, 1956 or Mossop and Culling, 1998). For very long interaural delays, meaningful information can only be extracted from the stimulus envelope and it is therefore determined by the modulator IPD. Blodgett *et al.* (1956) reported that the lateralization of broadband noise can be determined correctly for delays up to 20 ms. This is very close to the longest ITDs of about 18 ms, which can be lateralized correctly by the current implementation of the model with a modulation filter at 25 Hz. Theoretically, modulation filters at lower frequencies could account for even longer delays.

V. DISCUSSION

A. Psychoacoustic results

Matching the lateralization of a dichotic stimulus with an ILD is a fundamental experiment in binaural psychoacoustics. Experiment 1, for instance, is comparable to the pure-tone experiments of Sayers (1964). In his study, a linear dependence between ILD and IPD was found up to about 130°, where an ambiguous region began in which subjects matched the ILD either to an angle α at the right or to 2π-α at the left. Pilot experiments of this study confirmed this finding and therefore carrier IPDs were limited to a maximum of 135°, where left-right confusion occurred very infrequently. The same limit was chosen for the modulator IPDs in experiment 2. For three of the four subjects, the relation between matched ILD and carrier IPD observed in experiment 1 is linear with some flattening of the matching curve toward increasing IPDs. This is in good agreement with previously published data on pure tones (e.g., Sayers, 1964). Interestingly, the results were unaffected by a simultaneous diotic amplitude modulation (25, 50, and 100 Hz) of the

1-kHz carrier. The findings suggest that the diotic modulation offers no additional "midline cue" that might reduce the extent of lateralization of the whole stimulus.

In experiment 2, the modulator IPD was matched for stimuli with an in-phase carrier, reversing the roles of modulator and carrier when compared to experiment 1. Again, an increasing matched ILD was found for increasing IPDs with some saturation effects at 135°. The data can in principle be compared to the measurements of Bernstein and Trahiotis (1985b). They also measured isolated modulator IPDs but only up to ϕ_m=36°. For this modulator phase shift, the matched ILD was similar to the data of experiment 2 at 45°, except for their subject AJ, where the matching ILD was more than twice as large.

The interaction of carrier and modulator IPD was investigated in experiment 3 where both (opposed) cues were traded to obtain a central sound image. A common trading experiment for binaural cues is the time-intensity trading, where temporal (ITD/IPD) cues are traded against an opposed level (ILD) cue to perceive a midline image (Young and Carhart, 1974; Young, 1976; Trahiotis and Kappauf, 1978; Ruotolo *et al.*, 1979). In time-intensity trading, the maximum ILD is required to trade IPDs around 90° for pure tones. If the IPD is further increased, the ILD required for trading decreases again and reaches zero for an IPD of 180°, where the IPD produces an ambiguous, semi-focused image (Young, 1976). In the current experiment 3, the pure-tone (carrier) IPD was traded by a modulator IPD instead of an ILD. A pronounced maximum in the trading function was found for a carrier IPD of 45°. This coincides with the highest neural density in the best-IPD distributions found by Marquardt and McAlpine (2007).

The purpose of experiment 4 (trading with a two-tone complex) was to investigate the influence of the envelope waveform. According to Kohlrausch *et al.* (2000), the modulation detection sensitivity is subject-dependent with respect to several parameters. The modulator cue in the two-tone complex might be less salient than in the SAM tone due to shorter modulation troughs in its envelope. On the other hand, the modulator cue could also be more salient because of the steeper slopes in the envelope. The data reveal that for three out of four subjects, the envelope of the two-tone complex offers indeed a less salient cue, which could be attributed to the short modulation troughs. Only subject SE shows no decrease in the salience of the envelope cue. Transposed tones, however, offer both steep slopes and long modulation troughs and therefore cause very salient envelope cues (Bernstein and Trahiotis, 2007).

Experiment 5 revealed a level dependence of the relative importance of fine-structure and modulator IPD cues. The results show that the lateralization moves more toward the side of the lead as the sound level is increased, indicating that envelope cues become more prominent with increasing level.

B. Modeling lateralization

The suggested lateralization model is based on the physiologically motivated assumption of a π-limit for the internal, auditory representation of IPDs in different auditory bands. The assumption of the π-limit is based on the neuron distributions found by Crow *et al.* (1978) and Marquardt and McAlpine (2007).

Thompson *et al.* (2006) stated that with existing lateralization models it is not possible to explain the lateralization of 400-Hz wide 1.5-ms delayed noise centered around 500 Hz with these physiological constraints in mind. A separation of fine-structure and envelope cues for modeling lateralization with the assumption of π-limited neurons in binaural processing is required. The IPD model of Dietz *et al.* (2008) fulfills these requirements and was used as the basis of the current lateralization model. In order to account for the data of experiment 3, a neuron population with a physiological distribution of best IPDs had to be taken into account. An interim approach using only two pairs of IPD-sensitive neurons was not successful. In the framework of the recent discussion about multi-channel place coding vs two-channel hemisphere coding (McAlpine and Grothe, 2003; Harper and McAlpine, 2004; Phillips, 2008), the current model approach can be seen as a hybrid. The lateralization for a single channel is performed by place coding, while the combination of the channels is determined by the overall response in each hemisphere.

In comparison to other complex models of binaural processing (e.g., Breebaart *et al.*, 2001a), the current approach is purely focused on interaural temporal disparities and cannot account for combined interaural level and temporal effects. Additionally, the process of cue combination (step E in Fig. 1) could be criticized as being both speculative and simplified for the purpose of trading and matching experiments. Thus, the current combination of cues, for example, does not account for an increasing lateralization in the case of congruent fine-structure and envelope IPDs. However, an increased perceived lateral position was reported with both fine-structure and envelope IPDs sharing the same direction (Bernstein and Trahiotis, 1985a). In order to account for such effects a different weighting of $L_{\text{total}}^{\text{fine}}$ and $L_{\text{total}}^{\text{mod}}$ would be necessary allowing for a positive summation. Independent of the weighting, a fusion to one scalar lateralization value L_{total} is always somewhat artificial as it cannot describe any spatial extent of sound images. Some listeners reported that they perceived two sound images in trading conditions and estimated a rather imaginary midpoint to perform the task. Thus, a more complete description of the spatial impression should also include a measure of image width or compactness (e.g., Hess, 2006). All necessary information to describe the spatial impression is provided by the primary binaural feature extraction of the model (e.g., Fig. 1 step D). It is therefore possible to model the spatial distribution by extending the "perception" stage of the model (Fig. 1 steps E–G). In Sec. IV F, the lateralization of delayed noise was simulated for different bandwidths. The model output depends on the bandwidth in the same way as the data of Trahiotis and Bernstein (1986).

However, the parameter p^{fine} had to be smaller for the delayed-noise simulations than for any of the four subjects in the SAM-tone simulations. A hypothesis for this difference is that noise shows a continuous spectrum of modulation fre-

quencies while SAM tones have only a discrete modulation frequency. The single modulation filter did not cover all modulation frequencies of the noise; thus the influence of the modulator IPD was underestimated. In the simulation of Sec. IV F this was compensated for by an underestimation of p^{fine}. In order to exploit the whole modulation frequency range, a modulation filterbank would be necessary. This model extension is possible but not required for the SAM tones of the current study. The influence of noise bandwidth is also predicted correctly by the weighted-image model (Stern *et al.*, 1988). However, the weighted-image model requires long delay-lines of several milliseconds.

Findings of large subject variability in complex lateralization experiments (e.g., Bernstein and Trahiotis, 1985a or Trahiotis and Bernstein, 1986) can easily be modeled by the parameter p^{fine}. Conventional hardwired delay-line models cannot account for these differences. In addition, the intensity weighting of the separated cues employs a level dependence as it was also qualitatively observed in experiment 5.

C. Physiological validity of the model

A possible interpretation of the two separated fine-structure and envelope channels in the model is that in mammals the fine-structure is processed in the medial superior olive (MSO) while the modulation IPDs are processed in the lateral superior olive (LSO). There is physiological evidence for a specialization of the LSO to level and envelope (amplitude modulation) disparities (e.g., Joris and Yin, 1995 or Joris, 1996) and also several indications that the MSO is the dominant detector for fine-structure IPDs (e.g., Brand *et al.*, 2002). In species with large heads and low-frequency hearing, like humans or dogs, the MSO is much more pronounced than in species with small heads (Grothe *et al.*, 2001). Psychoacoustic just noticeable difference (JND) measurements of 500-Hz pure-tone and 4-kHz transposed-tone ITDs and ILDs (Furukawa, 2008) might further support the two-channel hypothesis. In transposed tones the JNDs of (envelope) ITDs and ILDs were found to add up according to a common-channel hypothesis whereas pure-tone JNDs of (fine-structure) ITD and ILD were found to add up according to partially independent channels. These two findings support the view that the envelope ITD is processed together with the ILD, while the fine-structure ITD is processed independently. On the other hand, it cannot be ruled out that cue integration works differently at low- and high-frequencies.

Another physiologically plausible way to explain separate access to fine-structure and envelope disparities can be obtained is the so-called DIFCOR (fine-structure) and SUMCOR (envelope) metrics (Joris, 2003). These metrics are obtained by the difference (DIFCOR) or the sum (SUMCOR) of two different cross-correlation techniques.

The distribution of best delays underlying the model was taken from Marquardt and McAlpine (2007). Due to the primary feature extraction of phase differences, the neurons that had their best delays outside the π-limit had to be discarded. Most of the discarded neurons would have been considered as "trough-type neurons" having their well-defined minima within the π-limit and two almost equally high side peaks a half-cycle away from the minimum. This description already holds for trough-type neurons with best delays in the interval $[2\pi/3;\pi]$ (Marquardt and McAlpine, 2007). An open question is whether these neurons are really tuned to the shallow side peaks, which are separated by 2π and therefore represent the same IPD, or, more likely, to the sharp minimum which is usually close to zero IPD. The latter would support an even stricter π-limit and the respective neurons could be interpreted as excitatory-inhibitory elements (e.g., Durlach, 1963; Breebaart *et al.*, 2001a) parallel to the excitatory-excitatory majority. In the corresponding IPD region ($>135°$), the psychoacoustic measurements of lateralization also became instable and ambiguous and had to be excluded from the main measurements. For instance, the perception of $\text{IPD}^{\text{mod}}=\pi$ was described as diffuse and not lateralized (Thompson and Dau, 2008). These observations can be interpreted as caused by the side peaks of trough-type neurons and they are in line with identical activity on both ipsi- and contralateral sides in the model. However, neither the model predictions nor the psychoacoustic data are distinctive enough to explain the role of peak-type vs trough-type neurons in lateralization.

As mentioned in the Introduction, Thompson *et al.* (2006) measured the response in the IC and not in the superior olivary complex where the MSO and the LSO are located. The 1500-μs delayed 500-Hz noise caused a higher response in the ipsilateral IC while the model output produces dominant ipsilateral fine-structure response (MSO) and contralateral modulator response (LSO). These results fit well to the MSO/LSO model interpretation since both ipsilateral MSO and contralateral LSO project excitatory to the ipsilateral IC (Loftus *et al.*, 2004). Therefore, the model supports both conclusions of the recent study by Thompson *et al.* (2006): the π-limit entails a new model approach and the integrated response in contralateral vs ipsilateral IC is not a good measure for lateralization.

A critical physiological parameter is the sharpness q of the response function. The response function was modeled as wide as possible, but with a sharpness of $q=2$ it is still much narrower than the response functions recorded in the MSO (e.g., Brand *et al.*, 2002). However, Fitzpatrick *et al.* (1997) found that subsequent neurons show a sharper tuning. Their recording from the thalamus of rabbits showed response functions that fit very well to $\cos^2(q(\text{IPD-BestIPD}_n))$ with $q=2$. Another possible fitting is a $\cos^4$ function (Harper and McAlpine, 2004) which would lead to similar results.

Overall, the most obvious and most important difference of the current model compared to other models is the separation of fine-structure and envelope cues. Psychoacoustic and physiologic findings support this approach. However, several questions remain, in particular, about the role of trough-type units and the extent to which independent processing is realized in the auditory system.

D. Relation of experimental to modeled data

Reviewing the experiments, it appears surprising that the individual differences for fine-structure IPD matching are as large as a factor of 2.5. This makes it questionable to analyze

mean values across subjects. An explanation for the differences in p^{fine} is that the carrier frequency of 1 kHz is already in a region where phase-locking in humans decreases. Therefore, the differences in the influence of the fine-structure could also be modeled by a subject-dependent variation in the phase-locking filter in the model. Currently, a cutoff frequency of 770 Hz and a fifth-order Butterworth lowpass filter are employed for all subjects. Individual differences could be modeled by varying both the cutoff frequency and the filter order or shape.

For the envelope-based lateralization, it might have been a coincidence that the variation was much less; however, most notably there was no correlation between the fine-structure factor p^{fine} and envelope factor p^{mod} since some subjects have a larger p^{fine} and one subject has a larger p^{mod}. Therefore, it was impossible to *match* the fine-structure lateralization with the envelope-based lateralization (or vice-versa) for all subjects in the same way. On the other hand, it was possible to *trade* each fine-structure cue with an envelope cue. For instance, subject HK traded a 135° fine-structure shift with as little as 10° envelope shift to the opposite direction even though the fine-structure shift alone required an ILD of 15 dB to be matched. In the same subject, a 10° envelope IPD only required a 1 or 2 dB ILD to be matched.

While all these properties show a good accordance between model and psychoacoustic data, two points have to be discussed in some more detail.

The first point is the dependence on modulation frequency in experiment 3. For higher modulation frequencies, all subjects needed larger modulator IPDs to compensate for a given carrier IPD. One might argue that this outcome is related to the choice of an IPD-axis for the modulator and could be compensated for by converting the ordinate to an ITD-axis. However, there are several arguments against it. First, the 25-Hz trading function would be significantly higher than the two other trading functions if the modulator IPDs are converted to ITD. Since the current average factor between the 100-Hz and the 25-Hz values is 2, it will be 0.5 on an ITD-axis, which is as far away from unity than with the current IPD-axis. Second, there was no frequency dependence in experiment 2 in terms of the modulator IPD. Therefore, the IPD seems to be the natural parameter for experiment 2 and it would be hard to motivate a change in the axis parameter for experiment 3. Third, the simple model hypothesis of an intensity weighting already accounts for the frequency dependence of the trading functions and can be explained in an easy way: the finite bandwidth of the auditory filters causes an attenuation of the side-bands of the SAM tones. The higher the modulation frequency is, the larger is the separation of the side-bands and the larger is the effect of attenuation, resulting in a decreased effective modulation depth. Thus, the energy in the modulation filters decreases with increasing modulation frequency and larger modulator IPDs are required to trade a given carrier IPD.

While experiments 1–3 tested different modulation frequencies, all with the sinusoidal envelope waveform and level, these features were varied in experiments 4 and 5. The reason why the model output depends on envelope waveform and level is the intensity weighting introduced to account for the modulation frequency dependence. The data from both experiments revealed the same trends but some additional variations among subjects. The variations can be explained by the nonlinear and individual increase in modulation detection sensitivity with level (Kohlrausch *et al.*, 2000). In order to model the data more precisely, individual data on modulation detection sensitivity for each subject at each condition would be required.

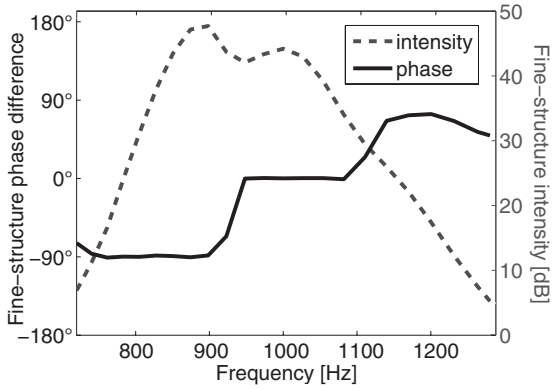


FIG. 9. The two critical parameters for deriving the intensity weighted mean IPD: intensity and fine-structure IPD as a function of center frequency. The input stimulus is a 1-kHz SAM tone with a 90° shift of the 100-Hz modulator and no carrier shift. The modulator shift induces a fine-structure shift in the off-frequency channels. The lowpass filter that limits the phase-locking reduces the weighting of the high-frequency channels. Therefore, the intensity weighted mean of the fine-structure IPD is negative.

The last point that needs to be discussed is the reported modulator-induced fine-structure shift in the off-frequency components. This effect, which can be seen in the simulation (Fig. 5, right panel), but not in the experimental data (Fig. 5, left panel), is caused by a modulator-induced fine-structure shift in the off-frequency channels. In Fig. 9, it can be seen that the 90° modulator shift (f_m=100 Hz) induces a fine-structure shift of −90° in the frequency region between 800 and 900 Hz. On the other hand it induces a positive shift between 1150 and 1250 Hz. However, the positive shift is a bit smaller and the intensity of the cue is much lower, due to the 770-Hz lowpass filter. Therefore, a net negative fine-structure cue remains, which is pulling the lateralization toward the center. The effect is stronger for higher modulation frequencies since these stimuli result in a lower intensity at the output of the modulation filter. Furthermore, the effect in the model is stronger if a stronger fine-structure weighting is applied (e.g., to model subject HK). A possible explanation why this phenomenon is not observed in the measurements is that the auditory system does not give so much weight to off-frequency channels or that the decrease in phase-locking is not as strong as is assumed in the model. Simulations at lower carrier frequencies or with less steep lowpass filters (for example, a first-order, 1-kHz lowpass as assumed in Dau *et al.*, 1996) would also reduce the effect in the model. Another explanation would be that the auditory system has learned to compensate for this natural asymmetry.

VI. CONCLUSIONS

The interaction of fine-structure and envelope interaural phase differences was investigated. Psychoacoustic data

were presented and a lateralization model based on IPD extraction was suggested. In the model, the separation of fine-structure and envelope information has proven to be a successful solution under the restrictive physiological side-condition of a π-limit for IPD extraction in the different auditory filters. The following conclusions can be drawn.

1. When carrier and modulator IPDs were set in opposition, for all subjects the largest modulator IPD required to offset a carrier IPD occurred when the carrier IPD was 45°. This result could be modeled assuming a population of binaural neurons with a physiological distribution of best IPDs clustered around 45°–50°.
2. Assuming a π-limited neuronal distribution of best IPDs, a separation of fine-structure and envelope cues appears to be required to successfully model the lateralization of binaural stimuli independent of the psychoacoustic measurements.
3. The assumption of separate processing for fine-structure and envelope cues allows the model to mimic the strong subject- and level-dependent differences as observed in the data.
4. By assuming an effectively analog processing for fine-structure and envelope cues the influence of modulator IPDs can be modeled correctly, however only for a fixed shape of the amplitude modulation.
5. The model can account for bandwidth dependencies since the intensity of envelope cues is influenced by the bandwidth.

ACKNOWLEDGMENTS

This study was supported by the DFG (SFB/TRR31 "The Active Auditory System") and the International Graduate School "Neurosensory Science, Systems and Applications." We would like to thank the members of the Medical Physics group and Birger Kollmeier for continuous support and fruitful discussions. We also thank Eric Thompson for proofreading the manuscript. D. Wesley Grantham, one anonymous reviewer, and Associate Editor Richard Freyman provided valuable input to improve the manuscript.

[1]Most properties of a narrow-band stimulus (e.g., tones, or wideband stimuli after peripheral filtering) can equally be described in terms of either ITD or IPD. ITD is the natural parameter to describe free field localization. However, in order to describe the model and the results of this study in the most convenient way and independent of the center frequency, the notation is mainly in terms of IPDs. Additionally the term "interaural temporal disparities" is used as a general expression that can be understood as either ITD or IPD.

[2]Even though the energy after the monaural hair-cell processing is always above or equal to 0-dB SPL (due to the additive noise), the energy in one of the two additional filters (fine-structure or envelope) may be below 0-dB SPL. This holds, e.g., for the envelope energy in pure-tone processing or the energy in the fine-structure filter for all frequencies above about 2 kHz.

Bernstein, L. R., and Trahiotis, C. (**1985a**). "Lateralization of low-frequency, complex waveforms: The use of envelope-based temporal disparities," J. Acoust. Soc. Am. **77**, 1868–1880.

Bernstein, L. R., and Trahiotis, C. (**1985b**). "Lateralization of sinusoidally amplitude-modulated tones: Effects of spectral locus and temporal variation," J. Acoust. Soc. Am. **78**, 514–523.

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. (**1996**). "The normalized correlation: Accounting for binaural detection across center frequency," J. Acoust. Soc. Am. **100**, 3774–3787.

Bernstein, L. R., and Trahiotis, C. (**2007**). "Why do transposed stimuli enhance binaural processing?: Interaural envelope correlation vs envelope normalized fourth moment," J. Acoust. Soc. Am. **121**, EL23–EL28.

Blodgett, H. C., Wilbanks, W. A., and Jeffress, L. A. (**1956**). "Effect of large interaural time differences upon the judgment of sidedness," J. Acoust. Soc. Am. **28**, 639–643.

Brand, A., Behrend, O., Marquardt, T., McAlpine, D., and Grothe, B. (**2002**). "Precise inhibition is essential for microsecond interaural time difference coding," Nature (London) **417**, 543–547.

Breebaart, J., van de Par, S., and Kohlrausch, A. (**2001a**). "Binaural processing model based on contralateral inhibition. I. Model structure," J. Acoust. Soc. Am. **110**, 1074–1088.

Breebaart, J., van de Par, S., and Kohlrausch, A. (**2001b**). "Binaural processing model based on contralateral inhibition. II. Dependence on spectral parameters," J. Acoust. Soc. Am. **110**, 1089–1104.

Breebaart, J., van de Par, S., and Kohlrausch, A. (**2001c**). "Binaural processing model based on contralateral inhibition. III. Dependence on temporal parameters," J. Acoust. Soc. Am. **110**, 1105–1117.

Buell, T. N., Griffin, S. J., and Bernstein, L. R. (**2008**). ""Listeners' sensitivity to "onset/offset" and "ongoing" interaural delays in high-frequency, sinusoidally amplitude-modulated tones," J. Acoust. Soc. Am. **123**, 279–294.

Colburn, H. S. (**1977**). "Theory of binaural interaction based on auditory-nerve data. II. Detection of tones in noise," J. Acoust. Soc. Am. **61**, 525–533.

Crow, G., Rupert, A. L., and Moushegian, G. (**1978**). "Phase locking in monaural and binaural medullary neurons: Implications for binaural phenomena," J. Acoust. Soc. Am. **64**, 493–501.

Dau, T., Kollmeier, B., and Kohlrausch, A. (**1997**). "Modeling auditory processing of amplitude modulations. I. Detection and masking with narrow-band carriers," J. Acoust. Soc. Am. **102**, 2892–2905.

Dau, T., Püschel, D., and Kohlrausch, A. (**1996**). "A quantitative model of the 'effective' signal processing in the auditory system: I. Model structure," J. Acoust. Soc. Am. **99**, 3615–3622.

Dietz, M., Ewert, S. D., Hohmann, V., and Kollmeier, B. (**2008**). "Coding of temporally fluctuating interaural timing disparities in a binaural processing model based on phase differences," Brain Res. **1220**, 234–245.

Dreyer, A., and Delgutte, B. (**2006**). "Phase locking of auditory-nerve fibers to the envelopes of high-frequency sounds: Implications for sound localization," J. Neurophysiol. **96**, 2327–2341.

Durlach, N. I. (**1963**). "Equalization and cancellation theory of binaural masking-level differences," J. Acoust. Soc. Am. **35**, 1206–1218.

Ewert, S. D., and Dau, T. (**2000**). "Characterizing frequency selectivity for envelope fluctuations," J. Acoust. Soc. Am. **108**, 1181–1196.

Faller, C., and Merimaa, J. (**2004**). "Source localization in complex listening situations: Selection of binaural cues based on interaural coherence," J. Acoust. Soc. Am. **116**, 3075–3089.

Fitzpatrick, D. C., Batra, R., Stanford, T. R., and Kuwada, S. (**1997**). "A neuronal population code for sound localization," Nature (London) **388**, 871–874.

Furukawa, S. (**2008**). "Detection of combined changes in interaural time and intensity differences: Segregated mechanisms in cue type and in operating frequency range?," J. Acoust. Soc. Am. **123**, 1602–1617.

Grothe, B., Covey, E., and Casseday, J. H. (**2001**). "Medial superior olive of the big brown bat: Neuronal responses to pure tones, amplitude modulations, and pulse trains," J. Neurophysiol. **86**, 2219–2230.

Halverson, H. (**1922**). "Binaural localization of phase and intensity," Am. J. Psychol. **33**, 178–212.

Hancock, K. E., and Delgutte, B. (**2004**). "A physiologically based model of interaural time difference discrimination," J. Neurosci. **24**, 7110–7117.

Harper, N. S., and McAlpine, D. (**2004**). "Optimal neural population coding of an auditory spatial cue," Nature (London) **430**, 682–686.

Henning, G. B. (**1974**). "Detectability of interaural delay in high-frequency complex waveforms," J. Acoust. Soc. Am. **55**, 84–90.

Hess, W. (**2006**). "time-variant binaural-activity characteristics as indicator of auditory spatial attributes," Ph.D. thesis, Bochum University, Germany.

Hirsh, I. (**1948**). "The influence of interaural phase on summation and inhibition," J. Acoust. Soc. Am. **20**, 536–544.

Hohmann, V. (**2002**). "Frequency analysis and synthesis using a gammatone

filterbank," Acta Acust. Acust. **88**, 433–442.
Jeffress, L. A. (**1948**). "A place theory of sound localization," J. Comp. Physiol. Psychol. **41**, 35–39.
Joris, P. X. (**1996**). "Envelope coding in the lateral superior olive. II. Characteristic delays and comparison with responses in the medial superior olive," J. Neurophysiol. **76**, 2137–2156.
Joris, P. X. (**2003**). "Interaural time sensitivity dominated by cochlea-induced envelope patterns," J. Neurosci. **23**, 6345–6350.
Joris, P. X., and Yin, T. C. (**1995**). "Envelope coding in the lateral superior olive. I. Sensitivity to interaural time differences," J. Neurophysiol. **73**, 1043–1062.
Kohlrausch, A., Fassel, R., and Dau, T. (**2000**). "The influence of carrier level and frequency on modulation and beat-detection thresholds for sinusoidal carriers," J. Acoust. Soc. Am. **108**, 723–734.
Kuhn, G. F. (**1977**). "Model for the interaural time differences in the azimuthal plane," J. Acoust. Soc. Am. **62**, 157–167.
Lindemann, W. (**1986**). "Extension of a binaural cross-correlation model by contralateral inhibition. I. Simulation of lateralization for stationary signals," J. Acoust. Soc. Am. **80**, 1608–1622.
Loftus, W. C., Bishop, D. C., Saint Marie, R. L., and Oliver, D. L. (**2004**). "Organization of binaural excitatory and inhibitory inputs to the inferior colliculus from the superior olive," J. Comp. Neurol. **472**, 330–344.
Marquardt, T., and McAlpine, D., **2007**, "A pi-limit for coding ITDs: Implications for binaural models," in *Hearing—From Sensory Processing to Perception*, edited by B. Kollmeier, G. Klump, V. Hohmann, U. Langemann, M. Mauermann, S. Uppenkamp, and J. Verhey (Springer, Berlin), pp. 312–318.
McAlpine, D., and Grothe, B. (**2003**). "Sound localization and delay lines—Do mammals fit the model?" Trends Neurosci. **26**, 347–350.
McAlpine, D., Jiang, D., and Palmer, A. R. (**2001**). "A neural code for low-frequency sound localization in mammals," Nat. Neurosci. **4**, 396–401.
McFadden, D., and Pasanen, E. G. (**1976**). "Lateralization at high frequencies based on interaural time differences," J. Acoust. Soc. Am. **59**, 634–639.
Millman, R. E., and Bacon, S. P. (**2008**). "The influence of spread of excitation on the detection of amplitude modulation imposed on sinusoidal carriers at high levels," J. Acoust. Soc. Am. **123**, 1008–1016.
Mossop, J. E., and Culling, J. (**1998**). "Lateralization of large interaural delays," J. Acoust. Soc. Am. **104**, 1574–1579.
Oxenham, A. J., and Moore, B. C. J. (**1994**). "Modeling the additivity of nonsimultaneous masking," Hear. Res. **80**, 105–118.
Palmer, A. R., and Russell, I. J. (**1986**). "Phase-locking in the cochlear nerve of the guinea pig and its relation to the receptor potential of the inner hair cells," Hear. Res. **24**, 1–15.
Patterson, R. D., Nimmo-Smith, I., Holdsworth, J., and Rice, P. (**1987**). "An efficient auditory filterbank based on the gammatone function," Paper presented at a Meeting of the IOC Speech Group on Auditory Modeling at RSRE, 14–15 December.
Phillips, D. P. (**2008**). "A perceptual architecture for sound lateralization in man," Hear. Res. **238**, 124–132.
Puria, S., Peake, W. T., and Rosowski, J. J. (**1997**). "Sound-pressure measurements in the cochlear vestibule of human-cadaver ears," J. Acoust. Soc. Am. **101**, 2754–2770.
Ruggero, M. A., and Rich, N. C. (**1991**). "Furosemide alters organ of Corti mechanics: Evidence for feedback of outer hair cells upon the basilar membrane," J. Neurosci. **11**, 1057–1067.
Ruotolo, B. R., Stern, R. M., and Colburn, H. S. (**1979**). "Discrimination of symmetric time-intensity traded binaural stimuli," J. Acoust. Soc. Am. **66**, 1733–1737.
Sayers, B. M. (**1964**). "Acoustic-image lateralization judgments with binaural tones," J. Acoust. Soc. Am. **36**, 923–926.
Sayers, B. M., and Cherry, E. C. (**1957**). "Mechanism of binaural fusion in the hearing of speech," J. Acoust. Soc. Am. **29**, 973–987.
Siveke, I., Ewert, S. D., Grothe, B., and Wiegrebe, L. (**2008**). "Psychophysical and physiological evidence for fast binaural processing," J. Neurosci. **28**, 2043–2052.
Siveke, I., Ewert, S. D., and Wiegrebe, L., **2007**, "Perceptual and physiological characteristics of binaural sluggishness," in *Hearing—From Sensory Processing to Perception*, edited by B. Kollmeier, G. Klump, V. Hohmann, U. Langemann, M. Mauermann, S. Uppenkamp, and J. Verhey (Springer, Berlin), pp. 354–357.
Stern, R. M., and Colburn, H. S. (**1978**). "Theory of binaural interaction based on auditory-nerve data. IV. A model for subjective lateral position," J. Acoust. Soc. Am. **64**, 127–140.
Stern, R. M., and Shear, G. D. (**1996**). "Lateralization and detection of low-frequency binaural stimuli: Effects of distribution of internal delay," J. Acoust. Soc. Am. **100**, 2278–2288.
Stern, R. M., Zeiberg, A. S., and Trahiotis, C. (**1988**). "Lateralization of binaural stimuli: A weighted-image model," J. Acoust. Soc. Am. **84**, 156–165.
Thompson, E. R., and Dau, T. (**2008**). "Binaural processing of modulated interaural level differences," J. Acoust. Soc. Am. **123**, 1017–1029.
Thompson, S. K., von Kriegstein, K., Deane-Pratt, A., Marquardt, T., Deichmann, R., Griffiths, T. D., and McAlpine, D. (**2006**). "Representation of interaural time delay in the human auditory midbrain," Nat. Neurosci. **9**, 1096–1098.
Trahiotis, C., and Bernstein, L. R. (**1986**). "Lateralization of bands of noise and sinusoidally amplitude-modulated tones: Effects of spectral locus and bandwidth," J. Acoust. Soc. Am. **79**, 1950–1957.
Trahiotis, C., and Kappauf, W. E. (**1978**). "Regression interpretation of differences in time-intensity trading ratios obtained in studies of laterality using the method of adjustment," J. Acoust. Soc. Am. **64**, 1041–1047.
Trahiotis, C., and Stern, R. M. (**1989**). "Lateralization of bands of noise: Effects of bandwidth and differences of interaural time and phase," J. Acoust. Soc. Am. **86**, 1285–1293.
van de Par, S., and Kohlrausch, A. (**1997**). "A new approach to comparing binaural masking level differences at low and high frequencies," J. Acoust. Soc. Am. **101**, 1671–1680.
Weiss, T. F., and Rose, C. (**1988**). "A comparison of synchronization filters in different auditory receptor organs," Hear. Res. **33**, 175–180.
Young, L. L. (**1976**). "Time-intensity trading functions for selected pure tones," J. Speech Hear. Res. **19**, 55–67.
Young, L. L. and Carhart, R. (**1974**). "Time-intensity trading functions for pure tones and a high-frequency AM signal," J. Acoust. Soc. Am. **56**, 605–609.
Zerbs, C. (**2000**). "Modeling the effective binaural signal processing in the auditory system," Ph.D. thesis, Oldenburg University, Germany.

Release from speech-on-speech masking in a front-and-back geometry

Neil L. Aaronson
Department of Physics and Astronomy, Michigan State University, Biomedical and Physical Sciences Building, East Lansing, Michigan 48824

Brad Rakerd
Department of Communicative Sciences and Disorders, Michigan State University, East Lansing, Michigan 48824

William M. Hartmann
Department of Physics and Astronomy, Michigan State University, Biomedical and Physical Sciences Building, East Lansing, Michigan 48824

(Received 13 October 2006; revised 6 November 2008; accepted 18 November 2008)

Informational masking of a target female talker by female distracters was measured with target and distracters presented from directly in front of the listener as a baseline condition. Next, it was found that if the distracters were also presented from directly in back of the listener, advanced or delayed by a few milliseconds with respect to the distracters in front, release from informational masking occurred. Release from informational masking was found for all delays within the Haas region of ± 50 ms, with peak release of about 3.5 dB. This peak occurred for a delay of ± 2 ms and it was shown to be the result of delay-and-add filtering. Release from energetic masking was also found, but only for delays of ± 0.5 ms or less.
 [DOI: 10.1121/1.3050276]

PACS number(s): 43.66.Qp, 43.66.Dc, 43.66.Pn [RLF] Pages: 1636–1648

I. INTRODUCTION

In a room full of talking people, it is generally possible to converse with just one. But the relative ease and success of the conversation will depend on a number of factors, including physical characteristics of the various talkers' voices and the content of their messages (Brungart *et al.*, 2001; Brungart and Simpson, 2007; Cherry, 1953; Yost, 1997). Spatial relationships among the talkers also play a role. It is easier to attend to a single talker among several who are speaking simultaneously if the target talker stands a distance away from the others. Hence, spatial separation between a target and any distracters can reduce the level of masking that is experienced by a listener (Bronkhorst, 2000; Duquesnoy, 1983; Peissig and Kollmeier, 1997). Hirsh (1950) proposed that if a listener perceived there to be a separation between target and distracters then that perception alone might be sufficient to trigger masking release. Freyman *et al.* (1999) took advantage of the precedence effect (Wallach *et al.*, 1949) to create such a perception, and, in fact, found measurable masking release as predicted. The specifics of their experiment were as follows.

First, in a baseline condition, a single loudspeaker directly in front of the listener presented the speech of multiple talkers. The listener's task was to attend to a single target talker and to ignore another interfering talker, the distracter. Next, in the precedence effect condition, a second loudspeaker off to the listener's right produced a copy of the distracter, shifted forward in time by 4 ms to lead the presentation of the distracter at the front speaker. Although the addition of this second loudspeaker increased the physical level of the distracting speech, it also shifted the perceived location of the distracting talker off to the listener's right side and well away from the target speaker's location, which remained in front. As a result, the intelligibility of the target speech was greatly improved. This experimental protocol has since been used in a number of studies of speech masking release. We will refer to it here as an added-delayed-distracter (ADD) experiment. Freyman *et al.* (1999) also found a limitation on masking release in the ADD experiment. Release was substantial for speech distracters but negligible for spectrally matched noise distracters. This suggested that the masking release seen with speech was largely a release from informational masking (IM). IM takes place centrally when there is a competition among co-occurring messages, as compared to energetic masking (EM), which takes place peripherally owing to spectral overlap among signals (Arbogast *et al.*, 2002; Brungart, 2001; Brungart *et al.*, 2001; Freyman *et al.*, 2001).

A time delay like the one just described in which the added distracters lead in time and invite perceptual separation via the precedence effect is defined here as a positive delay. Experiments have also been done with the distracters set to lag in time—a condition referred to here as negative delay. Freyman *et al.* (1999) found significant release from masking in an ADD experiment conducted with speech maskers and the negative delay time of −4 ms. In that instance, there was only a small difference in the perceived locations of the target speech and the distracter speech, but other spatial effects arose and these appear to have been able to support masking release as well. Specifically, a relatively diffuse auditory image of the distracters due to interaural

 0001-4966/2009/125(3)/1636/13/$25.00

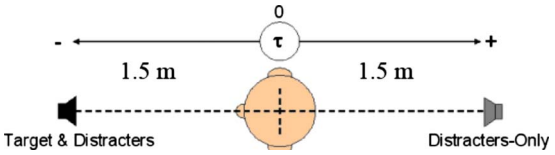


FIG. 1. (Color online) Arrangement of speakers relative to a listener for experiment 1. The delay between speakers is represented by τ, positive if the distracters from the back loudspeaker lead those from the front and negative if the distracters from the front loudspeaker lead those from the back.

disparities in the two-loudspeaker presentation was apparently important for differentiating the target talker from the distracters.

A number of ADD studies have since confirmed the presence of masking release for both positive and negative delays. Brungart *et al.* (2005) and Rakerd *et al.* (2006) extended the range of positive and negative delay times tested in ADD experiments out to values both much shorter and much longer than those that had been examined previously. One notable outcome of those studies was the finding that there can be release from EM with speech distracters, but only at very brief delay times. Another was the establishment of an upper time bound on IM release. The boundary is approximately 50 ms, and it appears to be set by the emergence of speech echoes (Haas, 1951).

In the original ADD study by Freyman *et al.* (1999) and in numerous studies since (e.g., Balakrishnan and Freyman, 2008; Freyman *et al.*, 2001, 2004, 2007; Rakerd *et al.*, 2006) leading and lagging signals have been presented from loudspeakers placed directly in front of a listener and off to the side, thereby modeling direct and reflected sounds distributed in the horizontal plane (HP). Direct sounds and their reflections can also arise in the median sagittal plane (MSP), and with that in mind the present ADD study was conducted with the front-back loudspeaker geometry shown in Fig. 1. An important distinction between the HP and the MSP concerns the role of binaural cues. Interaural differences in time and intensity are prominent and important for sound localization in the HP, but such differences are minimal or absent in the MSP owing to the equidistance of all points from a listener's two ears. In the MSP, listeners localize chiefly on the basis of spectral cues (Blauert, 1969, 1983), which has possible implications for performance in ADD experiments. The spectrum of a sound is complexly altered by the cooccurrence of a direct sound and its (positively or negatively) delayed reflection. For this reason, and because binaural cues are minimal in the MSP, masking release effects might be realized very differently in that plane than they are in the HP. On the other hand, previous experiments comparing sound localization and echo suppression for these two planes have found similar patterns of sensitivity (Litovsky *et al.*, 1997). A series of ADD experiments was run here to learn more about masking release in the MSP, and to compare the results with those of previous studies conducted in the HP.

II. EXPERIMENT 1: FRONT-BACK PRESENTATION WITH SPEECH DISTRACTERS

Experiment 1 measured release from speech-on-speech masking in an ADD experiment where the sound sources were directly in front and in back of the listener. Directly front and back are two locations that can be discriminated by listeners because of spectral differences (Blauert, 1969; Burger, 1958; Middlebrooks and Green, 1991). These locations are rarely confused by normal hearing listeners in broadband localization experiments (Wightman and Kistler, 1999). Methods used here were comparable to those used in a previous ADD study conducted with a HP geometry (Rakerd *et al.*, 2006).

A. Listeners

Four listeners, three male (listeners B, N, and S) and one female (listener K), participated in the experiment. Listeners K, N, and S were in their mid-twenties; and listener B was 52. All four listeners had normal hearing (pure tone thresholds at speech frequencies $\leqslant$15 dB hearing level at 0.5, 1, 2, and 4 kHz).

B. Anechoic room and experimental layout

Testing took place in an anechoic chamber, 3.0 m wide$\times$4.3 m long$\times$2.4 m high (IAC 107840). A listener was seated near the center of the chamber in a special chair, described below. One loudspeaker was placed directly in front of the listener, at ear height, 1.5 m from the center of the listener's head. Another loudspeaker was placed directly behind, also at ear height and 1.5 m from the head. This layout is shown in Fig. 1. It is referred to here as the *front-back geometry*.

C. Listener's chair

Rigorous measures were taken to prevent head motion and to ensure that each loudspeaker was equally distant from the listeners' ears. A wooden bite bar, 53 cm long, was attached to the chair, running parallel to the back of the chair. This bar was given a dark center line around its circumference at the center of its length and aligned approximately with the center of the chair. To ensure a constant alignment of the head, listeners were instructed to bite lightly on the bar and to maintain contact throughout the test facing the front loudspeaker. Prior to the test, listeners aligned the center line of their top incisors with the center line drawn on the bar using a small hand mirror.

D. Loudspeaker alignment

The loudspeaker azimuths were aligned individually, and were carefully centered on the listeners' midline, with the goal of minimizing interaural differences. The alignment procedure was as follows. Two small microphones were attached to the bite bar, one at each end. A sine tone was presented from the loudspeaker to be aligned. The outputs of the two microphones were observed simultaneously on a dual-channel oscilloscope outside the anechoic room, and the loudspeaker position was adjusted until the oscilloscope traces showed two sine tones with the same phase. A low frequency was used initially, and then successively higher frequencies were used for finer adjustments. After the final adjustment with a tone of 10 kHz, the estimated maximum

TABLE I. The full set of CRM sentences that were possible in this study. The listener is instructed to listen for Laker. The target talker always speaks this call sign.

{Talker}	"Ready	{call sign}	go to	{color}	{number}	now."
Target	"Ready	Laker	go to	Blue / Red / White / Green	One / Two / Three / Four	now."
Distracter1 / Distracter2	"Ready	Hopper / Ringo / Charlie	go to	Blue / Red / White / Green	One / Two / Three / Four	now".

error in angle was less than 0.5°. This corresponds to a maximum difference in arrival time of the sound between the left and right ears of less than 7 μs.

The loudspeaker distances (1.5 m from the center of the listener's head) were determined with a tape measure. This method had an estimated error of less than 1.5 cm, corresponding to a difference in arrival time between the front and back loudspeakers of less than 44 μs. In a delay-and-add filter, this delay corresponds to structure in the transfer function only above 11 kHz—well above the frequencies used in this experiment. A brief test of the perceptual effects of a small misalignment of this kind appears in Appendix A. This test indicated that the alignment procedure successfully minimized binaural difference cues.

E. Stimuli

The stimuli used for both targets and distracters were sentences taken from the coordinate response measure (CRM) corpus (Bolia *et al.*, 2000). For this experiment, each stimulus consisted of three female voices (the target and two distracters) issuing commands that followed the format: "Ready {call sign}, go to {color} {number} now." A chart of call signs, colors, and numbers allowed in these experiments is given in Table I. The target talker always used the call sign "Laker."

The voices of the three talkers were randomly chosen from among the four female voices available in the CRM. In any given stimulus, no two talkers shared any of the attributes of call sign, color, number, or individual female voice. With four colors and four numbers, the chance of guessing correctly becomes 1 in 16, or approximately 6%.

F. The task

Listeners were instructed to listen for the call sign Laker on each trial and to determine the color/number combination in the associated sentence. A stimulus with the Laker call sign was presented from the front on every trial. A liquid crystal display was mounted below the front loudspeaker. To help minimize any head motion, the listeners used a wireless gyroscopic mouse to control a pointer on the display without the need for a mouse pad or other surface. The listeners responded to each stimulus by clicking on the appropriately numbered button within the field of the appropriate color on the display. A response was considered correct only if the selected number and the color were both correct.

A single run consisted of five practice trials without feedback followed by 30 test trials. The listeners went through three runs for each experimental condition, with the order of runs randomized differently for each subject.

G. Front-only baselines

In a condition referred to here as the *front-only* (FO) *baseline*, the target and the distracters were presented exclusively from the front loudspeaker, with the level of each talker's speech fixed at 65 dB SPL. Thus, the level of the target was 0 dB relative to the level of each distracter. For a second reference point, the FO baseline experiment was repeated with the level of the target talker raised to +4 dB relative to the FO condition. This +4 dB signal-to-noise (S/N) ratio reference (the FO+4 dB condition) provided a way to express performance changes using a decibel scale.

As this ADD experiment was designed to search for perceptual effects over a wide range of time delays, it was desired that all listeners start with similar baseline performance. In a pilot test, it was found that three of the four listeners performed the baseline test near 30% correct, which was well above chance (6%). Listener K also performed above chance, but less well than the other listeners (~15% correct). The difference in baseline performance was eliminated by increasing the target level by 2 dB for listener K. All baseline and ADD data for listener K below reflect this 2 dB increase.[1]

H. Front-back ADD experiment

For all front-back ADD tests, the target and the distracters were presented from the front loudspeaker as for the FO baseline tests, and in addition, the distracters were presented from the back loudspeaker using the same level as in front (65 dB). There was a delay, τ, between the front and back distracters. The delay was varied over a wide range across the different conditions of the experiment. The set of delays employed was $\tau=\{\pm 32, \pm 8, \pm 2, \pm 0.5, 0\}$ ms. Positive delays indicate that the back loudspeaker (distracters only) led the front. Negative delays indicate that it lagged. Zero delay corresponds to synchrony in the presentation of the distracters from the two loudspeakers.

I. Results and analysis

Figure 2 shows the results of the front-back experiment for each listener, with the average across listeners given in the bottom panel. For the individual subjects, percent correct scores, averaged over three runs of 30 trials each, are given as a function of the delay, τ. Error bars represent the standard deviation over runs. A diagonally hatched rectangular stripe near the bottom of each panel shows the subject's average

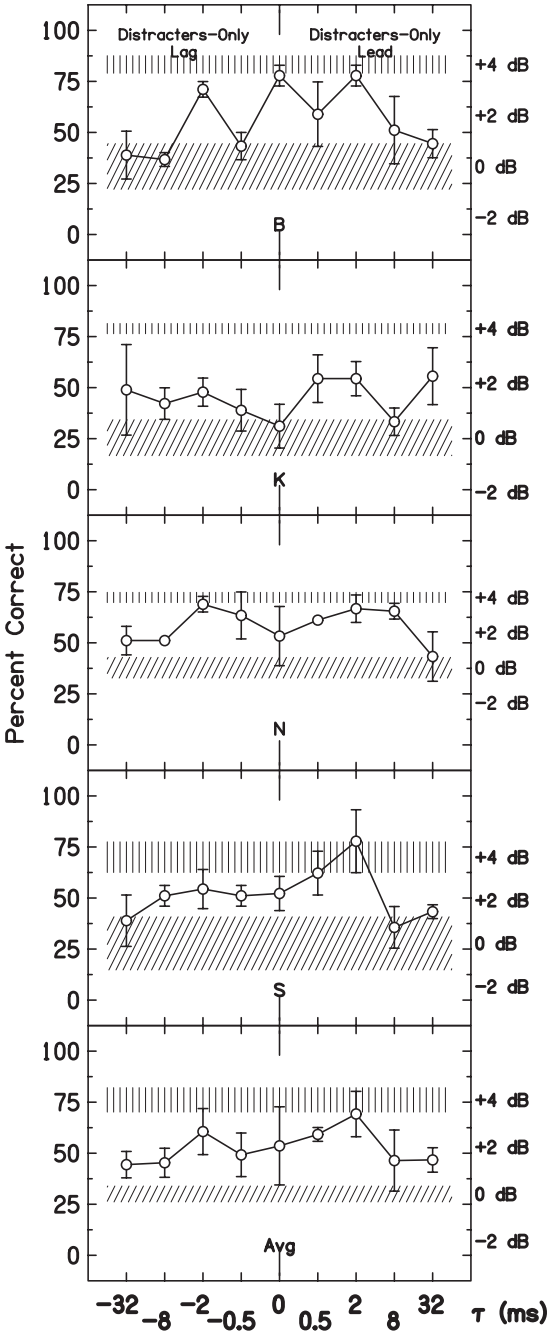


FIG. 2. Results of Experiment 1, the front-back experiment with speech distracters, for each of the four listeners and the average across listeners as a plot of percent correct vs delay, τ. The error bars are one standard deviation wide in each direction. The diagonally hatched rectangular stripe shows the result of the FO (single speaker) geometry, where the level of the target equals the level of each distracter voice. The width of the stripe spans the 95% confidence interval of average performance for the individual subject. For the average across listeners, the width represents the 95% confidence interval across the mean responses of individual listeners. The vertically hatched stripe is the same as the diagonally hatched except that the level of the target is boosted by 4 dB.

results for the FO baseline test [0 dB S/N ratio]. The width of the stripe represents a 95% confidence interval about the average baseline score. Another stripe (drawn with vertical hatch lines) is given for the FO test in which the S/N ratio was set to +4 dB (see right-hand axis of the figure). The bottom panel of Fig. 2 shows percent correct scores averaged across the four subjects. In that panel, error bars and the 95% confidence intervals for FO tests are based on the standard deviation over subjects.

Analysis of variance showed that front-back scores for all τ, $-32 \leqslant \tau \leqslant +32$ ms, were significantly greater than the FO baseline score ($p < 0.05$) except for $\tau = +8$ ms, where the score approached, but did not reach, significance ($p = 0.08$). These results provide strong evidence of release from masking in the front-back geometry for a wide range of delays comparable to those that might be encountered on an everyday basis in rooms.

For each listener, the measured difference in percent correct scores between the 0 and 4 dB FO reference conditions corresponded to a release of 4 dB (Sec. II G). These benchmark values were used to estimate the amount of unmasking in decibels on front-back tests by linear interpolation or extrapolation, as shown on the right-hand axes in Fig. 2. The delay times of $\tau = \{+2, -2\}$ ms showed the greatest release. Listeners displayed, on average, a release of 3.5 dB for $\tau = 2$ ms, and a release of about 2.5 dB for $\tau = -2$ ms. Measurable release of at least 1 dB extended out to $\tau = \pm 32$ ms.

The results of this experiment in the MSP agree with the results reported previously in the HP by Brungart *et al.* (2005) and Rakerd *et al.* (2006). Those experiments found masking release for delays as long as +32 and −32 ms, but not for 64 ms. Figure 2 shows that the release in the front-back ADD experiment strongly decreases at 32 ms. Therefore, it seems reasonable to conclude that about the same range of delays elicits a release from masking in both the HP and the MSP. However, at every value of the delay, the average release from masking is smaller in the MSP than in the HP. In the HP with two distracters, a maximum release of 11 dB was found (Rakerd *et al.*, 2006) compared to about 4 dB in the present experiment in the MSP. The 50 ms speech echo boundary found by Haas (1951), which Rakerd *et al.* held responsible for setting the upper bound on masking release in the HP, apparently sets it as well in the MSP.

It is likely that the release from masking seen in this experiment is aided by the ability of listeners to localize sounds in the MSP. Listeners use spectral structure in various frequency bands to localize sounds in the front and back (Blauert, 1969, 1983). Roffler and Butler (1968) showed that effective localization in the MSP is assisted by broadband stimuli with components above 8000 Hz, and fails for such stimuli without components above 2000 Hz. Because the CRM stimuli are low-pass filtered at 8000 Hz, energy in the 2000–8000 Hz range is present to contribute to front-back localization. The localization must employ the precedence effect, which is known to exist in the MSP (Litovsky *et al.*, 1997). A brief follow-up experiment, in which speech distracters were presented in the back loudspeaker only while the target remains in the front, appears in Appendix B and serves as a comparison for an entirely geometrical location change.

III. EXPERIMENT 2: FRONT-BACK PRESENTATION WITH SPEECH-SHAPED NOISE MASKERS

The masking of a speech signal by noise with similar spectral content is referred to as EM. This type of masking is

mainly attributed to physical interactions in the peripheral auditory channels (Kidd *et al.*, 1994). In the masking of speech on speech, EM may occur because of spectral overlap of the target signal and distracting speech, resulting in competition between the target and masker in the periphery of the auditory system. By contrast, IM does not require spectral overlap of signals (Arbogast *et al.*, 2002; Kidd *et al.*, 1994). IM in competing speech signals is the result of the difficulty a listener experiences in trying to distinguish a target speech message from distracting speech messages (Durlach *et al.*, 2003). Experiment 2 was conducted to determine the extent to which the speech-on-speech masking effects of Experiment 1 might be attributable to EM (and EM release).

Previous ADD studies examining masking release in the HP have found that release from EM occurs for very brief delays ($-0.5\ \text{ms} \leqslant \tau \leqslant +0.5\ \text{ms}$), and that EM release becomes negligible for longer delays (Brungart *et al.*, 2005; Freyman *et al.*, 2001, 1999; Rakerd *et al.*, 2006). In the HP, Rakerd *et al.* (2006) found an average release from EM of 2 dB in the very brief delay range. Brungart *et al.* (2005) also noted a significant release in similar conditions. The present study examined the time delay dependence of EM release in the MSP, specifically for the front-back geometry. Following previous studies, maskers used to test for EM release were speech-spectrum-shaped noises, with spectra matched to those of speech maskers.

A. Methods

Experiment 2 was an ADD experiment, identical to Experiment 1 with one exception. For Experiment 2, the distracters were continuous speech-spectrum-shaped noises that should be comparable to speech maskers in their ability to exert EM, but should exert no IM at all. The subjects of Experiment 2 were the same four listeners who participated in Experiment 1. They completed all tests for that experiment before beginning this one.

All distracters from the CRM corpus (i.e., those voices that spoke call signs other than Laker) were modified to derive equivalent speech-shaped noise samples, forming a "noise corpus." To do this, a discrete Fourier transform was applied to each individual speech file in its entirety. Then, each complex spectral component was multiplied by $e^{\iota\phi}$, where ϕ was a random variable uniformly distributed on the range $[-\pi, \pi]$. Thus, the phases of the frequency spectrum were randomized while the amplitudes remained unchanged. An inverse Fourier transform converted the modified spectrum back to the time domain. All listeners agreed that the result sounded like a swarm of bees, at roughly the same pitch as the original voices. Figure 3 shows an amplitude spectrum, averaged over five randomly chosen sentences, for each of the four talkers in the CRM. Each spectrum has a strong peak near 200 Hz indicating the fundamental component.

Pilot testing in the FO condition showed that at a S/N ratio of −10 dB, baseline performance with noise maskers was similar to that seen with speech distracters in Experiment 1. FO baseline tests were therefore conducted for all listeners at a −10 dB S/N ratio. For a second reference point

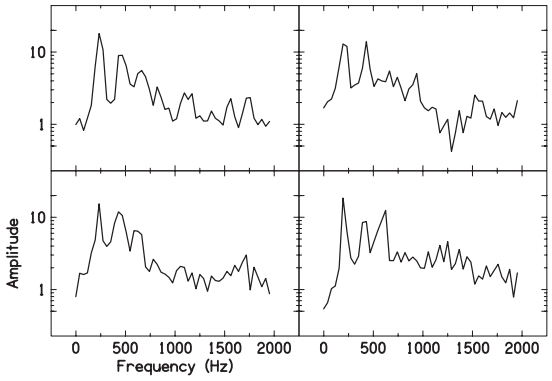


FIG. 3. Average amplitude spectra of voices. Each panel represents the average spectrum for one talker's voice, averaged over five utterances. The vertical axis has arbitrary amplitude units, the same for all four.

the speech level was increased to −6 dB S/N ratio. To test for release from EM in the front-back geometry, both the target speech and two noise maskers were presented from the front loudspeaker at −10 dB S/N ratio, and a delayed copy of the maskers was also presented from the rear loudspeaker. Altogether, there were nine test conditions, corresponding to different values of the delay, τ, the same nine that had been employed in Experiment 1.

B. Results and analysis

Figure 4 shows the results of Experiment 2. All listeners showed masking release (relative to baseline) at $\tau = \pm 0.5$ ms, and these eight points (four listeners and two delays) were all statistically significantly different from the baseline ($p < 0.05$). All listeners except for K also exhibited significant release when $\tau = 0$ ($p < 0.05$). Listeners N and B displayed a release of about 4 dB for $\tau = 0$. For listener S, the release was approximately 3 dB.

The results for Experiment 2, averaged over all four listeners, are shown in the bottom panel of Fig. 4. Statistically significant release from masking ($p < 0.05$) occurred for all three values of τ in the region $|\tau| \leqslant 0.5$ ms ($p < 0.05$) and for no other value outside that range. The average release from EM at $\tau = 0$ was about 2.5 dB. A release of 2.5 dB was also found for $\tau = 0.5$ ms, and a release of 1.5 dB was found for $\tau = -0.5$ ms. There was no evidence of masking release for longer delays. For $\tau = \pm 32$ ms and $\tau = \pm 8$ ms, the average performance was, in fact, significantly below the performance in the baseline condition ($p < 0.05$), presumably because the masking power in front-back trials was double that for FO trials. Overall, these results agree with the HP experiments by Brungart *et al.* (2005) and Rakerd *et al.* (2006), which also found significant EM release for very brief delays (−2, −0.5, 0, and 0.5 ms) but not for longer delays. For brief delays, the magnitude of the release found in the front-back geometry was similar to that found in the HP.

The noise masker results of Experiment 2 (Fig. 4) on EM can be compared with the speech masker results of Experiment 1 (Fig. 2). The following points are notable: (1) The range of delays for which the release appears in Experiment 2 (−0.5 to 0.5 ms) was far more limited than in Experiment 1. (2) No release occurred in Experiment 2 at $\tau = \pm 2$ ms,

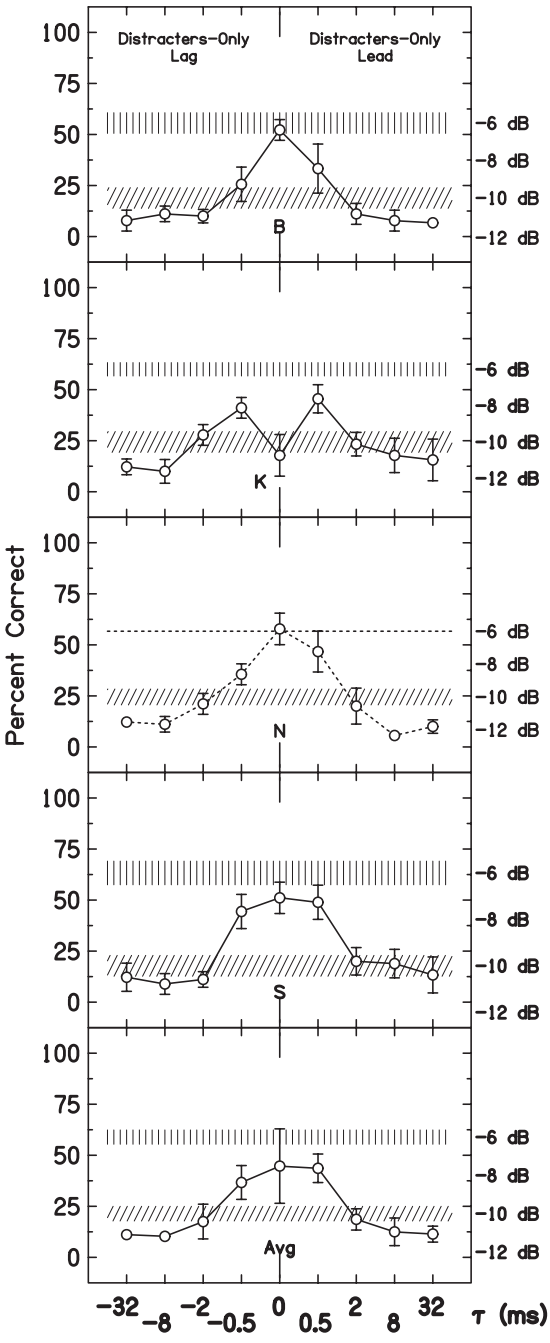


FIG. 4. Individual and average results of Experiment 2, the front-back experiment with noise distracters, similar in form to Fig. 2, as a plot of percent correct vs delay, τ. For listener N, who showed no variation within three runs in performance of the boosted FO condition, a dashed straight line represents the average percent correct.

where Experiment 1 showed the greatest release. (3) The release at $\tau=\{-0.5,0.5\}$ ms was statistically the same in Experiments 1 and 2 ($p>0.716$ in a two-sample t-test of zero difference).

We conjecture that the release seen in Experiment 2 at $\tau=0$ ms and the release seen at ± 0.5 ms occurred for different reasons. For $\tau=0$ ms, it seems likely that the release from EM is a localization effect, with the noise maskers perceived to be located separately from the target due to summing localization. The release from EM at $\tau=\{-0.5,0.5\}$ ms is more likely due to delay-and-add (comb) filtering (Hartmann, 1998) because the "window" of summing localization is beginning to close at these delay times (Blauert, 1983). As a consequence of delay-and-add filtering, the masking noises had a broad spectral valley from 0 to 2 kHz, centered at 1 kHz. In an informal ADD experiment where listeners adjusted the delay of noise maskers, a delay near 0.5 ms was found to be particularly effective in unmasking target female speech. The conjecture that different mechanisms lead to release for the different delays might explain why all listeners exhibited a release at $\tau=\{-0.5,0.5\}$ ms, but, unlike the other listeners, listener K failed to show a release at $\tau=0$ ms. According to the conjecture, all listeners were able to utilize the spectral mechanism that leads to a release at $\tau=\pm 0.5$ ms, but listener K failed to make use of the localization mechanism that other listeners used to achieve a release from masking at $\tau=0$ ms. Implications of EM results are further addressed in Sec. VI B.

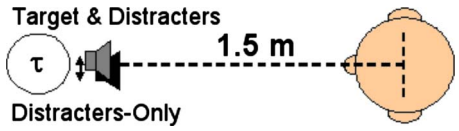


FIG. 5. (Color online) Arrangement of speakers relative to a listener for Experiment 3. The delay between speakers is represented by τ.

IV. EXPERIMENT 3: FRONT-FRONT PRESENTATION WITH SPEECH DISTRACTERS

Interaural differences were minimized in Experiment 1, which means that the masking release for speech distracters found there was chiefly due to spectral effects. These spectral effects may have had either or both of two origins. One source was the spatial nature of the front-back layout itself. The head-related transfer functions for sources in front and in back are quite different (Blauert, 1983), and listeners almost certainly gained some localization information from their head-related transfer functions (HRTFs). This spectral localization information in turn may have mediated speech masking release.

The other source of spectral differences was delay-and-add filtering. Experiment 3 was conducted to separate out the contributions of HRTFs and delay-and-add filtering. To do this, a new test layout was established, referred to here as "front-front" (see Fig. 5). The front-front layout deprived listeners of any spatial cues but retained spectral cues caused by delay-and-add filtering.

A. Methods

To create the front-front layout, the loudspeaker that had previously been in back was moved and placed on top of the front loudspeaker so as to make them collocated.[2] The loudspeaker alignment, method of data collection, time delays, and the target and masking stimuli for Experiment 3 were the same as in Experiment 1. The four listeners for Experiment 3 were the same as for Experiment 1 as well. The FO test was repeated here to provide a measurement of baseline performance contemporaneous with the front-front ADD test.

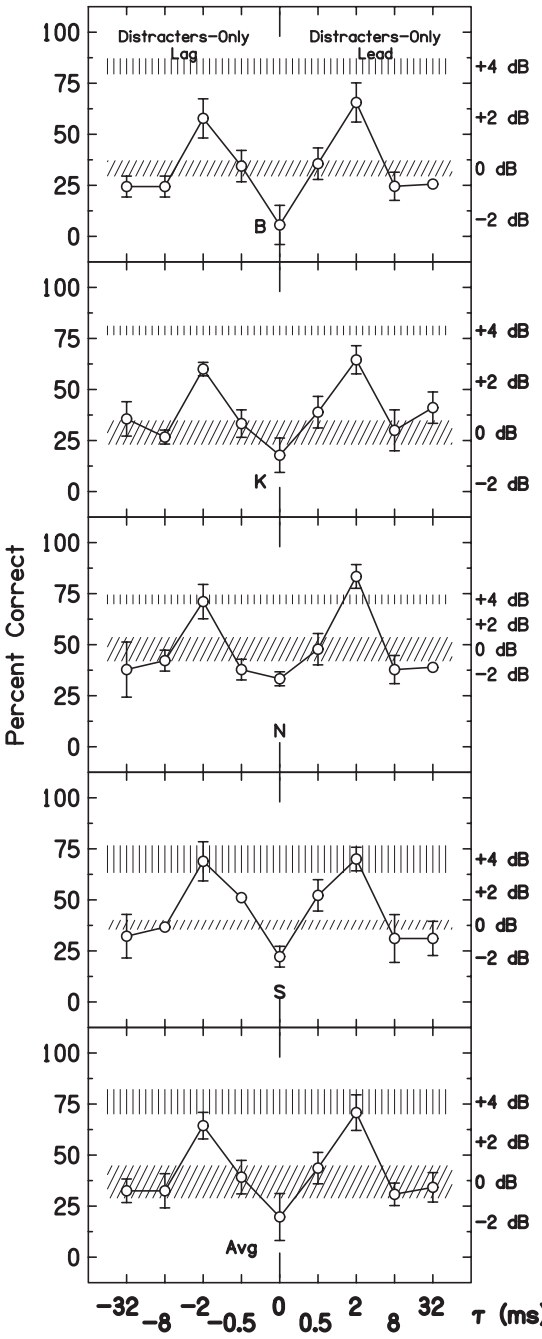


FIG. 6. Individual and average results of Experiment 3, the front-front experiment with speech distracters, similar in form to Fig. 2, as a plot of percent correct vs delay, τ.

B. Results and discussion

Figure 6 shows the results for the front-front ADD test. This figure is in every way parallel to Fig. 2 for the front-back ADD test.

The shape of the functions in Fig. 6 was remarkably similar across the four listeners. All listeners exhibited substantial release from masking for the delays of $\tau = \{+2, -2\}$ ms. This was most dramatically demonstrated by listener N, who showed nearly 6 dB of release for $\tau = 2$ ms. The smallest release seen at ± 2 ms was for listener K who showed a release of 2.5 dB at $\tau = -2$ ms. No listener showed any evidence of masking release for any other delay.

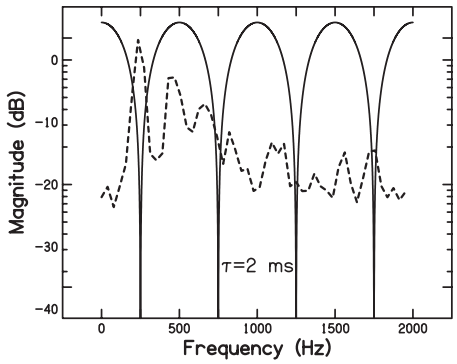


FIG. 7. The solid lines show the theoretical amplitude response of an ideal delay-and-add filter with delay $\tau = 2$ ms. Dips occur at 250 Hz and every additional 500 Hz thereafter. Peaks occur at every integer multiple of 500 Hz. Superimposed in dashed lines is the average amplitude spectrum of one of the four female CRM voices, which is taken from the upper left panel of Fig. 3.

Another point of agreement among listeners is that they all performed worse than baseline for a delay of $\tau = 0$. Here the distracter presentations perfectly coincided at the leading and lagging speakers and were therefore 6 dB more intense than at baseline.

The average results in the bottom panel of Fig. 6 clearly show significant release from masking at $\tau = \pm 2$ ms ($p < 0.05$) and no release at any other delay value. The peak in performance for $\tau = 2$ ms corresponded to a 3.5 dB release in masking, while the peak for $\tau = -2$ ms corresponded to a release of nearly 3.0 dB. These two decibel values are essentially the same, indicating symmetry about $\tau = 0$.

1. Delay-and-add filtering with speech maskers

The improved performance for $\tau = \pm 2$ ms in Experiment 3 is very likely due to delay-and-add filtering of the distracters in this experimental setup. The transfer function of a delay-and-add filter with a delay of 2 ms is shown in Fig. 7, superimposed on the average spectrum of a typical female voice from the CRM. For a delay of 2 ms, peaks occur at integer multiples of 500 Hz, and dips occur at 250 Hz and every additional 500 Hz thereafter.

The first dip in the delay-and-add spectrum for a delay of $\tau = \pm 2$ ms may be especially important. As shown in Fig. 7, this dip is close to the average fundamental frequency of female voices. Apparently, suppressing the energy in the fundamental component introduced timbre differences in the distracters that helped the listeners distinguish between the target talker and the distracters, leading to a release from masking of the target. The release seen at 2 ms in Experiment 3 (no spatial cues) was as large as the release at 2 ms in Experiment 1 (front-back spatial cues). It seems likely that the peaks at ± 2 ms seen in Experiment 1 were the result of the delay-and-add filtering effect, as made evident in Experiment 3, but spatial cues may also have played some role.

2. Delay-and-add filtering with noise maskers

In contrast to the results with speech maskers, no release was seen at $\tau = \pm 2$ ms with noise maskers in Experiment 2, though presumably delay-and-add filtering had a similar ef-

fect on the spectrum of the maskers in that experiment. Delay-and-add filtering of the noise maskers removes power from certain parts of the spectrum but adds power to other parts of the spectrum. Therefore, delay-and-add filtering would not be expected to lead to a release from EM unless the delay afforded some strategic advantage.

Unmasking the fundamental frequency of the target would not lead to better intelligibility of the target speech, since the relevant information in this speech task is not contained in the fundamental.

3. More on speech maskers

Support for the notion that a dip at the fundamental caused by delay-and-add filtering contributes to a release from masking by speech is found in the results of Brungart *et al.* (2005). In that study, a similar experiment to the present one was performed in a virtual auditory environment with a single *male* speech distracter and a target taken from the CRM corpus (this was referred to as the F-FF condition). Brungart *et al.* (2005) found a peak in release from masking for a delay of 4 ms, distinct from no release for 2 and 16 ms. This peak is similar to the peak seen in the present experiment for 2 ms, distinct from no release seen for 0.5 and 8 ms. A delay-and-add filter with a delay of 4 ms has a first dip at 125 Hz, near the expected fundamental frequency for a male talker, as used by Brungart *et al.* (2005). By comparison, a delay-and-add filter with a delay of 2 ms has a first dip at 250 Hz, near the expected fundamental frequency for female talkers, as used in the present experiments. Brungart *et al.* (2005) also reported release at delays of 0, 0.25, 0.5, and 1 ms of magnitudes similar to that of the release at 4 ms, whereas no release for 0 or 0.5 ms was found in the present experiment. This difference may be due to the fact that the present experiment was performed with two distracters instead of the single distracter employed by Brungart *et al.* (2005).

V. EXPERIMENT 4: SPECTRAL STRUCTURE—A FOLLOW-UP TO EXPERIMENTS 1 AND 3

Delay-and-add filtering in Experiment 3 produced multiple dips in the spectra of distracters. This experiment asked which of those dips was most important for masking release. If one dip or another is particularly important in eliciting a release from masking, this would give insight into the underlying mechanisms by which such a release from masking is attained.

A. Methods

A digital filter was designed to "mimic" the first dip in the amplitude response of a delay-and-add filter with delay τ=2 ms. This was a finite impulse response filter of order 336, with 0.04% ripple in the passband. The amplitude response of this filter is shown in Fig. 8(a). The dip in the filter's amplitude response was centered at 250 Hz and had a depth of approximately 39 dB. A second filter was designed with a dip at 750 Hz in order to mimic only the second dip in the delay-and-add filter. This filter [Fig. 8(b)] was FIR of order 392, with 0.02% passband ripple, and a dip at 750 Hz

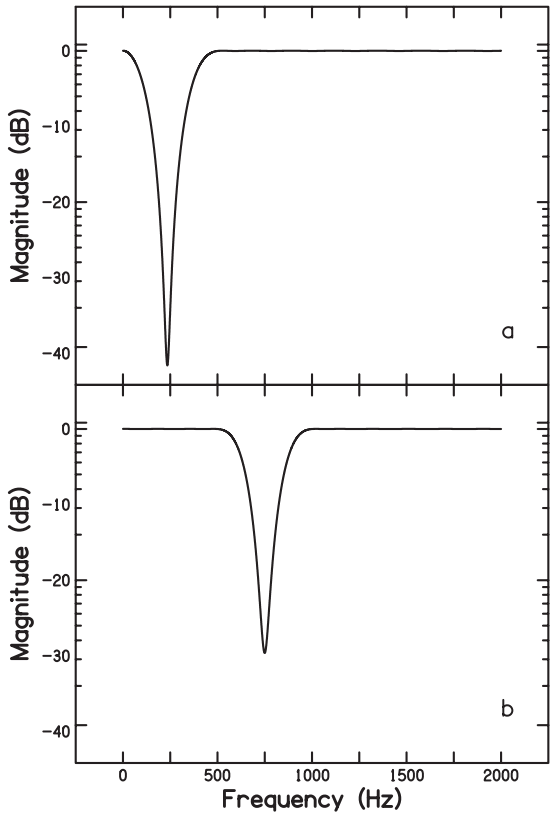


FIG. 8. The top graph shows the amplitude response of a filter imitating the first dip in a delay-and-add filter with delay τ=2 ms. This is a FIR filter of order 336. The bottom graph shows the amplitude response of a filter imitating the second dip in a delay-and-add filter with delay τ=2 ms. This is a FIR filter of order 392.

of 32 dB. By comparison, the first two spectral dips actually measured for the front-front geometry in the anechoic room used for these experiments occurred at 250 and 750 Hz, with depths of 22 and 12 dB, respectively.

The entire CRM stimulus set was processed with each digital filter to create two separate filtered corpi, one with energy removed at 250 Hz, and the other with energy removed at 750 Hz. These stimuli were then used individually as the distracters in two separate FO experiments. For each experiment, listeners completed three runs consisting of five practice trials (without feedback) followed immediately by 30 test trials. To simulate the effective target-to-distracter level difference of Experiment 1, wherein both loudspeakers were actively producing distracting speech, the level of the target talker was reduced here by 6 dB. The listeners in this experiment were the same as in all previous experiments.

B. Results and analysis

The results of Experiment 4 are shown in Table II for each listener and for the average across listeners. "FF" refers to the results of Experiment 3, speech-on-speech masking, with a delay of τ=2 ms, where the greatest release occurred. "FO 250 Hz" refers to the FO test wherein the distracters have a notch at 250 Hz per the first dip in a delay-and-add filter with delay τ=2 ms, and similarly for the condition la-

TABLE II. Experiment 4: Percentage of correct responses (plus or minus one standard deviation) for each listener by condition. (a) FO baseline condition. [(b) and (c)] Simulated delay-and-add filtering with a 2-ms delay. (d) Front-front condition with 2-ms delay—data taken from Experiment 3. Averages and standard deviations for each listener are calculated across three runs. Averages and standard deviations across listeners are shown in the bottom row.

Listener	(a) FO 0 dB	(b) FO 750 Hz	(c) FO 250 Hz	(d) FF
B	33 ± 3%	24 ± 8%	60 ± 3%	66 ± 9%
K	29 ± 5%	33 ± 7%	51 ± 8%	64 ± 7%
N	48 ± 5%	47 ± 0%	64 ± 7%	83 ± 6%
S	38 ± 2%	20 ± 0%	59 ± 8%	70 ± 6%
Avg.	37 ± 8%	31 ± 11%	67 ± 9%	71 ± 9%

beled "FO 750 Hz." The "FO 0 dB" condition refers to the baseline condition where target and unfiltered distracters are presented from the front loudspeaker only.

When the filter dip was at 250 Hz, all listeners showed an improvement in performance over the FO baseline, FO 0 dB. The average magnitude of the release was 2 dB. When the dip was at 750 Hz, none of the listeners showed any improvement in performance, and listener S performed consistently worse than in the FO baseline condition.

The release found for distracters with a notch at 250 Hz supports the idea that delay-and-add filtering was responsible for the release from masking demonstrated in Experiment 1. The lack of release found for distracters with a notch at 750 Hz suggests that only the first dip in the spectrum of delay-and-add filtering is responsible for this effect. However, since performance in this experiment in no case reached the level found in the front-front geometry for delay τ=2 ms [Table II, column (d)], this first spectral dip may not be solely responsible for the release in masking shown in Experiment 3. An alternative explanation for the smaller release seen in Experiment 4 is that the 6 dB target reduction used in Experiment 4 may have underestimated the effective S/N ratio in Experiment 3.

Experiment 4 indicates that the release from masking seen in Experiment 3, uniquely at delays of ±2 ms, was the result of eliminating the fundamental component of the distracting speech. An experiment by Freyman *et al.* (2005) supports this conclusion. That experiment began with a baseline condition wherein target and distracters were both high-pass filtered so as to remove the first few harmonics. When the fundamental was added back to the target, performance improved by about 2–3 dB. That experiment is quite similar to Experiment 4 in that it demonstrated a release from masking that takes place because of the lack of energy in the fundamental of the distracters compared to that of the target.

VI. SUMMARY

This article describes added delayed distractor (ADD) experiments in a front-back geometry, where care was taken to minimize interaural differences.

A. Speech distracters

Experiment 1 tested the ability of listeners to segregate target speech from distracting speech when both were presented from directly ahead, and when an additional time-shifted copy of the distracting speech was presented from behind. Thus, this experiment continued an ADD paradigm begun by Freyman *et al.* (1999), extending it into the median sagittal plane (MSP).

Experiment 1 showed that listeners experienced release from masking in the MSP for all delays tested between −32 and +32 ms. The magnitude of release was on the order of 2–4 dB (see Fig. 2), and the peak release occurred for delays of $\tau = \pm 2$ ms. Comparison to previous ADD studies conducted in the horizontal plane (HP) (Brungart *et al.*, 2005; Rakerd *et al.*, 2006) indicated that release occurs over a similar range of delays in both planes, though the magnitude of the release in the MSP is significantly less than that seen in the HP (8–11 dB). Greater release in the HP is not surprising since it seems likely that speech segregation could only be helped by binaural cues, which were strong in the HP, but were minimized in the MSP. Though binaural discrepancies (e.g., anatomical asymmetries) may have made some contribution to masking release in the front-back geometry, Appendix A shows that, by themselves, they could not have accounted for the release as observed.

The finding that this release occurred for a wide range of delays (as long as ±32 ms) is an important result of Experiment 1. Further, performance was found to be approximately independent of the sign of the delay. It did not matter which distracter led, either the front (coincident with the target) or the back, even for the longest delays.

Additional experiments were done in order to gain insight into the results of Experiment 1. Experiment 2 employed noises to measure energetic masking (EM) release in a front-back geometry. Experiments 3 and 4 examined the role of delay-and-add filtering.

B. Noise maskers

Experiment 2 was identical to Experiment 1 except that continuous noise maskers were used in place of speech distracters. In contrast to Experiment 1, Experiment 2 (Fig. 4) showed that EM release occurs in the MSP only for short delays, τ=0 ms and $\tau = \pm 0.5$ ms. The magnitude of the EM release and the range of delays over which it occurred were similar to the results previously seen in the HP (Freyman *et al.*, 2001; Rakerd *et al.*, 2006).

The results for noise maskers in Experiment 2 were also similar to the results of a previous study with noise using head-related transfer functions (HRTFs) and headphones to simulate a MSP geometry (Brungart *et al.*, 2005). In that study, which used both male and female target voices, a significant release from EM was found for delays of $\tau = \{-0.5, +1.0, +2.0\}$ ms. Brungart *et al.* (2005) explained that poor performance with a noise masker in an ADD experiment is expected, since the second presentation of the masker adds noise power to the masker, but that some delays may lead to a release due to delay-and-add filtering. Short delays in a delay-and-add filter lead to broad spectral valleys through which a listener may perceive an unfiltered target.

The release from EM in Experiment 2 can be understood from some combination of several effects. One effect is delay-and-add filtering. A delay of $\tau = \pm 0.5$ ms in the masking noise leads to a broad valley centered on 1000 Hz, giving the listener improved access to an important spectral region for the target speech. This is a plausible explanation for the release demonstrated by all listeners in Experiment 2 at $\tau = \pm 0.5$ ms. However, at $\tau = 0$ ms, no delay-and-add filtering occurs, and the release seen in Experiment 2 at this delay must have some other explanation. A second possibility is that release from EM at $\tau = 0$ ms, or for the entire range, $|\tau| \leq 0.5$ ms, is a localization effect caused by summing localization. Summing localization occurs for delays less than 1 ms, and is known to occur in the MSP (Litovsky *et al.*, 1997). Summing localization may shift the perceived location of the maskers away from the target, probably at the same time making the maskers more diffuse, and thus different from the target. A localization effect leading to a release from EM is, however, not expected because localization is thought to mediate release from IM and not from EM (Arbogast *et al.*, 2002).

As applied to Experiment 2, the possibility that a localization effect leads to a release from EM requires that summing localization produces a larger release from EM than is produced by the law of the first wavefront. That law is a part of the localization precedence effect. It says that the location of the leading source dominates. The problem is that the law of the first wavefront for broadband noise, for instance, at a delay of 4 ms, is very strong. However, no other explanations for the observed release are particularly forthcoming, especially as it occurs for a delay of $\tau = 0$.

C. The role of delay-and-add filtering

Experiments 3 and 4 examined the role of delay-and-add filtering in the results of Experiment 1 by removing the spatial aspect of the ADD experiment. In Experiment 3, the back and front loudspeakers from Experiment 1 were placed together in front so that the target, distracters, and added distracters were collocated. Otherwise, Experiment 3 was identical to Experiment 1. In this way, the effect of delay-and-add filtering was separated from spatial effects. The results (Fig. 6) show a release only for delays of $\tau = \pm 2$ ms, matching the delays at which peak release occurred in Experiment 1. Experiment 4 used digital signal processing techniques to show that the release for $\tau = \pm 2$ ms occurs mainly because the first dip in the delay-and-add filter occurs near the fundamental frequency of the distracting speech. The distracting speech becomes distinguishable from the target when its timbre is changed by attenuating the fundamental component.

D. Symmetry

The results of all the experiments showed a notable symmetry about the zero-delay condition. In Experiment 1, with a speech target and speech distracters, the average data shown in Fig. 2 were approximately symmetrical for all delays, 0.5–32 ms. Symmetry between positive and negative delays would not be expected *a priori* if the release from masking is primarily driven by the precedence effect. However, it is possible that the imperfect behavior of the localization precedence effect that is exhibited at negative delays is adequate to promote release from masking. For every delay showing appreciable release, the release was slightly stronger when the source in back led the source in front—positive delay. This small asymmetry might be attributed to the localization precedence effect.

E. Implications

1. Front-back

Previously, it was shown that a release from IM occurs in ADD experiments in the HP (Brungart *et al.*, 2005; Freyman *et al.*, 2001, 2005, 1999; Rakerd *et al.*, 2006), and binaural cues were held primarily responsible for this masking release. It has been shown now that such a release occurs as well in the MSP when binaural cues are minimized to the extent that they are expected to be unimportant. When distracters are presented from in front and in back the greatest release from masking occurs for a delay of ± 2 ms. Experiments with distracters only in front show that this peak owes much of its importance to delay-and-add filtering. Apart from that, the front-back experiment shows release from masking for a wide range of delays, at least out to 32 ms, and this release is likely caused by the ability of listeners to localize the distracters (or to delocalize them) using the localization cues that are available in the MSP, namely, spectral cues. Localization in the MSP is weaker than in the HP, and thus it was not surprising to find that the release that can be achieved in an ADD condition is smaller in the MSP than in the HP. The presence of masking release out to long delays, both positive and negative, reveals a behavior similar to that noticed in the HP (Rakerd *et al.*, 2006) and suggests that a similar general mechanism is at work to achieve a release from informational masking (IM) in both cases.

2. Precedence

In many previous studies, the precedence effect (Litovsky *et al.*, 1999, 1997) has been cited as the main mechanism by which release from IM is achieved in ADD experiments. By moving the *perceived* location of the distracters away from the target (i.e., for positive delays) localization of the distracters separately from the target allows for the perceptual segregation of the two. However, for large negative delays, the precedence effect will place the perceived location of the distracters near the target. According to a simple in-

terpretation of the precedence effect the distracters should nearly coincide with the target leading to no benefit to the listener trying to hear out the target speech. However, this study, as well as studies of the HP by Brungart *et al.* (2005) and Rakerd *et al.* (2006), shows a release from IM for a large range of negative delays. Furthermore, in both the HP and the MSP, the magnitude of the release is similar for both positive and negative delays. Freyman *et al.* (1999) argued that even a small shift in apparent location, caused by the negative delays, could mediate some release from masking. Other effects, such as perceived diffuseness and timbre changes caused by delay-and-add filtering, also likely contribute.

ACKNOWLEDGMENTS

The authors are grateful to Peter Xinya Zhang, who helped develop the special chair and speaker alignment system. Associate editor Richard Freyman and anonymous reviewers greatly improved the manuscript. Work was supported by the NIDCD Grant No. DC00181.

APPENDIX A: TEST FOR INTERAURAL DIFFERENCES

The methods used in this article had the goal of eliminating interaural differences by confining the sources to the front-back dimension. However, interaural differences cannot be completely eliminated, no matter how accurately the experimental system is aligned. Nevertheless, we believe that the small interaural differences that exist in a geometry such as ours have no perceptual importance (see also Middlebrooks and Green, 1991).

To test for the perceptual effects of interaural differences, which might arise because the listener is inadvertently misaligned or because of individual anatomical asymmetry, a headphone experiment was run in which misalignments were deliberately introduced such that the ears were not equidistant from each loudspeaker.

This test used a misalignment equivalent to a 5° rotation of the listener. Accordingly, one ear was effectively 65 μs closer to the front loudspeaker and 65 μs farther away from the back loudspeaker, for a total difference in arrival time of 130 μs. The same difference, but with opposite sign, occurred in the other ear.

Seven listeners, including B and N from the previous experiments, were presented with CRM stimuli of the same type used in Experiment 1 (target plus two speech distracters) through Sennheiser HD 414 headphones. The task was identical to that of previous experiments. The distracters were passed through delay-and-add filters before they were combined with the target speech. Two experimental scenarios were tested—one in which the delay between the maskers was 870 μs in the left ear and 1130 μs in the right, the "different-delays" scenario; and a second condition in which the delay between the maskers was the same in both ears 1000 μs, the "same-delays" scenario.

A reference delay of 1000 μs (1 ms) was chosen because that delay is representative of delays used in Experiment 1, and because it is beyond the range at which release from EM occurs (Experiment 2). For 1000 μs there is substantial release (but not the greatest release) from IM. These delay scenarios were tested at SNRs of −4, 0, and +4 dB, for a total of $2 \times 3 = 6$ conditions.

Differences in performance between the scenarios were examined within and across listeners for each SNR. The most relevant SNR is 0 dB, at which Experiment 1 was performed. At 0 dB, the average listener performed marginally better with different delays than with same delays. The mean improvement in percent correct, plus and minus one standard deviation, was $7 \pm 16\%$. This was not a significant improvement (one-sided Wilcoxon signed rank test, $N=5$, $W^+=3.0$, $p=0.140$). There were also differences in performance across listeners—three listeners performed slightly better with different delays, two performed slightly better with same delays, and two showed no difference in performance. For lower SNR, there was a greater improvement in performance under the different-delays condition, but the difference in performance did not reach the 0.05 level of significance ($N=5$, $W^+=1.5$, $p=0.069$). For higher SNR, the differences in performance between the same-delays and different-delays conditions were entirely negligible ($N=6$, $W^+=11.0$, $p=0.583$). Note here that a worst-case scenario (a 5° rotation of the listener) has been assumed. In reality, the error in alignment was almost certainly smaller. Thus, binaural discrepancies may have contributed to the unmasking observed in our experiment, but these discrepancies cannot account for the statistically significant unmasking obtained in those experiments.

APPENDIX B: SEPARATED-SOURCE PRESENTATION WITH SPEECH DISTRACTERS

The positive delay conditions of Experiment 1, particularly the conditions for which $\tau > 1$ ms, can be expected to elicit a precedence effect shift in the perceived location of the distracters. The precedence effect should shift the perceived location of the distracters from front to back. For comparison, it is interesting to investigate the masking release caused by a real geometrical shift. Several experiments have been performed, which have measured the release from both EM and IM when the distracters are moved from the front, where they were collocated with the speech target, to the back (Freyman *et al.*, 2005; Plomp, 1976; Zurek, 1993).

To test the effect of a real physical shift of the distracters using the setup of the previous experiments and the CRM stimuli, the experiment of Freyman *et al.* (2005) was repeated. Three conditions were tested: (1) the FO baseline, as in Experiment 1, with S/N ratio of 0 dB; (2) the FO baseline, as in Experiment 1, with S/N ratio of 4 dB; and (3) a separated-source test with the target presented from the front loudspeaker and the speech distracters presented from the

TABLE III. Separated-source experiment: Percentage of correct responses (plus or minus one standard deviation) for each listener by condition. (a) FO baseline condition. (b) FO condition with +4 dB target level relative to a single distracter. (c) Target in front and distracters in back. Averages and standard deviations for each listener are calculated across five runs. Averages and standard deviations across listeners are shown in the bottom row.

Listener	(a) FO	(b) FO+4 dB	(c) F-B
A	26 ± 8.3%	71 ± 5.0%	75 ± 8.4%
B	34 ± 8.6%	82 ± 4.5%	66 ± 7.8%
E	22 ± 9.0%	75 ± 3.8%	58 ± 6.9%
N	38 ± 9.0%	73 ± 4.7%	75 ± 8.3%
Avg.	30 ± 7.4%	75 ± 4.8%	68 ± 7.4%

back loudspeaker at a S/N ratio of 0 dB. The first two conditions were the FO and FO+4 dB conditions. The third condition was new.

Four listeners were tested, including listeners B and N from Experiment 1 and two inexperienced listeners, A and E. Listeners sat quietly and faced the front loudspeaker without moving their heads throughout each run. Listeners were positioned such that their ears were half way between the front and back loudspeakers. Each run consisted of five practice trials followed by 30 trials for which data were collected and scored as in Experiment 1. Each listener went through five runs of each condition.

The results of the experiment are presented in Table III. These may be informally compared to the results of Experiment 1. Performance in the separated-source condition was slightly worse than that in the FO+4 dB condition, indicating a release from masking somewhat less than 4 dB as a result of spatially separating the target and distracters. In detail, the average percent correct performance across listeners in the separated condition was 75 ± 4.8%, an increase over the FO baseline performance of 38 ± 7.4%. This increase in performance in the F-B condition with target in front and distractors in back corresponds to an average release of 3.5 dB. This is comparable in magnitude to the release seen in Experiment 1 between the FO baseline performance and the best performance at any other delay for the average listener (this occurred at τ=2 ms, where the average release was 3.5 dB).

[1]For listener K, the condition where the target was at 67 dB and each distracter was at 65 dB was the "0 dB SNR" condition, and other values of SNR are referenced to this baseline. For instance, the +4 dB condition corresponds to a target at a level of 71 dB and each distracter at 65 dB for listener K. Thus, values of SNR are reported relative to the initial level of the target, which was 65 dB for listeners B, N, and S; and 67 dB for listener K.

[2]The distance between the centers of the two speakers was 8.9 cm, which corresponded to a difference in vertical angle of 3.3°. Therefore, the arrival time discrepancy between the signals from the two speakers was less than 10 μs, producing a negligible effect on the delay-and-add filtering.

Arbogast, T. L., Mason, C. R., and Kidd, G., Jr. (**2002**). "The effect of spatial separation on informational and energetic masking of speech," J. Acoust. Soc. Am. **112**, 2086–2098.

Balakrishnan, U., and Freyman, R. L. (**2008**). "Speech detection in spatial and nonspatial speech maskers," J. Acoust. Soc. Am. **123**, 2680–2691.

Blauert, J. (**1969/70**). "Sound localization in the median plane," Acustica **22**, 205–213.

Blauert, J. (**1983**). *Spatial Hearing* (MIT, Cambridge, MA).

Bolia, R. S., Nelson, W. T., Ericson, M. A., and Simpson, B. D. (**2000**). "A speech corpus for multitalker communications research," J. Acoust. Soc. Am. **107**, 1065–1066.

Bronkhorst, A. W. (**2000**). "The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions," Acta Acust. **86**, 117–128.

Brungart, D. (**2001**). "Informational and energetic masking effects in the perception of two simultaneous talkers," J. Acoust. Soc. Am. **109**, 1101–1109.

Brungart, D. S., and Simpson, B. D. (**2007**). "Effect of target-masker similarity on across-ear interference in a dichotic cocktail-party listening task," J. Acoust. Soc. Am. **112**, 1724–1734.

Brungart, D., Simpson, B., Ericson, M., and Scott, K. (**2001**). "Informational and energetic masking effects in the perception of multiple simultaneous talkers," J. Acoust. Soc. Am. **110**, 2527–2538.

Brungart, D., Simpson, B., and Freyman, R. L. (**2005**). "Precedence-based speech segregation in a virtual auditory environment," J. Acoust. Soc. Am. **118**, 3241–3251.

Burger, J. F. (**1958**). "Front-back discrimination of the hearing system," Acustica **8**, 301–302.

Cherry, E. C. (**1953**). "Some experiments on the recognition of speech with one and two ears," J. Acoust. Soc. Am. **25**, 975–979.

Duquesnoy, A. J. (**1983**). "Effect of a single interfering noise or speech source upon the binaural sentence intelligibility of aged persons," J. Acoust. Soc. Am. **74**, 739–743.

Durlach, N. I., Mason, C. R., Kidd, G. J., Arbogast, T. L., Colburn, H. S., and Shinn-Cunningham, B. G. (**2003**). "Note on informational masking (1)," J. Acoust. Soc. Am. **113**, 2984–2987.

Freyman, R. L., Balakrishnan, U., and Helfer, K. S. (**2001**). "Spatial release from informational masking in speech recognition," J. Acoust. Soc. Am. **109**, 2112–2122.

Freyman, R. L., Balakrishnan, U., and Helfer, K. S. (**2004**). "Effect of number of masking talkers and auditory priming on informational masking in speech recognition," J. Acoust. Soc. Am. **115**, 2246–2256.

Freyman, R. L., Helfer, K. S., and Balakrishnan, U. (**2005**). "Spatial and spectral factors in release from informational masking in speech recognition," Acta Acust. **91**, 537–545.

Freyman, R. L., Helfer, K. S., and Balakrishnan, U. (**2007**). "Variability and uncertainty in masking by competing speech," J. Acoust. Soc. Am. **121**, 1040–1046.

Freyman, R. L., Helfer, K. S., McCall, D. D., and Clifton, R. K. (**1999**). "The role of perceived spatial separation in the unmasking of speech," J. Acoust. Soc. Am. **106**, 3578–3588.

Haas, H. (**1951**). "On the influence of a single echo on the intelligibility of speech," Acustica **1**, 49–58.

Hartmann, W. M. (**1998**). *Signals, Sound, and Sensation*, 2nd ed. (Springer-Verlag, New York).

Hirsh, I. J. (**1950**). "Relation between localization and intelligibility," J. Acoust. Soc. Am. **22**, 196–200.

Kidd, G. J., Mason, C. R., Deliwala, P. S., Woods, W. S., and Colburn, H. S. (**1994**). "Reducing informational masking by sound segregation," J. Acoust. Soc. Am. **95**, 3475–3480.

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., Rakerd, B., Yin, T. C. T., and Hartmann, W. M. (**1997**). "Psychophysical and physiological evidence for a precedence effect in the median sagittal plane," J. Neurophysiol. **77**, 2223–2226.

Middlebrooks, J. C., and Green, D. M. (**1991**). "Sound localization by human listeners," Annu. Rev. Psychol. **42**, 135–159.

Peissig, J., and Kollmeier, B. (**1997**). "Directivity of binaural noise reduction in spatial multiple noise-source arrangements for normal and impaired listeners," J. Acoust. Soc. Am. **101**, 1660–1670.

Plomp, R. (**1976**). "Binaural and monaural speech intelligibility of connected discourse in reverberation as a function of azimuth of a single competing sound source (speech or noise)," Acustica **34**, 200–211.

Rakerd, B., Aaronson, N. L., and Hartmann, W. M. (**2006**). "Release from speech-on-speech masking by repeated presentation of the masker," J. Acoust. Soc. Am. **119**, 1597–1605.

Roffler, S. K., and Butler, R. A. (**1968**). "Factors that influence the localization of sound in the vertical plane," J. Acoust. Soc. Am. **48**, 1255–1259.

Wallach, H., Newman, E. B., and Rosenzweig, M. R. (**1949**). "The precedence effect in sound localization," Am. J. Psychol. **57**, 315–336.

Wightman, F. L., and Kistler, D. J. (**1999**). "Resolution of front-back ambiguity in spatial hearing by listener and source movement," J. Acoust. Soc. Am. **105**, 2841–2853.
Yost, W. A. (**1997**). "The cocktail party problem: Forty years later," in *Binaural and Spatial Hearing in Real and Virtual Environments*, edited by R. H. Gilkey and T. R. Anderson (Erlbaum, Hillsdale, NJ), pp. 329–347.
Zurek, P. M. (**1993**). "Binaural advantages and directional effects in speech intelligibility," in *Acoustical Factors Affecting Hearing Aid Performance*, 2nd ed., edited by G. A. Studebaker and I. Hockberg (Allyn and Bacon, Needham Heights, MA).

Limits of temporal pitch in cochlear implants

Ying-Yee Kong[a)]
Department of Speech Language Pathology & Audiology, Northeastern University, Boston, Massachusetts 02115 and MRC Cognition and Brain Sciences Unit, 15 Chaucer Road, Cambridge CB2 7EF, United Kingdom

John M. Deeks
MRC Cognition and Brain Sciences Unit, 15 Chaucer Road, Cambridge CB2 7EF, United Kingdom

Patrick R. Axon
Department of Otoneurological and Skull Base Surgery, Addenbrooke's Hospital, Cambridge CB2 2QQ, United Kingdom

Robert P. Carlyon
MRC Cognition and Brain Sciences Unit, 15 Chaucer Road, Cambridge CB2 7EF, United Kingdom

(Received 15 January 2008; revised 5 December 2008; accepted 13 December 2008)

A common finding in the cochlear implant literature is that the upper limit of rate discrimination on a single channel is about 300 pps. The present study investigated rate discrimination using a procedure in which, in each block of two-interval trials, the standard could have one of the five baseline rates (100, 200, 300, 400, and 500 pps) and the signal rate was a given percentage higher than the standard. Eight Med-El C40+ subjects took part. The pattern of results was different than those reported previously: six Med-El subjects performed better at medium rates (200–300 pps) compared to both lower (100 pps) and higher (400–500 pps) rates. A similar pattern of results was obtained both with the method of constant stimuli and for 5000-pps pulse trains amplitude modulated at rates between 100 and 500 Hz. Compared to an unmatched group of eight Nucleus CI24 listeners tested using a similar paradigm and stimuli, Med-El subjects performed significantly better at 300 pps and higher but slightly worse at 100 pps. These results are discussed in relation to evidence on the limits of temporal pitch at low and high rates in normal-hearing listeners.
 [DOI: 10.1121/1.3068457]

PACS number(s): 43.66.Ts, 43.66.Hg [RLF] Pages: 1649–1657

I. INTRODUCTION

Perhaps the major remaining challenges for cochlear implant (CI) researchers are to improve the perception of pitch and the ability of listeners to extract speech from competing backgrounds. These two issues may be related, given that pitch differences between different sources provide a major cue for segregation in normal-hearing (NH) listeners, and that such differences are not readily exploited by CI users (Stickney *et al.*, 2004, 2007).

In a typical envelope extraction algorithm used in a CI speech processor, pitch is encoded in the temporal envelope with a modulation rate equal to the fundamental frequency (F0) of a sound. For sounds presented via these speech processors, pitch perception is generally very poor (e.g., Gfeller and Lansing, 1991; Kong *et al.*, 2004; McDermott, 2004; Vandali *et al.*, 2005). Slightly better performance can be achieved when the speech processor is bypassed and a simple stimulus, such as a train of equal-amplitude pulses, is presented to one channel of a CI (e.g., Pijl and Schwarz, 1995; Pijl, 1997). Even then, rate difference limens (DLs) are orders of magnitude higher than for low-frequency pure tones in NH listeners, and, typically, performance deteriorates markedly at pulse rates greater than about 300 pps (e.g., Shannon, 1983; Tong *et al.*, 1983; Townshend *et al.*, 1987; McDermott and McKay, 1997; Zeng, 2002). The poor temporal pitch perception in conjunction with the deterioration of rate discrimination at higher rates reflects a limitation that is likely to discourage attempts to improve CI speech processors by encoding the temporal fine structure of the signal (Wilson *et al.*, 2004).

The upper limit of 300 pps is remarkably lower than that observed in NH listeners presented with filtered pulse trains. Carlyon and Deeks (2002) investigated the upper limit of rate discrimination in a group of NH listeners using alternating-phase harmonic complexes. When the stimuli were filtered into the very high frequency region between 7800 and 10800 Hz, NH listeners could detect differences in pulse rate relative to a 712-pps baseline rate, an upper limit higher than that reported in CI listeners. The reasons for the discrepancies in performance between CI and NH listeners remain unclear. It has been suggested that accurate pitch perception may depend on a match between place and temporal cues, and that the mismatch between these two cues in CIs may limit discrimination performance in CI listeners (Oxenham *et al.*, 2004; Moore and Carlyon, 2005). In addition, or alternatively, stimulation by electric pulse trains does not

a)Author to whom correspondence should be addressed. Electronic mail: yykong@neu.edu

preserve the phase transitions near the peak of the traveling wave in the normal cochlea, information that may be important for pitch estimation (Kim *et al.*, 1980).

The goal of this study was to investigate the upper limit of temporal pitch and elucidate the mechanisms underlying temporal pitch perception in electric stimulation, with a view toward testing the effect of novel stimulus manipulations that could shed light on the basis for this limitation. We tested a group of listeners implanted with the Med-El COMBI 40+ implant, which has a longer electrode array than other devices, allowing one to measure performance across a wider range of cochlear sites. If perception of temporal pitch is dependent on a match between place and temporal cues, then the apical electrodes might show better rate discrimination at lower rates and the limits of temporal pitch would increase as the stimulating electrode progressively moved toward the basal locations.

Three experiments were conducted to examine temporal pitch perception in Med-El COMBI 40+ implant users. Experiment 1 measured rate discrimination performance as a function of baseline rate. We developed a procedure that allows us to quickly measure rate discrimination as a function of baseline rate by choosing a reasonably large rate difference (ΔR) of 35% between standards and signals. The assumption is that if subjects have the ability to discriminate differences in rate, they should be able to accomplish this at 35% ΔR. The limit of temporal pitch will then be revealed at the baseline rate where peformance approaches chance level. A potential advantage of this procedure is that, by mixing several baseline rates within a test block in our procedure, we can minimize the possibility that subjects will learn to use non-pitch-related cues to perform the task, a potential advantage over the conventional method of constant stimuli (Fechner, 1966) and the traditional adaptive procedure (Levitt, 1971), in which the same standard is used on every trial. The results revealed an unusual nonmonotonic function in some subjects, with best performance at a baseline rate of 200–400 pps, inconsistent with previous reports on patients implanted with Nucleus device (e.g., Shannon, 1983; Tong *et al.*, 1983; Townshend *et al.*, 1987; Zeng, 2002). This pattern of results was verified when listeners were tested with a different procedure that fixed the baseline rate and varied the ΔR within a test block. To further investigate the perception of temporal pitch, Experiment 2 measured rate discrimination with sinusoidually amplitude-modulated (SAM) stimuli in the same group of listeners. Amplitude-modulation (AM) rate discrimination was similar to that with equal-amplitude pulse trains, suggesting that the nonmonotonic function observed in some subjects was due neither to stimulus error nor to some idiosyncratic responses to the pulse trains; rather, it reflects a genuine limitation on the ability of these subjects to perceive pitch at different rates. Experiment 3 investigated the effect of place of stimulation on the limits of temporal pitch. Results showed a lack of consistent electrode effect on the rate discrimination performance and the shape of the rate discrimination function across rates. Taken together, the decrease in performance at the lowest (100 pps) and highest (500 pps) baseline rates may indicate the lower and upper bounds of temporal pitch in some of our implant subjects. Issues regarding the lower and upper limits of temporal pitch are discussed in relation to evidence on the limits of temporal pitch in NH listeners.

II. EXPERIMENT 1: RATE DISCRIMINATION

A. Method

1. Subjects

Eight Med-El COMBI 40+ CI users (M1–M8; four males) were recruited. Their ages ranged from 36 to 68 years with a mean of 52.25 years. All subjects were implanted with the standard long electrode array with a full insertion up to 30 mm into the cochlea. The electrode array contained 12 electrodes, spaced 2.4 mm apart. All of them, except for M5, were postlingually deafened adults who had at least one year of implant usage prior to the testing. Subject M5 became severely hearing impaired at age three due to meningitis. They were all regular hearing aid users prior to implantation. Table I provides additional information regarding hearing history, etiology of hearing loss, and the duration of implant use for each subject.

2. Stimuli

The Med-El COMBI 40+ implant uses only monopolar stimulation with a reference electrode positioned beneath the temporal muscle. The electrical stimuli were 300-ms biphasic pulse trains, and there was no interphase gap. Stimulation rates were fixed at 100, 200, 300, 400, and 500 pps for the five baseline rate conditions. The interstimulus interval was 400 ms. Electric stimuli were presented at subjects' maximal comfortable level (MCL). Pulse phase durations were 26.7, 40, or 100 μs depending on the loudness growth function of each subject. For some subjects (e.g., M1, M6, and M7), longer-duration pulses were used due to the fact that, at low pulse rates, loudness did not reach the MCL at the device compliance limit with the shorter pulse phase duration. The pulse phase duration for each subject is indicated in Table I. The signal rate was fixed at 35% higher than the baseline rate, yielding 135, 270, 405, 540, and 675 pps for the baseline rate conditions of 100, 200, 300, 400, and 500 pps, respectively. The decision to use a 35% ΔR was based on the rate discrimination results reported previously. For example, Zeng (2002) reported that rate DLs (71% correct point) were about 10–30 pps for a baseline of 100 pps, and about 100–300 pps for a baseline of 300 pps. Subjects should be able to perform well above chance for at least some rates within the block with this reasonably large ΔR. Additionally, two subjects (M3 and M7) were tested with signal rates 20% higher than the baseline rate. Before the experiment, all stimuli were calibrated using a "detector box," which replicates the electronics of the implanted part of the COMBI 40+, and which was connected to a LeCroy LT344 digital storage oscilloscope. Stimuli were controlled using a research interface box (RIB) and associated software provided by Med-El. Because of speed limitations associated with the RIB software and hardware available at the time, the delay between the button press after each trial and the start of the next trial was approximately 2.5 s. This delay did not depend on the pulse rates used in each trial.

TABLE I. Subject details.

Subject	Device	Age	Gender	Ear	Pulse phase duration (μs)	Etiology	Dur. profound/severe HL (yr)[a]	Dur. CI use (month)	Strategy
M1	Med-El	56	M	R	100	Progressive+viral	32	30	CIS+
M2	Med-El	56	F	L	40	Progressive+radiotherapy	22	15	CIS+
M3	Med-El	47	F	R	26.7	Congenital progressive	30	14	CIS+
M4	Med-El	68	M	R	26.7	Otosclerosis	35	14	CIS+
M5	Med-El	50	M	L	40	Meningitis	47	24	CIS+
M6	Med-El	48	F	L	40	Unknown	6	18	CIS+
M7	Med-El	36	M	R	40	Meningitis	21	30	CIS+
M8	Med-El	57	F	R	26.7	Progressive+chemotherapy	36	15	CIS+
C1	CI24M	50	F	L	45	Idiopathic or measles	44	69	SPEAK
C2	CI24M	64	M	L	45	Congenital progressive	64	69	SPEAK
C3	CI24M	69	M	R	45	Otosclerosis	44	66	ACE
C4	CI24M	76	M	R	45	Otosclerosis+noise	51	54	SPEAK
C5	CI24M	67	M	L	45	Progressive unknown	19	45	SPEAK
C6	CI24M	75	F	L	45	Progressive unknown	7	48	SPEAK
C7	CI24M	59	F	L	45	Chronic suppurative otitis media	9	84	SPEAK
C8	CI24R	62	M	L	45	Sudden hearing loss	11	11	SPEAK

[a]Years of profound or severe hearing loss prior to implantation.

3. Procedure

Rate discrimination was performed on a single electrode using a two-interval two-alternative force-choice (2I2AFC) procedure, in which subjects were asked to choose the interval that contained the higher pitched sound. Visual feedback was provided after each trial. Percent correct discrimination was measured as a function of the five baseline rates.

The standard and signal stimuli were loudness balanced before the experiment, using the method described by Landsberger and McKay (2005). A pair of stimuli was presented to each subject. At the initial presentation, one of the stimuli in the pair was fixed at MCL and the other was presented at a much quieter level. The subject adjusted the level of the quieter sound until the loudness of the two stimuli was the same. They were encouraged to raise the level of the variable sound until it was slightly louder than the fixed sound and then reduce the level until the loudness of the two sounds was the same. First, all baseline rates were loudness matched to the baseline rate of 300 pps (e.g., 300 versus 500 pps). Then the signal rate was loudness matched to the baseline rate for each of the baseline rate conditions (e.g., 100 versus 135 pps). It should be noted that, despite our effort to minimize the use of loudness cues, we still cannot completely rule out the use of such cues for some subjects. In each block, all five different baseline rates were presented in random order. Fifty trials were presented within each block, with ten trials per baseline rate. A total of ten blocks were presented to each subject, resulting in 100 trials per data point. Electric pulses were delivered to the middle electrode (El 7), except for subject M1. On El 7, M1's loudness did not reach the MCL even for a long pulse phase duration of 100 μs. Therefore, El 8 was used for this subject.

B. Results and discussion

Percent correct discrimination for each of the five baseline rates was obtained for each subject (see Fig. 1). A 95% confidence interval was plotted for the individual data (see the vertical bars). On average, these subjects achieved 70%, 83%, 85%, 75%, and 71% correct for the baseline rate conditions of 100, 200, 300, 400, and 500 pps, respectively. Averaged across runs and subjects (the mean data in Fig. 1), a repeated-measures analysis of variance (ANOVA) revealed a significant rate effect [$F(4,28)=4.90$, $p<0.05$] and a significant quadratic trend [$F(1,7)=20.22$, $p<0.005$], indicating a nonmonotonic performance as a function of rate. Paired t-tests with Bonferroni corrections were performed for each subject to compare the performance between the lowest rate

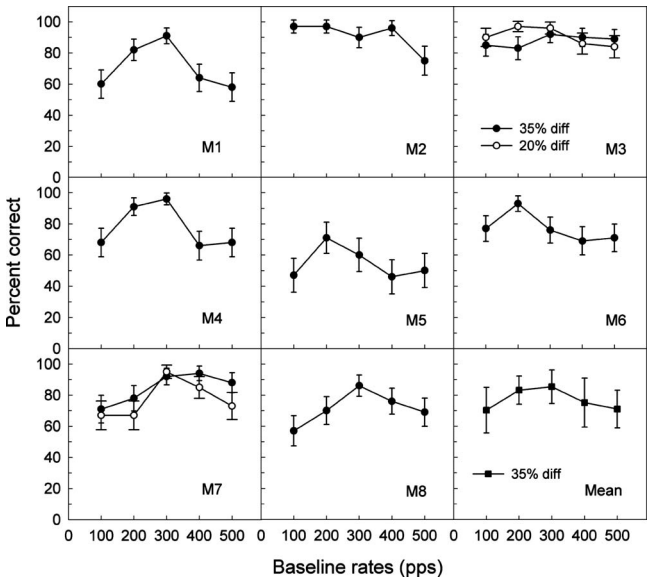


FIG. 1. Rate discrimination performance in eight Med-El C40+ users using the mixed-block procedure. Solid circles: signal rates 35% higher than the baseline rate. Open circles: signal rates 20% higher than the baseline rate. Vertical bars represent 95% confidence intervals of the mean.

of 100 pps and one of the three mid-frequency rates (200–400 pps) that yielded the best discrimination score, as well as between the mid-frequency rate (200 or 300 pps) and the highest rate at 500 pps. Results revealed that five out of the eight subjects (M1, M4, M5, M7, and M8) performed significantly better at 200 or 300 pps compared to the lower rate of 100 pps, and six subjects (M1, M2, M4, M5, M6, and M8) performed significantly better in the mid-frequency range compared to the highest rate of 500 pps. This pattern of results differed from the common reports in the implant literature of steadily decreasing performance with increasing pulse rate (e.g., Shannon, 1983; Tong *et al.*, 1983; Townshend *et al.*, 1987; van Hoesel and Clark, 1997; Zeng, 2002). The only study on rate discrimination on Mel-El users was by Baumann and Nobbe (2004), but they did not test rates below 200 pps. Second, all of our subjects performed significantly higher than chance level (50% correct in a 2I2AFC task) at 300 pps and six out of the eight subjects (except for M1 and M5) performed significantly above chance at 500 pps reflected by the 95% confidence intervals (the vertical bars) in Fig. 1. Although comparisons across studies using different procedures and stimuli should be made with some caution, this tentatively suggests that our subjects had a rather higher upper limit of rate discrimination than previously reported (e.g., Shannon, 1983; Tong *et al.*, 1983; Zeng, 2002).

To summarize, most subjects showed an unusual nonmonotonic function with best performance at baseline rates of 200–400 pps. One subject, M2, showed the typical low-pass characteristic. However, performance for M2 approached ceiling at the low rates (100 and 200 pps), which may have concealed the actual discriminability differences between the two rate conditions. Subject M3 showed a relatively flat function with almost perfect performance at all rates tested. The true pattern of performance from this subject is undetermined, also due to this ceiling effect.

Possible reasons for the unusual pattern of results are as follows:

1. Stimulus error

Although our stimuli were calibrated on a digital storage scope via a detector box, it is possible that in a real implant, the stimuli delivered from the electrodes differed from that programmed. To verify that no pulses were "dropped" by the implant and that the input-output (I/O) function of the device was as specified, we recorded the response to our stimuli in our subjects using surface electrodes. Our recordings showed that the frequencies of the recorded signals were identical to the input stimuli, and the I/O function of the potentials was proportional to the I/O function measured using the detector box, down to the lowest level studied.

2. Test method

Our method of mixing various baseline rates within a single block has not been used by other researchers. One could attribute the unusual pattern of results reported in this study to our novel test procedure (we refer this procedure as "mixed-block" procedure). For example, the baseline rate changed from trial to trial, which may create an effect of frequency uncertainty. On average, the change in baseline rate between any two consecutive trials would have been larger for the extreme than for the intermediate rates, and this nonsensory factor may have affected the pattern of results observed. It has been shown that auditory perception can be affected by the edge effect and by the uncertainty of the signal frequencies. For example, Braida *et al.* (1984) demonstrated that intensity resolution changed near the extremes of the experimental intensity range. Scharf *et al.* (1987) showed that detection of a pure tone in noise improved when the frequency of the signal was fixed compared to a condition where the signal frequency was uncertain.

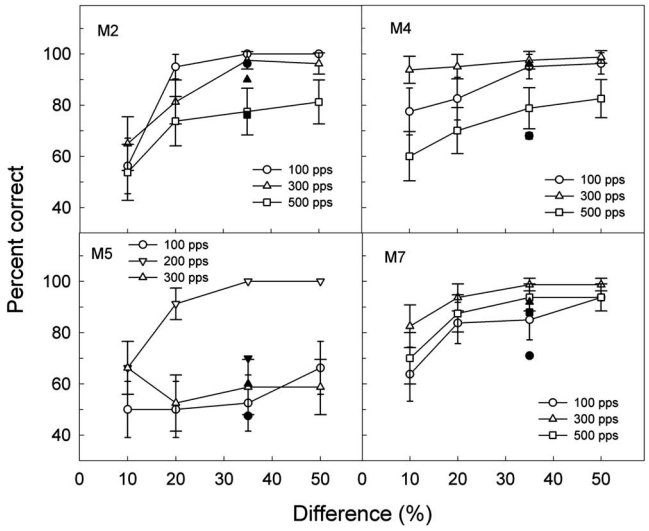


FIG. 2. Percent correct rate discrimination as a function of the percentage difference between signal and baseline rates (100 pps: open circles; 200 pps: open inverted triangles; 300 pps: open upright triangles; and 500 pps: open squares). Results using the mixed-block procedure with ΔR at 35% are indicated by the closed symbols. Vertical bars represent 95% confidence intervals of the mean.

We carried out a follow-up experiment using the traditional method of constant stimuli to test if our mixed-block procedure was responsible for the unusual pattern of results. We tested four subjects (M2, M4, M5, and M7) with the same test stimuli and 2I2AFC procedure, except that this time the experiment was blocked with only one baseline rate presented within each block. The signal rates were fixed at 10%, 20%, 35%, and 50% higher than the baseline rate. The four signal rates were presented within one block in random order. We refer this procedure as the "fixed-block" procedure. Each subject was tested for a total of 16 blocks for each baseline rate condition, resulting in 80 trials per data point on the psychometric function.

Figure 2 shows individual rate discrimination performance for each baseline rate. Results from the mixed-block procedure are replotted on the same graph indicated by the closed symbols to facilitate comparison. The main finding is that the pattern of results was similar between the mixed-block procedure and the fixed-block procedure for each subject. Subject M2, who exhibited a low-pass characteristic

with the mixed-block procedure, also showed the same characteristic with the fixed-block procedure. The remaining three subjects, who exhibited nonmonotonic functions with the mixed-block procedure, also showed the same pattern with the fixed-block procedure. However, all subjects performed better at all baseline rate conditions with the fixed-block procedure compared to the mixed-block procedure at 35% difference. Excluding the 200-pps performance for subject M5, the average difference in performance between procedures was less than ten percentage points. This small difference may be due to (1) a learning effect, because all subjects were tested with the mixed-block procedure prior to the fixed-block procedure and/or (2) the nonsensory factors that we have mentioned previously. Regardless, we emphasize that the patterns of results did not differ between test procedures.

3. Device and/or patient differences

Our study is the first to measure rate discrimination at 100 pps with the Med-El device. Baumann and Nobbe (2004) measured rate discrimination with Med-El implants but only did so at 200 pps and higher rates, whereas most of the previous studies on rate discrimination were performed with users of the Nucleus device. To examine the device differences that may have contributed to the unique patterns of results, we tested eight Nucleus CI24 implant listeners (C1–C8; five males) with the same mixed-block procedure. These subjects were between 50 and 76 years of age with a mean of 65.25 years. Detailed subject information is shown in Table I. The mean duration of profound/severe hearing loss was similar between the Med-El and CI24 implant users (Med-El: 28.6 years; CI24: 31 years). The ages of implantation were also similar between the two groups, ranging from 34 to 67 years of age in the Med-El group and from 44 to 71 years of age in the CI24 group. However, the mean duration of implant use was longer for the CI24 subjects than the Med-El subjects (Med-El: 20 months; CI24: 55.75 months). Test stimuli were similar to those used in Med-El subjects except that the biphasic pulse trains were 200 ms in duration with a fixed pulse phase duration of 45 μs. The stimulation electrode was also the middle electrode (El 11) of the array. All subjects were presented with 10 blocks of trials, resulting in 100 trials per baseline rate, except for C5 who only was tested with 5 blocks due to time constraints.

Figure 3 shows the rate discrimination performance in CI24 users. Four of the eight subjects (C2, C4, C6, and C8) showed the classic low-pass function, in which performance dropped off at a higher rate at 300 pps and was close to chance at rates above 300 pps. However, two subjects (C1 and C7) showed a non-monotonic pattern, with performance at 200 pps significantly higher than that at 100 pps ($p < 0.05$). Note that the nonmonotonic trend is present in C5, but did not reach statistical significance, possibly due to the insufficient number of runs (only 50 trials per data point). One subject, C3 showed an unusual ("U" shape) pattern of performance with the best rate discrimination at the lowest rate of 100 pps, followed by the highest rates of 400 and 500 pps. Further measures with this subject, using different

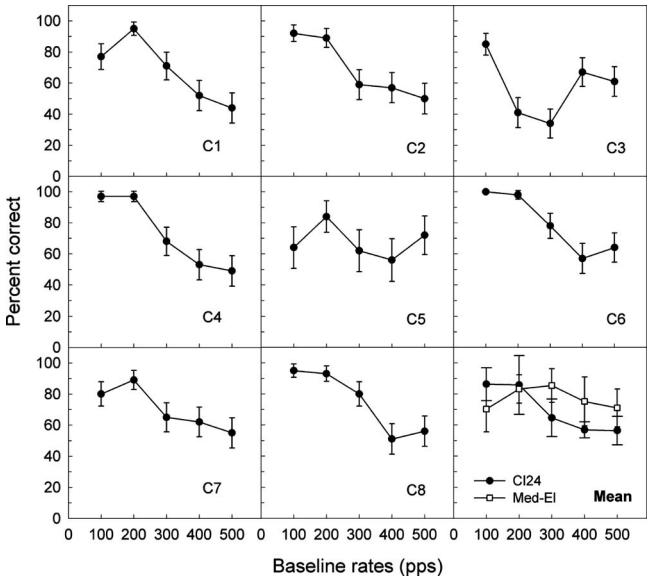


FIG. 3. Rate discrimination as a function of baseline rates in eight CI24 subjects (C1–C8). The mean data from the eight Med-El subjects (see Fig. 1) are replotted in the bottom rightmost panel along with the mean data from the eight CI24 subjects.

stimulus durations, revealed that this pattern was highly repeatable. When averaged across subjects (solid symbols in the bottom rightmost panel), the performance function shows a low-pass characteristic, with the discrimination performance (65% correct) approaching chance level at 300 pps. The mean data (35% ΔR) from eight Med-El subjects are replotted on the same panel (open squares) to facilitate comparisons. A repeated-measures ANOVA performed on the CI24 data revealed a significant rate effect [$F(4,28)=11.80$, $p<0.005$], but a linear trend [$F(1,7)=27.79$, $p<0.005$], unlike the quadratic trend observed in the Med-El group. A mixed-design two-factor (device $\times$ rate) ANOVA on the data revealed a significant interaction between device and rate [$F(4,56=8.61$, $p<0.005$]. Subsequent t-tests revealed significantly better performance for the Med-El group at 300 pps [$t(14)=3.15$, $p=0.007$], 400 pps [$t(14)=2.84$, $p=0.013$], and 500 pps [$t(14)=2.56$, $p=0.023$], and better performance for the CI24 group at 100 pps [$t(14)=2.25$, $p=0.041$].

The reason for the different pattern of results between these two groups of subjects is unclear. We should interpret the group difference with caution. It is noted that the age range, age of implantation, and the duration of severe/profound hearing loss in both groups are roughly the same. One difference is that the Med-El group was implanted more recently (within the past 3–4 years), while the CI24 group was implanted in an earlier period (2–5 years before the Med-El group). Another is the fact that the two groups of subjects experienced different speech-processing strategies in everyday life, with the CI24 patients using SPEAK or ACE, and the Med-El patients using CIS+. In our study, it could not be determined whether differences in rate discrimination performance between the two groups were genuinely due to the differences in the devices or other unknown differences in the demography in our sample.

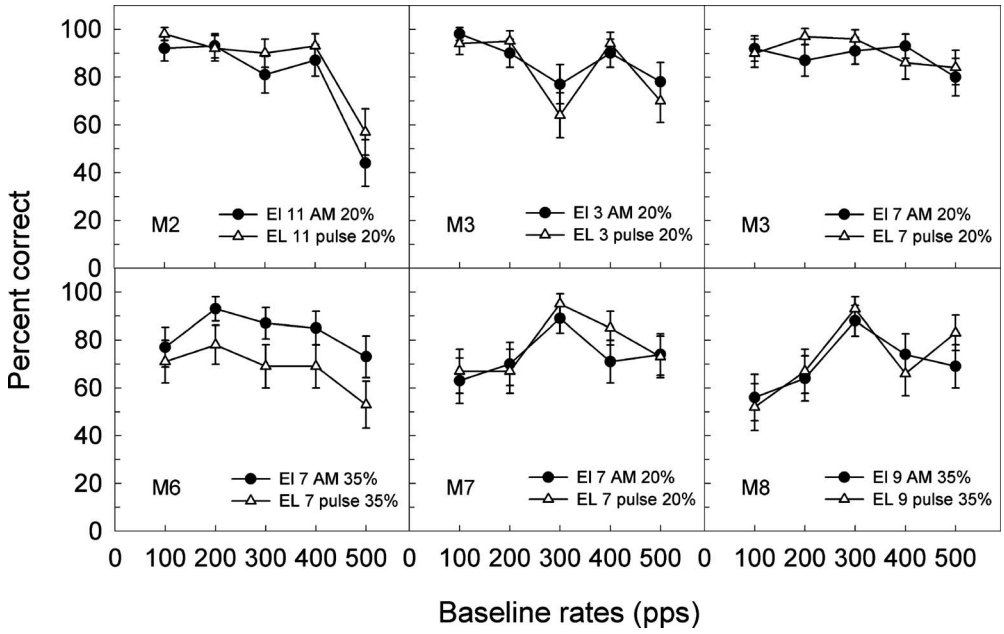


FIG. 4. Rate discrimination with SAM pulse train (closed circles) and equal-amplitude pulse train (open triangles) in five subjects. Note that subject M3 was tested on two different electrodes (El 3 and El 7). Vertical bars represent 95% confidence intervals of the mean.

III. EXPERIMENT 2: RATE DISCRIMINATION WITH SINUSOIDUALLY AMPLITUDE-MODULATED STIMULI

A. Rationale

Pulse-rate discrimination on a single electrode (experiment 1) is believed to be mediated by timing cues encoded in the auditory nerve. In acoustic stimulation, a change in pitch percept can be elicited using amplitude-modulated noise by varying the rate of modulation up to about 500 Hz (Burns and Viemesiter, 1976, 1981). In electric stimulation, a temporally coded pitch percept can also be elicited using a SAM pulse train, where the perceived pitch corresponds to the timing interval between the peaks of the envelope (McKay *et al.*, 1994; McDermott and McKay, 1997). Here, we further investigated the limits of temporal pitch using SAM pulse trains.

B. Methods

Five of the Med-El users took part. Each subject was tested on both rate discrimination with unmodulated pulse trains (same as experiment 1) and discrimination of modulation rates of SAM stimuli. Prior to the experiment, the detection threshold and the MCL of the unmodulated high-rate pulse trains were measured for each subject. During the threshold measurement, some subjects were able to hear an artifactual percept consisting of a brief but very soft click when stimulation was started. A threshold was subsequently determined as the lowest level at which the subject heard a longer-duration stimulus with a tonal quality instead of a short click. For the SAM stimuli, the pulse rate of the carrier was set at 5000 pps and the pulse phase duration at 26 μs. The high-rate pulse train was then modulated with a sinusoid having a starting phase of $3\pi/2$, with maxima and minima corresponding to the MCL and threshold (T) levels of an unmodulated 5000-pps pulse train. The baseline modulator frequencies were 100, 200, 300, 400, and 500 Hz.

The experimental procedure was also the same as experiment 1. A total of ten blocks were presented to each subject, resulting in 100 trials per data point. However, the difference between the baseline rate and the signal rate differed in different subjects. To avoid the ceiling effect, subjects M2, M3, and M7 were tested with the signal rate at 20% higher than the standard rate, smaller than the rate difference used in experiment 1. The remaining two subjects, M6 and M8, were tested with ΔR at 35%. Note that we tested subject M3 on two electrodes (apical El 3 and middle El 7), subject M2 on El 11, and subject M8 on El 9, a different electrode than experiment 1. Subjects M6 and M7 were tested on the middle electrode El 7. Subjects M2, M3, and M8 performed the pulse rate and modulation rate discrimination tasks in two separate sessions within a month of each other. Subjects M6 and M7 performed both tasks within the same test session. Subject M6's pulse-rate discrimination (same test condition as in experiment 1) was re-measured in this experiment to minimize any learning effect since the first measurement was performed three months prior to this experiment. Subjects M2, M6, and M7 were tested with pulse-rate discrimination first, followed by AM-rate discrimination. Subjects M3 (on both electrodes) and M8 were tested in the reverse order.

C. Results and discussion

Figure 4 shows the percent correct rate discrimination as a function of baseline rates in both tasks for each subject. Modulation-rate discrimination and pulse-rate discrimination performance were indicated by the closed circles and open triangles, respectively. First, visual inspection of the data revealed similar patterns of results between the SAM stimuli

and unmodulated pulse train stimuli for all subjects. Subject M2 showed a low-pass characteristic for both psychometric functions. Subjects M6, M7, and M8 showed nonmonotonic functions, with the best performance at the same baseline rate for both tasks (M6 at 200 pps, M7 and M8 at 300 pps). Subject M3 performed the worst for both tasks at the 300 pps condition on El 3 and was close to ceiling at all rates on El 7. Second, not only the patterns of results but also the overall level of performance was similar for unmodulated and modulated stimuli. Averaged across subjects and across runs for each subject, differences in performance between the two stimuli were within 1 percentage point for all rates, and the difference was not significant at any rate ($p > 0.05$). The comparison between AM-rate and pulse-rate discrimination for individual subjects revealed significant differences in only two subjects, M2 [$F(1,9)=6.53$, $p<0.05$] and M6 [$F(1,9)=26.58$, $p<0.005$]. Subject M2 showed slightly better performance (ranging from 6 to 9 percentage points) with equal-amplitude pulse train than with the AM stimuli. However, subject M6 showed consistently better AM-rate discrimination (ranging from 6 to 20 percentage points) than pulse-rate discrimination.

The similar patterns of results between equal-amplitude and AM pulse trains suggest that the observed variations in performance with baseline rate were not due to subjects latching onto a loudness-difference cue, such as might have been caused by imperfect loudness balancing at some rates. However, our results differ from those reported by Baumann and Nobbe (2004), who reported that AM-rate discrimination was significantly worse than pulse-rate discrimination ability at rates above 200 pps. This discrepancy could be due either to differences in procedure—for example, their use of an adaptive procedure, an odd-man-out task, and level rove—or to stimulus differences. In both studies the carrier was sinusoidally amplitude modulated on a linear scale between threshold and MCL, and stimuli were generated and controlled using the RIB software and hardware. However, as noted in Sec. III B, it is possible to greatly underestimate threshold using the RIB interface, due to an audible artifact associated with playing each stimulus. If this had not been controlled for in their study, then only the broad peaks of each modulator may have been audible, thereby resulting in a shallower envelope slope over the audible range of amplitudes.

IV. EXPERIMENT 3: LIMITS OF TEMPORAL PITCH AT DIFFERENT COCHLEAR LOCATIONS

A. Rationale

Two lines of evidence from NH listeners suggest that the processing of pulse rate may depend on cochlear location, and that the effects of location and rate may interact. First, Krumbholz and colleagues (Krumbholz *et al.*, 2000; Pressnitzer *et al.*, 2001) reported that the frequency of the lower limit of temporal pitch increased with the frequency region of the cochlea. Second, based on the evidence of the poor frequency discrimination performance with transposed tones presented in NH listeners (Oxenham *et al.*, 2004), it has been suggested that the poor rate discrimination performance with electric stimulation may be due to the mismatch between place and rate of stimulation. If so, then apical electrodes might show better rate discrimination at lower rates and the limits of temporal pitch would increase as the stimulating electrode progressively moved toward the basal locations.

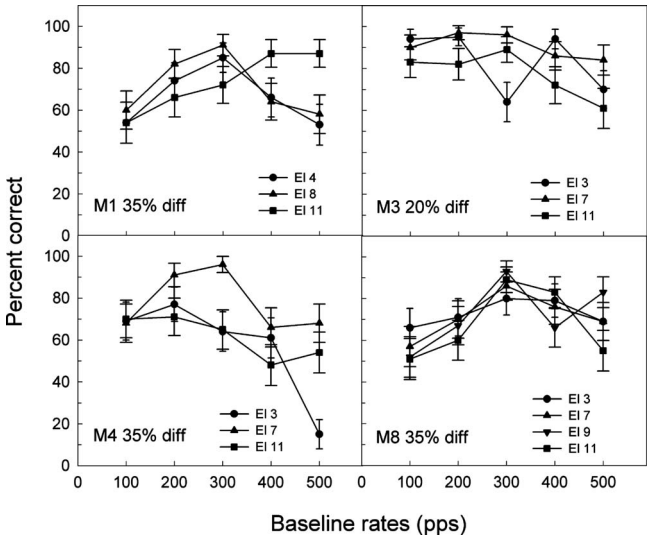


FIG. 5. Rate discrimination on different electrodes on four subjects. Note that the signal rates are 35% higher than the baseline rates for subjects M1, M4, and M8, but only 20% higher for subject M3. The smaller difference between baseline rate and the signal rate used on M3 was to avoid the ceiling effect. Vertical bars represent 95% confidence intervals of the mean.

B. Methods

Four Med-El implant users (M1, M3, M4, and M8) took part. In addition to the middle electrode tested in experiment 1, all subjects were subsequently tested with two or three electrodes: apical electrodes (El 3 or 4) were tested first followed by basal electrodes (El 9 and/or El 11). Stimuli and test procedure were the same as in experiment 1. For each subject, all test sessions for this experiment took place within a 1 month time period.

C. Results and discussion

Figure 5 shows rate discrimination results on all the electrodes tested for each subject. Rate discrimination results on the middle electrode (El 7 or 8) are replotted from Fig. 1 (experiment 1). In addition, results from subject M3 on electrode 3 and from subject M8 on electrode 9 were the same as those obtained in experiment 2. In general, subjects' ability to discriminate rate difference varied depending on the electrode stimulated. However, there was no consistent pattern of how the performance changed as the electrode position moved from the apical location to the basal location. For example, while subject M1 exhibited better rate discrimination at higher rates on the basal electrode than the apical and middle electrodes, subjects M3 and M4 showed the opposite pattern. The effect of pulse rate on performance also differed across electrodes for subjects M1, M3, and M4. For subject M1, the unusual peaky functions were obtained in El 4 and 8, but a high-pass characteristic monotonically increasing function was obtained in El 11. Subject M4, on the other hand,

showed a low-pass monotonically decreasing function in El 3 and 11, but the middle electrode showed a peak at 300 pps. It was noted that the performance at the highest rate in El 3 was significantly below the chance level of 50%, suggesting a possible pitch reversal at this rate for this electrode. Subject M8, however, did not show differences in performance across electrodes. When averaged across runs and subjects, a two-way repeated-measures ANOVA revealed a nonsignificant main effect of electrode [$F(2,6)=2.82$, $p>0.1$] and a nonsignificant two-way interaction between rate and electrode [$F(8,24)=1.41$, $p>0.1$].

The lack of a consistent electrode effect on the rate discrimination performance is similar to the findings of Baumann and Nobbe (2004) and Zeng (2002). This suggests that either the limitation in rate processing at higher rates does not result entirely from a mismatch between place and rate of stimulation, or that good rate discrimination at high rates requires a more accurate rate-place match than has been achieved so far. With the Med-El device, which provides a deeper insertion to the cochlea up to 30 mm from the base, one might expect that rate discrimination at lower rates would improve on the apical electrodes due to a better match between rate and place. However, our results did not support this hypothesis nor did they support the notion of an effect of cochlear region on the lower and the upper limit of temporal pitch.

V. GENERAL DISCUSSION

One of the major findings in our study is that most of the Med-El subjects showed an unexpected nonmonotonic pattern of rate discrimination performance, with the intermediate baseline rates (200–400 pps) producing better scores than the lowest (100 pps) and the highest (500 pps) baseline rates tested. Our follow-up studies and additional calibrations ruled out the possibility that this pattern of results is due to stimulus errors or to the use of a new test procedure. It was also noted that this pattern of results was not *entirely* specific to the Med-El subjects in our sample; although there was a device × rate interaction, three out of the eight CI24 subjects showed better performance at 200 pps than at 100 pps. There is also some previous evidence in the literature which showed superior rate discrimination performance as the baseline rate was increased above 100 pps. McKay *et al.* (1994) measured rate pitch difference limens (PDLs) on a group of Nucleus 22 subjects using a two-alternative forced-choice procedure without level rove. The PDL was defined as percentage rate change for which the two stimuli were judged different in pitch 75% of the time. One of their subjects, subject 6, showed 4%–5% PDL for baseline rates of 175 and 200 pps and 15% PDL for baseline rates of 100 and 150 pps. For this subject, the PDL was about 10 percentage points lower in the mid-rate range (175–200 pps) than in the low-rate range (100–150 pps). The poorer performance at 100 pps compared to 200 and 300 pps in our study suggests that for our CI listeners a lower limit of temporal pitch is at about 100 pps. Some of our Med-El subjects (M1, M4, M7, and M8) who exhibited very poor rate discrimination at 100 pps reported that 100 pps pulse trains sounded like a "frog croak," a description similar to the mixture of pitch and flutter reported by NH listeners at the lower limit of temporal pitch around 30 Hz (Krumbholz *et al.*, 2000). Still, it should be noted that the 30 Hz limit in NH listeners is lower than the lowest rate of 100 pps studied here. Despite the fact that the lower limit increases with increasing the frequency region of the cochlea in NH listeners, Krumbholz *et al.* (2000) argued that the lower limit of temporal pitch was determined by the characteristics of a temporal pitch mechanism, rather than by the effects of filtering in the peripheral auditory system. The fact that some of our CI subjects showed deterioration of rate discrimination performance at 100 pps across electrodes further supports a central mechanism on the lower limit of temporal pitch.

The relatively good performance of some of our Med-El subjects at high rates (≥300 pps) is somewhat surprising given the commonly reported upper limit of rate discrimination of about 300 pps. Despite the fact that most CI listeners have shown a lack of rate discrimination ability above 300 pps, exceptions have been reported previously. Two subjects in a study by Hochmair-Desoyer *et al.* (1983) and one of the five subjects (SR2) in a study by Wilson *et al.* (1997) showed an ability to perceive temporal pitch for rates up to 500–1000 Hz. In addition, one subject (subject 5) in the study of McKay *et al.* (1994) showed rate pitch DLs of about 2%, 8%, and 9% for baseline rates of 300, 400, and 500 pps, respectively. It is unclear why these subjects were able to perceive rate pitch at high rates. Our results are also consistent with the perception of rate pitch reported in NH listeners. Using alternating-phase harmonic complexes filtered into the frequency region between 7800 and 10800 Hz, Carlyon and Deeks (2002) reported that NH listeners could detect an approximately 27% differences in pulse rate relative to a 712-pps baseline rate. Hence, the upper limit of temporal pitch in the best CI subjects is roughly similar to that observed in NH listeners. Temporal pitch perception at high rates (up to 800–1000 Hz) is physiologically plausible in the case of electrical stimulation. The response of the auditory nerve to electrical stimulation is highly synchronized, at least in recently deafened animals, and firing occurs on every stimulation cycle up to rates of 800 Hz (van den Honert and Stypulkowski, 1984; Hartmann and Klinke, 1990; Javel, 1990), thus preserving periodicity information of the stimulus. Although it is still unclear why discrimination performance at high rates can vary substantially across subjects and electrodes, it is worth noting that the similarity in the pattern of results obtained with "regular" pulse trains and with amplitude-modulated high-rate pulse trains suggests that the upper limit of temporal pitch is unlikely to depend strongly on any particular pattern of auditory-nerve response. Furthermore, the fact that the pattern of results was similar for each subject for the two stimulus types, even though that pattern differed significantly across subjects and electrodes, suggests that limits on temporal pitch perception depend more strongly on the particular subject and electrode tested than on the details of the stimulus used.

Our findings on temporal pitch perception in CI listeners have potential implications for CI speech processing design. The current assumption is that since CI users cannot follow

rate change above 300 pps, they will not benefit from a processing strategy that encodes temporal fine structure of a signal. First, unlike the results in the majority reports which tested patients with the Nucleus devices, our results showed that some CI listeners have relatively good temporal pitch perception with rate DL $<20\%$ for rates up to 300 pps, and some listeners even detect rate changes at higher rates above 300 pps. Second, the finding referred to above, in which the pattern of results depends more on the subject than on the stimulus, suggests that an approach in which signal-processing strategies are matched to the temporal processing capacity of each patient may be more fruitful than attempts to modify the stimulus in order to improve temporal processing. The better rate discrimination at high rates exhibited by some subjects might possibly translate into better speech recognition performance in competing backgrounds if these subjects can use temporal cues to better perceive differences in voice pitch. A useful topic for future research is whether good rate discrimination at high rates is a predictor of good speech recognition in noise, once the effects of potential covariants such as duration of deafness have been excluded.

ACKNOWLEDGMENTS

We are grateful to all subjects for their participation in these experiments. We thank Dr. Hedwig Gockel and Professor Colette McKay for useful discussions and Ms. Cathy Lynch for assistance in subject recruitment. We also thank Dr. Richard Freyman, Dr. Bob Shannon, and an anonymous reviewer for their many helpful suggestions. This work was supported by the Otology Research Fund.

Baumann, U., and Nobbe, A. (**2004**). "Pulse rate discrimination with deeply inserted electrode arrays," Hear. Res. **196**, 49–57.

Braida, L. D., Lim, J. S., Berliner, J. E., Durlach, N. I., Rabinowitz, W. M., and Purks, S. R. (**1984**). "Intensity perception. XIII. Perceptual anchor model of context-coding," J. Acoust. Soc. Am. **76**, 722–731.

Burns, E. M., and Viemeister, N. F. (**1976**). "Nonspectral pitch," J. Acoust. Soc. Am. **60**, 863–869.

Burns, E. M., and Viemeister, N. F. (**1981**). "Played-again SAM: Further observations on the pitch of amplitude-modulated noise," J. Acoust. Soc. Am. **70**, 1655–1660.

Carlyon, R. P., and Deeks, J. M. (**2002**). "Limitations on rate discrimination," J. Acoust. Soc. Am. **112**, 1009–1025.

Fechner, G. (**1966**). *Elements of Psychophysics*, translated by H. E. Adler and edited by D. H. Howes and E. G. Boring (Holt, Rinehart, and Winston, New York).

Gfeller, K., and Lansing, C. R. (**1991**). "Melodic, rhythmic, and timbral perception of adult cochlear implant users," J. Speech Hear. Res. **34**, 916–920.

Hartmann, R., and Klinke, R. (**1990**). "Response characteristics of nerve fibers to patterned electrical stimulation," in *Cochlear Implants: Models of the Electrically Stimulated Ear*, edited by J. M. Miller and F. A. Spelman (Springer-Verlag, New York), pp. 135–159.

Hochmair-Desoyer, I. J., Hochmair, E. S., Burian, K., and Stiglbrunner, H. K. (**1983**). "Percepts from the Vienna cochlear prosthesis," Ann. N.Y. Acad. Sci. **405**, 295–306.

Javel, E. (**1990**). "Acoustic and electrical encoding of temporal information," in *Cochlear Implants: Models of the Electrically Stimulated Ear*, edited by J. M. Miller and F. A. Spelman (Springer-Verlag, New York), pp. 247–295.

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

Kong, Y.-Y., Cruz, R., Jones, J. A., and Zeng, F.-G. (**2004**). "Music perception with temporal cues in acoustic and electric hearing," Ear Hear. **25**, 173–185.

Krumbholz, K., Patterson, R. D., and Pressnitzer, D. (**2000**). "The lower limit of pitch as determined by rate discrimination," J. Acoust. Soc. Am. **108**, 1170–1180.

Landsberger, D. M., and McKay, C. M. (**2005**). "Perceptual differences between low and high rates of stimulation on single electrodes for cochlear implantees," J. Acoust. Soc. Am. **117**, 319–327.

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

McDermott, H. J. (**2004**). "Music perception with cochlear implants: A review," Trends Amplif. **8**, 49–82.

McDermott, H. J., and McKay, C. M. (**1997**). "Musical pitch perception with electrical stimulation of the cochlea," J. Acoust. Soc. Am. **101**, 1622–1631.

McKay, C. M., McDermott, H. J., and Clark, G. M. (**1994**). "Pitch percepts associated with amplitude-modulated current pulse trains in cochlear implantees," J. Acoust. Soc. Am. **96**, 2664–2673.

Moore, B. C. J., and Carlyon, R. P. (**2005**). "Perception of pitch by people with cochlear hearing loss and by cochlear implant users," in *Pitch: Neural Coding and Perception*, edited by C. J. Plack and A. J. Oxenham (Springer, New York), pp. 234–277.

Oxenham, A. J., Bernstein, J. G., and Penagos, H. (**2004**). "Correct tonotopic representation is necessary for complex pitch perception," Proc. Natl. Acad. Sci. U.S.A. **101**, 1421–1425.

Pijl, S. (**1997**). "Labeling of musical interval size by cochlear implant patients and normally hearing subjects," Ear Hear. **18**, 364–372.

Pijl, S., and Schwarz, D. W. (**1995**). "Melody recognition and musical interval perception by deaf subjects stimulated with electrical pulse trains through single cochlear implant electrodes," J. Acoust. Soc. Am. **98**, 886–895.

Pressnitzer, D., Patterson, R. D., and Krumbholz, K. (**2001**). "The lower limit of melodic pitch," J. Acoust. Soc. Am. **109**, 2074–2084.

Scharf, B., Quigley, S., Aoki, C., Peachey, N., and Reeves, A. (**1987**). "Focused auditory attention and frequency selectivity," Percept. Psychophys. **42**, 215–223.

Shannon, R. V. (**1983**). "Multichannel electrical stimulation of the auditory nerve in man: I. Basic psychophysics," Hear. Res. **11**, 157–189.

Stickney, G. S., Assmann, P. F., Chang, J., and Zeng, F.-G. (**2007**). "Effects of cochlear implant processing and fundamental frequency on the intelligibility of competing sentences," J. Acoust. Soc. Am. **122**, 1069–1078.

Stickney, G. S., Zeng, F. G., Litovsky, R., and Assmann, P. F. (**2004**). "Cochlear implant speech recognition with speech maskers," J. Acoust. Soc. Am. **116**, 1081–1091.

Tong, Y. C., Blamey, P. J., Dowell, R. C., and Clark, G. M. (**1983**). "Psychophysical studies evaluating the feasibility of a speech processing strategy for a multiple-channel cochlear implant," J. Acoust. Soc. Am. **74**, 73–80.

Townshend, B., Cotter, N., van Compernolle, D., and White, R. L. (**1987**). "Pitch perception by cochlear implant subjects," J. Acoust. Soc. Am. **82**, 106–115.

van den Honert, C., and Stypulkowski, P. H. (**1984**). "Physiological properties of the electrically stimulated auditory nerve. II. Single fiber recordings," Hear. Res. **14**, 225–243.

van Hoesel, R. J., and Clark, G. M. (**1997**). "Psychophysical studies with two binaural cochlear implant subjects," J. Acoust. Soc. Am. **102**, 495–507.

Vandali, A. E., Sucher, C., Tsang, D. J., McKay, C. M., Chew, J. W., and McDermott, H. J. (**2005**). "Pitch ranking ability of cochlear implant recipients: a comparison of sound-processing strategies," J. Acoust. Soc. Am. **117**, 3126–3138.

Wilson, B. S., Sun, X., Schatzer, R., and Wolford, R. D. (**2004**). "Representation of fine structure or fine frequency information with cochlear implants," International Congress Series **1273**, 3–6.

Wilson, B. S., Zerbi, M., Finley, C. C., Lawson, D. T., and van den Honert, C. (**1997**). "Speech processors for auditory prostheses: Relationships between temporal patterns of nerve activity and pitch judgments for cochlear implant patients," 8th Quarterly Progress Report, NIH Project N01-DC-5-2103.

Zeng, F.-G. (**2002**). "Temporal pitch in electric hearing," Hear. Res. **174**, 101–106.

Low-frequency speech cues and simulated electric-acoustic hearing

Christopher A. Brown and Sid P. Bacon
Department of Speech and Hearing Science, Psychoacoustics Laboratory, Arizona State University, P.O. Box 870102, Tempe, Arizona 85287-0102

(Received 26 December 2007; revised 8 December 2008; accepted 9 December 2008)

The addition of low-frequency acoustic information to real or simulated electric stimulation (so-called electric-acoustic stimulation or EAS) often results in large improvements in intelligibility, particularly in competing backgrounds. This may reflect the availability of fundamental frequency (F0) information in the acoustic region. The contributions of F0 and the amplitude envelope (as well as voicing) of speech to simulated EAS was examined by replacing the low-frequency speech with a tone that was modulated in frequency to track the F0 of the speech, in amplitude with the envelope of the low-frequency speech, or both. A four-channel vocoder simulated electric hearing. Significant benefit over vocoder alone was observed with the addition of a tone carrying F0 or envelope cues, and both cues combined typically provided significantly more benefit than either alone. The intelligibility improvement over vocoder was between 24 and 57 percentage points, and was unaffected by the presence of a tone carrying these cues from a background talker. These results confirm the importance of the F0 of target speech for EAS (in simulation). They indicate that significant benefit can be provided by a tone carrying F0 and amplitude envelope cues. The results support a glimpsing account of EAS and argue against segregation.
 [DOI: 10.1121/1.3068441]

PACS number(s): 43.66.Ts, 43.71.Ky [BCM] Pages: 1658–1665

I. INTRODUCTION

Recently, individuals with residual hearing restricted to the low frequencies (below about 500–750 Hz) have been implanted with a relatively short electrode array designed to preserve as much of the residual hearing as possible in the apical region (Gantz *et al.*, 2005; Gantz and Turner, 2003, 2004; Turner *et al.*, 2004; von Ilberg *et al.*, 1999). These individuals, in addition to full-insertion implant users who have some residual hearing in the nonimplanted ear, have the potential to combine the electric and acoustic sources of information. For both simulated and real implant processing, the addition of low-frequency acoustic stimulation often enhances speech understanding, particularly when listening to speech in the presence of competing speech (Dorman *et al.*, 2005; Kong *et al.*, 2005; Turner *et al.*, 2004). The benefit of this so-called electric-acoustic stimulation (EAS) occurs even when the acoustic stimulation alone provides little or no intelligibility (i.e., no words correctly identified).

Although little is known about the auditory processing or the acoustic cues underlying this effect, some (e.g., Chang *et al.*, 2006; Kong *et al.*, 2005; Qin and Oxenham, 2006) have suggested that listeners combine the relatively weak pitch information conveyed by the electric stimulation with the stronger pitch cue from the target talker's fundamental frequency (F0) or voice pitch in the low-frequency acoustic region to segregate target and background. It has been thought for some time that F0 aids in the segregation of competing talkers (e.g., Assmann, 1999; Assmann and Summerfield, 1990; Bird and Darwin, 1997; Brokx and Nooteboom, 1982; Culling and Darwin, 1993). Recent reports (Chang *et al.*, 2006; Qin and Oxenham, 2006) have shown indirectly that F0 is likely to play an important role independent of any role that the first formant may play. For example, the addition to vocoder stimulation of 300 Hz low-pass speech, which itself should not contain much if any first formant information (Hillenbrand *et al.*, 1995) or yield any intelligibility, improved speech intelligibility in a competing background (Chang *et al.*, 2006; Qin and Oxenham, 2006).

However, the question remains of what low-frequency cues are responsible for the EAS effect. Kong and Carlyon (2007) simulated EAS conditions, and found that voicing and amplitude envelope information provided benefit over vocoder alone. On the other hand, F0 cues provided no additional benefit at any SNR tested. They argued against F0 as a cue for segregation, and suggested that, in addition to the voicing cue, the amplitude envelope may help listeners by indicating when to listen or "glimpse" the target.

While several papers have suggested F0 as a cue for EAS, the supporting evidence has been relatively circumstantial. The primary goal of the present study was to evaluate directly the importance of F0 for EAS. A secondary goal was to evaluate the importance of the amplitude envelope of the acoustic stimulus in the EAS effect, as well as the importance of combining F0 and the amplitude envelope. To do this, we replaced the target speech in the low-frequency region with a tone that was modulated either in frequency to track the dynamic changes in the target talker's F0, in amplitude with the amplitude envelope of the low-pass target speech, or both in frequency and amplitude. There is evidence (Faulkner *et al.*, 1992) that this kind of processing can provide an aid to lip reading for hearing-impaired listeners, particularly those with limited frequency selectivity, though it is unclear whether it can yield an EAS benefit.

 0001-4966/2009/125(3)/1658/8/$25.00

In addition to the theoretical importance of determining the contribution to intelligibility by F0, there may be practical benefits as well: if EAS benefit can be demonstrated with a low-frequency tone carrying F0 (and/or the amplitude envelope), it is possible that hearing-impaired listeners with especially elevated low-frequency thresholds could benefit more from the tonal cue than from speech itself because the entire cue could be made audible due to the concentration of all the energy into a narrow frequency region, whereas only a portion of the broader band speech might be.

II. EXPERIMENT 1

Experiment 1 examined the contribution of the dynamic changes in F0 to the benefit in intelligibility from simulated EAS by replacing the low-pass speech with a tone that was modulated in frequency to track the changes in F0 that occur across an utterance. Because we expected the amplitude envelope of the low-pass speech to contribute to intelligibility as well, we included conditions in which a tone equal in frequency to the mean F0 of the target talker was modulated by the amplitude envelope of the low-pass target speech. An additional set of conditions combined the F0 and the envelope cues.

A. Method

1. Subjects

Data were collected from 25 (15 females, 10 males) fluent speakers of English, who ranged in age from 26 to 38 years and who were compensated either monetarily or with course credit for their time. All 25 listeners had pure-tone air-conduction thresholds ≤20 dB HL (ANSI, 1996) at octave and half-octave frequencies from 250 to 6000 Hz in the right ear, which was used exclusively.

2. Stimuli

Prior to testing, the dynamic changes in the target talker's F0 were extracted from each sentence using the YIN algorithm (de Cheveigné and Kawahara, 2002) with a 40 ms window size and 10 ms step size. In addition, the onsets and offsets of voicing in each utterance were extracted manually, with 10 ms raised-cosine ramps applied to the transitions.

Target stimuli consisted of the IEEE sentences (IEEE, 1969) produced by a female talker with a mean F0 of 184 Hz. Backgrounds were the AZBIO sentences (Spahr and Dorman, 2004) produced by a male (mean F0=92 Hz) or a female (mean F0=224 Hz) talker, four-talker babble (Auditec, 1997), or generic speech-shaped noise (low passed at 800 Hz, using a first-order Butterworth filter). The target speech began 150 ms after the onset of the background and ended 150 ms before the background offset. Two single-talker background sentences were concatenated when necessary. Prior to processing, the level of the broadband target speech was adjusted to 70 dB SPL and the rms level of the background stimuli was adjusted to achieve a +10 dB SNR, which was shown in pilot testing to produce about 30% correct in vocoder-only test conditions. This allowed sufficient room for improvement when a low-frequency cue was added.

The information conveyed by electric stimulation was simulated using a four-channel vocoder that employed sinusoidal carriers. The signal was bandpass filtered into four frequency bands. The logarithmically spaced cutoff frequencies of the contiguous vocoder bands were 750, 1234, 2031, 3342, and 5500 Hz. The envelope of each band was extracted by half-wave rectification and low-pass filtering (sixth-order Butterworth, cutoff frequency of 400 Hz or half the bandwidth, whichever was less). This envelope was used to modulate the amplitude of a tone at the arithmetic center of the band (the frequencies of the carrier tones were 992, 1633, 2687, and 4421 Hz). This thus simulates a 20 mm insertion depth, appropriate for "hybrid" EAS in which the electric and acoustic stimulation occur in the same ear.

The low-frequency region consisted of either target speech low-pass filtered at 500 Hz (tenth-order Butterworth) or a tone whose mean frequency equaled the mean F0 of the target talker for each sentence (overall mean F0=184 Hz). The tone was unmodulated (except for the modulation due to the onsets and offsets of voicing; this voicing cue was present in all of the tone conditions) or modulated either in frequency with the dynamic F0 changes in each target utterance, in amplitude with the envelope of the 500 Hz low-pass speech [obtained via half-wave rectification and low-pass filtering at 16 Hz (second-order Butterworth)] or both in frequency and amplitude. In all cases, the tone was audible only when voicing occurred, and the level of the tone was adjusted to be equal in rms to that of the 500 Hz low-pass speech. Note that, as was the case for the study by Kong and Carlyon (2007), the background was never present in the low-frequency region. This was done because it allowed a more sensitive measure of the contributions of each low-frequency cue of interest.

All processing was done digitally via software routines in MATLAB, and stimuli were presented monaurally using an Echo Gina 3G sound card (16 bit precision, 44.1 kHz sampling rate), Tucker Davis PA5 attenuators, and Sennheiser HD250 headphones.

3. Conditions

The output of the four-channel vocoder (target plus background) was either presented alone (V), combined with the 500 Hz low-pass target speech ($V/500$) or combined with a tone that was either unmodulated (except for voicing; V/T), modulated in frequency by the dynamic change in F0 (V/T_{F0}), modulated in amplitude by the envelope of the low-pass speech (V/T_{env}), or modulated in both frequency and amplitude ($V/T_{F0\text{-}env}$). In addition, the 500 Hz low-pass target speech and each of the tonal cues were presented in isolation without the vocoder stimulation.

4. Procedure

Participants were seated in a double-walled sound booth with one of two experimenters, who scored responses and controlled stimulus presentation. One experimenter was aware of the experimental details of each condition as it was presented, and one was not.[1] Responses were made verbally,

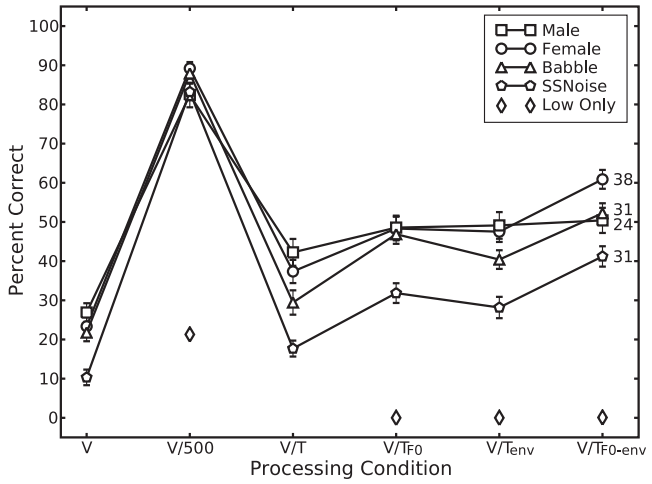


FIG. 1. Mean percent correct scores, ±1 standard error. Each curve represents a different background, which was present in the vocoder region only. Processing conditions are along the *x* axis. The output of the vocoder was presented alone (V) or with different low-frequency cues, which included 500 Hz low-pass speech ($V/500$), a tone with only voicing applied (V/T), a tone modulated in frequency by the fundamental of the target talker (V/T_{F0}), a tone modulated in amplitude by the envelope of the low-pass speech (V/T_{env}), and a tone modulated in both frequency and amplitude ($V/T_{F0\text{-}env}$). Mean intelligibility provided by the low-frequency cues themselves is depicted by diamonds. Values to the right of the $V/T_{F0\text{-}env}$ data points are percentage points of improvement over vocoder alone. Although the processing variable, which is depicted along the *x* axis, is not a continuous variable, the different levels of processing within each background are connected with lines for clarity.

and participants were instructed to repeat as much of the target sentence as they could. No feedback was provided.

Participants first heard ten unprocessed broadband target sentences presented in quiet, followed by ten sentences in quiet processed with the four-channel vocoder, to familiarize them with the target talker's voice and with the vocoder processing. Participants then heard 100 sentences of the target talker at a SNR of +10 (babble background) processed through the four-channel vocoder, combined with the low-frequency tone modulated in both frequency and amplitude ($V/T_{F0\text{-}env}$).[2]

There were 50 keywords (ten sentences) per test condition, and the presentation order of the conditions was randomized for each subject. No sentence was heard more than once.

B. Results

Figure 1 shows the mean percent correct results. Each curve represents performance in a different background, and error bars represent ±1 standard error. The different processing conditions are represented along the *x* axis. Diamonds represent performance when the respective low-frequency cue was presented alone.

A two-factor repeated-measures analysis of variance, with background and processing conditions as the main effects, revealed significant differences ($p<0.001$) within each variable. A *post hoc* Tukey analysis using a Holm–Bonferroni correction on the background variable showed all pairwise differences to be significant except male and female, and male and babble. A Tukey analysis was conducted on the different processing conditions as well; significant differences (adjusted $p<0.001$) were found between each pair of groups, except V/T_{F0} and V/T_{env}. These effects are described in more detail below.

The improvement in performance observed from the V (vocoder-only) conditions to the $V/500$ (vocoder plus 500 Hz low-pass target speech) conditions was on average, about 65 percentage points. This improvement demonstrates, in simulation, the EAS effect of combining vocoder stimulation with low-frequency speech, which itself provided only about 20% intelligibility (diamond marker at $V/500$). The improvement in performance over V in the V/T (vocoder plus tone carrying the voicing cue only) conditions averaged about 11 percentage points across backgrounds. This effect indicates that the voicing cue is informative under these conditions. The effect of the dynamic changes in $F0$ on intelligibility—above and beyond the effects of voicing—can be seen by comparing scores for V/T_{F0} with those for V/T. Across backgrounds, the improvement averaged about 13 percentage points. Similarly, the addition of a tone modulated with the envelope of the low-pass target speech to the vocoder (V/T_{env}) produced about 11 percentage points of improvement relative to V/T. Both of these differences were statistically significant ($p<0.001$). Finally, when the tone was modulated in both frequency and amplitude and combined with vocoder ($V/T_{F0\text{-}env}$), improvement over V/T averaged about 20 percentage points. Note that the tonal cues by themselves were not sufficient for any words to be reported correctly (diamond markers at V/T_{F0}, V/T_{env}, and $V/T_{F0\text{-}env}$).

For three of the four backgrounds (female talker, male talker, and speech-shaped noise), the contributions to intelligibility of $F0$ and the amplitude envelope of low-pass speech were statistically equivalent (adjusted $p>0.42$), and each cue was statistically more effective than the voicing cue alone. In addition, in these three backgrounds, the combination of F0 and amplitude envelope cues provided significant benefit over the amplitude envelope cue alone (adjusted $p<0.01$).

The amount of improvement in the $V/T_{F0\text{-}env}$ condition relative to the V condition for each background is given at the far right in Fig. 1. The largest improvement was seen in the female background (38 percentage points; circles), while the amount of improvement was 24 percentage points in the male background (squares), and about 31 percentage points in both the multitalker babble (triangles) and speech-shaped noise backgrounds (pentagons).

When the background was a male talker, nearly all of the benefit over vocoder only provided in the tone conditions was due to the voicing cue. Neither F0, nor the amplitude envelope, nor the combination of the two cues provided significantly more benefit than voicing alone (recall that the voicing cue was present in all of the tone conditions). It is unclear why this pattern of results was obtained only when the background was male.

C. Discussion

For three of the four backgrounds (female talker, babble, and speech-shaped noise), the pattern of results was similar:

F0 and the amplitude envelope of low-pass speech contributed equal, and somewhat independent and additive sources of information. In addition, each cue alone provided significant benefit over the voicing cue (recall that voicing was present in all tonal conditions), demonstrating that both F0 and the amplitude envelope are useful cues in simulated EAS conditions. These findings contrast with a recent finding (Kong and Carlyon, 2007) that showed no additional improvement in intelligibility due to the dynamic changes in F0 over the combined voicing and amplitude envelope cues.

On the other hand, performance with the male background showed a pattern of results that is more similar to that found by Kong and Carlyon (2007), in that nearly all of the improvement observed in the tone conditions could be attributed to the voicing cue, and F0 contributed no further benefit. It is unclear why this pattern of results is observed only for the male background in the current experiment and thus why the pattern of results here is generally different from that in Kong and Carlyon (2007). There are various procedural differences (e.g., sentence materials, number of vocoder channels, and carrier type in the low-frequency region) that may have contributed to the different pattern of results in the two studies. This is the focus of follow-up experiments currently under way.

1. Effects of background

Our results with speech-shaped noise are also inconsistent with those reported in the literature. Turner *et al.* (2004) showed an EAS benefit to speech intelligibility when low-frequency speech was combined with both real and simulated electric stimulations when the background was a competing talker, but not speech-shaped noise. We show a benefit under our simulated EAS conditions (compare *V* with *V*/500) with both types of backgrounds. This discrepancy may be due to our decision not to include the background in the low-frequency region.

Previous studies (e.g., Stickney *et al.*, 2004) have reported that a competing talker produces poorer speech intelligibility than speech-shaped noise in vocoder-alone processing. We did not obtain similar results; however, in that speech-shaped noise produced the most masking (compare different backgrounds in the V processing condition). Our results can be explained by the generic nature of our speech-shaped noise. Because it was simply low-pass filtered at 800 Hz with a first-order Butterworth filter, it had more energy in the frequency range encompassed by our vocoder (750–5500 Hz) than the speech backgrounds we used, and thus was a more effective masker.

2. Explaining the benefits of EAS

Our results show that the additional low-frequency voicing, amplitude envelope, and F0 cues can more or less independently contribute to speech intelligibility in simulated EAS. F0 has been thought for some time to aid in segregating competing talkers (e.g., Assmann, 1999; Assmann and Summerfield, 1990; Bird and Darwin, 1997; Brokx and Nooteboom, 1982; Culling and Darwin, 1993). Indeed, several recent reports (e.g., Chang *et al.*, 2006; Qin and Oxenham, 2006; Kong *et al.*, 2005) have suggested that F0 may aid in segregation of target and background in simulated EAS. However, the current results provide indirect evidence that indicate that segregation is not responsible for the benefits observed. For example, if the benefit due to F0 is explained by segregation, we might expect that the conditions which contained the greatest F0 difference between target and masker would yield the greatest benefit when target F0 information is added. However, the F0 difference between target and background was greatest when the background was male (F0 difference of 92 Hz), yet the addition of target F0 information had the least benefit with this background. Thus, the results of experiment 1 are not consistent with the segregation as an explanation for the benefits of F0 under EAS conditions.

There are other ways F0 may provide benefit as well. F0 has been shown to be important for several linguistic cues, including consonant voicing (Boothroyd *et al.*, 1988; Holt *et al.*, 2001), lexical boundaries (Spitzer *et al.*, 2007), and contextual emphasis (Fry, 1955) as well as manner (Faulkner and Rosen, 1999). It is unclear, however, to what extent any of these linguistic cues may have contributed to the simulated EAS effects observed here.

The effects of both amplitude envelope and voicing (which is a significant component of the amplitude envelope) are not surprising. In general, it is plausible that envelope information in the low-frequency region from either the target or the masker can improve speech intelligibility by providing an indication of when to listen, even at moments when the overall SNR is poor, since at any given moment a relatively favorable SNR is more likely when either the target level is relatively high or the masker level is relatively low (recall that in the present experiment, the low-frequency region did not contain the masker). This glimpsing cue was suggested by Kong and Carlyon (2007) as a possible explanation for the benefit observed in EAS.

Glimpsing might also at least partly explain the effects of F0. That is, F0 might provide an indication of when to listen, much like that provided by the amplitude envelope (Kong and Carlyon, 2007). If F0 and envelope cues both indicate a favorable time to listen, the two cues should be correlated. We evaluated this by comparing the fluctuations in amplitude with those in F0 across ten utterances. Only the voiced segments of each sentence were used, and the F0 track was equated in rms with the envelope track. The lowest *r* value obtained was 0.29, while the highest was 0.76, and the mean *r* was 0.52. In general, during the voiced portioned of the sentences we examined, as the amplitude envelope increases, so does F0. This analysis thus suggests that increases in F0 may indicate moments during an utterance in which favorable SNRs are more likely. The fact that they are not more highly correlated is consistent with the finding that F0 and the amplitude envelope appear to provide at least somewhat independent and additive benefit to vocoder alone.

We have clearly demonstrated a benefit in intelligibility due to the presence of F0 in simulated EAS, at least in three of four backgrounds tested. However, it is important to note that as with Kong and Carlyon (2007), the background was never present in the low-frequency region in any of these

conditions. While not ecologically valid, we chose this design because it allowed us a more sensitive measure of the contributions of the cues of interest. However, it is important to determine the effect that the background will have on the improvement observed due to the tone. Experiment 2 addressed this issue by examining the improvement due to the target tone in the low-frequency region with and without the presence of a background tone.

III. EXPERIMENT 2

Experiment 2 was designed to examine the effects of having background information in the low-frequency region. We included conditions in which vocoder stimulation was combined with a tone modulated with F0 and amplitude envelope information from the target talker, a tone carrying these cues from the background talker, or two tones, one carrying the information from the target and one from the background. This allowed us to compare directly the effects of the cues from each source. In addition, conditions in which the vocoder was combined with target and background low-pass speech were included as well, to allow for direct comparison. Thus, we replicated and extended the conditions from Experiment 1 to include conditions containing the low-frequency stimulus representing the background (either low-pass background speech or a tone tracking the F0 and amplitude envelope of the background speech).

A. Method

1. Subjects

Twelve normal-hearing listeners (11 females, 1 male) ranging in age from 26 to 38 years were paid an hourly wage for their services. The criteria for inclusion were identical to those used for Experiment 1, although a different group of listeners was recruited.

2. Stimuli

The processing and hardware were identical to those used in experiment 1. Prior to conducting experiment 1, we had extracted F0 and voicing data from a male (mean F0 =127) production of the CUNY sentences (Boothroyd *et al.*, 1985), as well as from male (mean F0=90) and female (mean F0=184; the target talker in experiment 1) productions of the IEEE sentences for earlier pilot experiments. Because the design of experiment 2 called for F0 data from the background as well as the target, we switched to CUNY target sentences because we were then able to use the two sets of IEEE sentences as background. A consequence of our switch to the CUNY sentence set is that we were able to extend our results from experiment 1 to high-context target materials.

3. Conditions

A target sentence was combined with a background (male or female talker) and processed through a four-channel vocoder (see experiment 1). The output of the vocoder was presented either alone (V) or with a low-frequency cue. In three conditions, the low-frequency cue consisted of 500 Hz low-pass speech, and was either target [or signal (S)] speech alone ($V/S500$),[3] background (B) speech alone ($V/B500$), or both target and background speech ($V/SB500$). Three other conditions combined the vocoder output with either a tone that was modulated in frequency with the target talker's F0 and amplitude with the envelope of the 500 Hz low-pass target speech ($V/S_{\text{F0-env}}$), a tone modulated in the same way using the background's F0 and low-pass envelope ($V/B_{\text{F0-env}}$), or both tones combined ($V/SB_{\text{F0-env}}$).

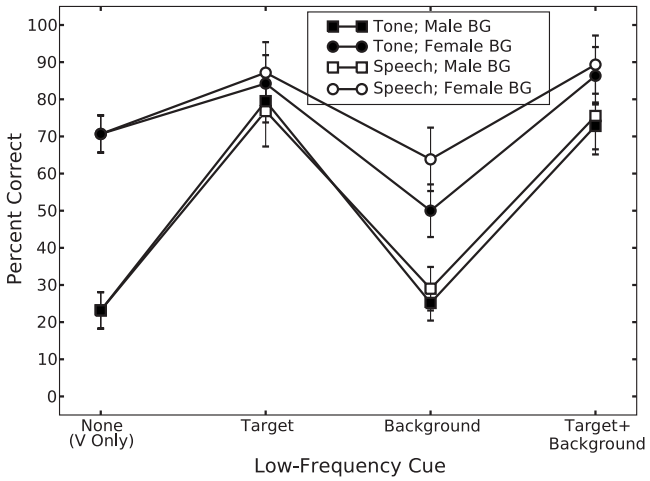


FIG. 2. Mean percent correct scores, ±1 standard error. Squares and circles indicate performance when the background was a male and a female, respectively. Unfilled and filled symbols indicate performance when the low-frequency cue was speech and a modulated tone, respectively. Different low-frequency stimulus conditions are represented along the x axis, and are no low-frequency stimulus (none), target only (target), background only (background), or target and background (target+background). Although the low-frequency stimulus variable, which is depicted along the x axis, is not a continuous variable, the different levels of this variable within each background gender/low-frequency cue combination are connected with lines for clarity.

4. Procedure

The procedure, including the pretest exposure to the vocoder plus target tone condition, was identical to that used in Experiment 1.

B. Results

Figure 2 shows the mean percent correct results. Circles represent performance with a female background and squares represent performance with a male background. Unfilled symbols represent performance when speech was present in the low-frequency region and filled symbols represent performance when a tone or tones were present. The different target/background combinations presented in the low-frequency region are represented along the x axis. There was either no stimulus (vocoder only; none), target only (target), background only (background), or both target and background (target+background). Error bars represent ±1 standard error.

A three-factor repeated-measures analysis of variance was conducted, with background gender (male or female), low-frequency stimulus (none, target, background, or target +background), and low-frequency processing (tone or

speech) conditions as the main effects. There were significant differences within the stimulus and gender variables ($p < 0.001$). Neither the three-way interaction nor any of the two-way interactions were statistically significant (adjusted $p > 0.42$), except the interaction between gender and low-frequency stimulus ($p = 0.001$). This significant interaction is likely due to the differences in performance observed in the vocoder-only conditions (none), which showed the male background to be a more effective masker than the female background. The processing variable was not significant ($p = 0.22$). The lack of a significant difference for processing indicates that a tone conveying both the F0 and the envelope of the target speech provided as much benefit as the low-pass speech itself, whether or not the background was present.

A *post hoc* Tukey analysis was conducted on group means. Within the male background (squares), the presence of the target in the low-frequency region (speech or tone) provided about 55 percentage points of improvement over vocoder alone, regardless of whether the background was present (target+background) or not (target). This improvement was statistically significant (adjusted $p < 0.001$). On the other hand, the low-frequency cue (speech or tone) had no statistically significant effect on intelligibility, regardless of whether or not the target was present (target+background) or not (background) (adjusted $p > 0.99$). Within the female background (circles), there were no statistically significant changes in intelligibility due to either the target or the background, whether the low-frequency stimulus was speech or tones. This effect was likely due to the overall high performance with the female background, which may not have allowed enough room to observe the improvements due to the low-frequency cues.

C. Discussion

The results of Experiment 2 indicate that the EAS benefits of target F0 and amplitude envelope information are not effected by the presence of background F0 and amplitude envelope information. This can be seen by comparing performance in "target" with performance in "target+background." Indeed, the presence of the background in the low-frequency region had no statistically significant effect on performance in any of the comparisons we examined. In other words, performance was equivalent whether the background was present or not. Note, however, the female-background tone condition in which performance declined by about 20 percentage points from vocoder only performance (compare filled circles in "none" and "background").

1. Effects of gender

When the background was a female talker (circles), the amount of improvement due to the low-frequency target stimulus (either speech or tone) was about 15 percentage points. With the male background (squares), which was a much more effective masker in the vocoder-only condition, the improvement in performance when a target stimulus was present in the low-frequency region was about 55 percentage points. At first glance, the difference in masking effectiveness between the female and male backgrounds seems surprising, given that several earlier studies have reported a lack of such an effect for CI patients and normal-hearing subjects listening in simulation (e.g., Qin and Oxenham, 2003; Stickney *et al.*, 2004). This apparent discrepancy may be explained by our use of a sinusoidal vocoder, as most of the studies that report no differences in masking based on gender have used noise-excited vocoders. Consistent with this explanation are recent results from Cullington and Zeng (2008). They used a sinusoidal vocoder and found that speech reception thresholds of a male talker were about 10 dB better in the presence of a female background than in the presence of a male background using a sinusoidal vocoder.

The background gender effect also contrasts with the results from experiment 1, which showed about the same amounts of masking for the male and female backgrounds. This discrepancy is more difficult to explain. F0 separation cannot account for the effects observed; the F0 difference between the target and the male and female backgrounds from experiment 1 are about 102 and 40 Hz, respectively, while for experiment 2 they are 37 and 57 Hz, respectively. Thus, if F0 separation were a critical factor, one would expect a larger background gender effect in experiment 1, but this outcome was not observed. We have conducted a brief pilot experiment[4] to address this issue, which demonstrates that with a sinusoidal vocoder, the ability of one talker to mask another may not be accurately predicted by gender or F0 difference. More work is clearly needed, and we are developing experiments to more fully characterize the relationship between gender, F0 separation, and intelligibility under vocoded conditions.

2. Effects of sentence materials

We have demonstrated that a tone carrying the F0 and amplitude envelope cues from the target can provide significant benefit in simulated EAS, and that the effects of a tone carrying these same cues from a background talker were not significant. In addition to the lack of an effect of low-frequency background cues, experiment 2 provided several other interesting results. First, the amount of benefit due to the tone in experiment 1 was between 24 and 38 percentage points, while in experiment 2 the amount of benefit was as much as 57 percentage points. Second, a tone carrying F0 and the low-frequency amplitude envelope of the target talker provided as much benefit as the low-pass target speech (compare open and filled symbols in the same low-frequency stimulus condition in Fig. 2).

One possible explanation for the difference across experiments 1 and 2 in the amounts of benefit provided by the tone may be related to the amount of context in the target sentences. The CUNY sentences (target materials in experiment 2) are considered to have high context, whereas the IEEE sentences (target materials in experiment 1) are considered to have low context (Duchnowski *et al.*, 2000; Grant and Walden, 1995). It has been shown (e.g., Bell *et al.*, 1992) that the use of high-context sentence materials reduces the dynamic range of intelligibility as compared to low-context sentences, so that a given increase in the amount of information provided yields a correspondingly greater change in per-

cent correct. Another factor may have been the production styles of the target talkers. The IEEE sentence sets were produced with a conversational style, while the CUNY sentence set we used was produced at a slower rate, using a more highly articulated speaking style. The differences in context and speaking styles between the sentence materials may be responsible for the difference in the amount of improvement due to the tone observed between experiments 1 and 2.

IV. GENERAL DISCUSSION

The results of the present experiments demonstrate that F0 can be an important cue under simulated EAS listening. Both F0 and the amplitude envelope contributed significantly to the EAS benefit, and when these two cues were combined, a benefit of as much as 57 percentage points was observed over vocoder only. When sentence context was high, the presence of the tone provided as much benefit as low-pass speech, and this benefit was not adversely affected by the presence of a tone carrying the F0 and amplitude envelope of the low-pass background speech.

The results from Experiment 2, in which the presence of background F0 information provided no benefit to intelligibility, do not support the argument that EAS benefit is due to listeners' improved ability to segregate target and background, as has been proposed (Chang *et al.*, 2006; Qin and Oxenham, 2006; Kong *et al.*, 2005). We would have expected the background to provide some benefit if segregation were the explanation.

On the other hand, intelligibility has been shown to be adversely affected when F0 is flattened (Assman, 1999; Laures and Weismer, 1999) or inverted (Culling *et al.*, 2003). These results have led these investigators to conclude that F0 may help indicate where to listen in an utterance by providing information about where content words are located. This explanation is consistent with the glimpsing account of EAS benefit suggested by Kong and Carlyon (2007), and is a plausible account of our results, although our experiments were not designed specifically to test this.

The benefit provided by a tone carrying F0 and envelope cues observed here in simulations of EAS holds promise for implant listeners, particularly those with significantly elevated thresholds in the low-frequency region. These listeners may not normally benefit from EAS because of their inability to adequately hear a sufficient bandwidth of the speech in the low-frequency region, even with amplification. On the other hand, higher sensation levels (and consequently the potential for some EAS benefit) may be achieved if the low-frequency stimulus were a relatively narrowband tone, modulated in frequency with F0 and in amplitude with the amplitude envelope. It may be possible to construct a processor that extracts F0 in real time, and then applies it (as well as the amplitude envelope) to a tone in the low-frequency acoustic region. Note that this approach is different from previous attempts at exploiting F0 information with electric stimulation. For example, Rosen and Ball (1986) found little benefit from a single-channel implant that conveyed F0. The processor described here would present F0 information in the acoustic region, which would be combined with multiple-channel electrical stimulation in the higher frequency region. This approach is more similar to that used by Faulkner *et al.* (1992), who found that an F0-modulated tone was helpful to profoundly hearing-impaired listeners as an aid to lip reading. The processor described here could greatly expand the population of cochlear implant users who stand to benefit from EAS to include individuals who have very little residual low-frequency hearing. Of course, the efficacy of such a processor would depend on whether the effects observed here in simulation emerge with real implant patients. In that regard, Brown and Bacon (2008) provided promising preliminary data collected on EAS patients that showed that combining electric stimulation with an acoustically presented tone modulated in frequency and amplitude (as done here) can be an effective means of achieving the benefits of EAS.

V. SUMMARY

A tone modulated either in frequency to track the F0 of the target talker, or in amplitude with the amplitude envelope of the target talker provides significant benefit in simulated EAS.

A tone modulated in both frequency and amplitude ($T_{\text{F0-env}}$) generally provides more benefit than either cue alone.

The presence of the tone ($T_{\text{F0-env}}$), under these simulated conditions, resulted in improvements in intelligibility of between 23 and 57 percentage points over vocoder alone. Intelligibility was not affected by the presence of a tone that tracked the F0 and envelope of a background.

ACKNOWLEDGMENTS

This research was supported by grants from the National Institute of Deafness and Other Communication Disorders (NIDCD Grant Nos. DC01376 and DC008329 awarded to S.P.B.). The authors gratefully acknowledge Michael Dorman for insightful discussions about this work, Robert Carlyon and an anonymous reviewer for helpful comments on earlier versions of the manuscript, and Bethany Stover for assistance with data collection.

[1]The experimenter who had knowledge of the conditions during testing ran 11 of the participants. The one who did not ran 14 participants. A two-factor analysis of variance was conducted with experimenter as one factor, and processing condition as the other (only the V and $V/T_{\text{F0-env}}$ conditions were included, since any bias would presumably be between these conditions). The interaction term in this analysis was not significant, $p=0.68$, indicating that bias in scoring by the knowledgeable experimenter was not a significant factor in the pattern of results.

[2]Pilot experiments using multi-talker babble background (SNR of +10) indicated that there was a learning effect in the $V/T_{\text{F0-env}}$ condition with a performance asymptote at around 80 sentences (about 30 percentage points of improvement was observed), but no learning was observed in the V condition (400 sentences of exposure yielded about 4 percentage points of improvement). This led to our decision to provide 100 sentences of practice in the $V/T_{\text{F0-env}}$ condition prior to data collection. However, a reviewer expressed concern that the improvements in performance observed in the $V/T_{\text{F0-env}}$ condition, relative to the other tonal conditions, may have been due to this practice. To address this, we compared the learning effects of each of the four tonal conditions (V/T, V/T_{F0}, V/T_{env}, and $V/T_{\text{F0-env}}$) by exposing a separate set of subjects (three per group) to one of the conditions for 100 sentences, then testing them on V, $V/500$,

V/T, V/T_{env}, V/T_{F0}, and $V/T_{F0\text{-}env}$. While we cannot perform inferential statistics on these data due to the small sample size, we can report that overall the pattern of results was similar to that observed in experiment 1 and, moreover, within each learning condition, mean performance was always better in the $V/T_{F0\text{-}env}$ condition than in the condition in which subjects received learning. Subjects who received learning in V/T averaged 25 percent correct when tested in V/T, and 39 percent correct in $V/T_{F0\text{-}env}$. Subjects exposed to V/T_{F0} during learning scored 45 percent correct in V/T_{F0} and 61 percent correct in $V/T_{F0\text{-}env}$. And subjects exposed to V/T_{env} scored 17 percent correct in V/T_{env} and 60 percent correct in $V/T_{F0\text{-}env}$. We therefore conclude that our results were not biased by presenting 100 sentences of the $V/T_{F0\text{-}env}$ condition prior to testing.

[3]In experiment 1 we used T to refer to the tone in the low-frequency region that carried the F0 and envelope information. In experiment 2, T could refer to tones carrying information about the target, the background, or both. Thus, we use S and B to differentiate the target or "signal" from the background.

[4]Under vocoder-only conditions, we combined a female target (mean F0 =211) with either of two female backgrounds (mean F0s of 184 and 240). Thus, the mean F0 "distance" between the target and each background talker was roughly equal (27 and 29 Hz). We found that for the six participants we tested, the background talker whose F0 was lower was a much better masker than the talker with the higher mean F0. While certainly not conclusive, this result demonstrates that neither simple F0 distance nor gender categorization may be enough to predict the amount of masking a particular background might yield with a particular target.

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

Assmann, P. F. (**1999**). "Fundamental frequency and the intelligibility of competing voices," Proceedings of the 14th International Congress of Phonetic Science San Francisco, August.

Assmann, P. F., and Summerfield, Q. (**1990**). "Modeling the perception of concurrent vowels: Vowels with different fundamental frequencies," J. Acoust. Soc. Am. **88**, 680–697.

Auditec (**1997**). "Auditory Tests (Revised)," Compact Disc, Auditec, St. Louis.

Bell, T. S., Dirks, D. D., and Trine, T. D. (**1992**). "Frequency-importance functions for words in high- and low-context sentences," J. Speech Hear. Res. **35**, 950–959.

Bird, J., and Darwin, C. J. (**1997**). "Effects of a difference in fundamental frequency in separating two sentences," Paper for the 11th International Conference on Hear., Granthan, UK, August.

Boothroyd, A., Hanin, L., and Hnath, T. (**1985**). "A sentence test of speech perception: reliability, set equivalence, and short term learning (Internal Rep. No. RCI 10)," New York: Speech.

Boothroyd, A., Hnath-Chisolm, T., Hanin, L., and Kishon-Rabin, L. (**1988**). "Voice fundamental frequency as an auditory supplement to the speechreading of sentences," Ear Hear. **9**, 306–312.

Brokx, J. P. L., and Nooteboom, S. G. (**1982**). "Intonation and the perceptual separation of simultaneous voices," J. Phonetics **10**, 23–36.

Brown, C. A., and Bacon, S. P. (**2008**). "A new approach to electric-acoustic stimulation," J. Acoust. Soc. Am. **123**, 3054.

Chang, J. E., Bai, J. Y., and Zeng, F. (**2006**). "Unintelligible low-frequency sound enhances simulated cochlear-implant speech recognition in noise," IEEE Trans. Biomed. Eng. **53**, 2598–2601.

Culling, J. F., and Darwin, C. J. (**1993**). "Perceptual separation of simultaneous vowels: Within and across-formant grouping by F0," J. Acoust. Soc. Am. **93**, 3454–3467.

Culling, J. F., Hodder, K. I., and Toh, C. Y. (**2003**). "Effects of reverberation on perceptual segregation of competing voices," J. Acoust. Soc. Am. **114**, 2871–2876.

Cullington, H. E., and Zeng, F. (**2008**). "Speech recognition with varying numbers and types of competing talkers by normal-hearing, cochlear-implant, and implant simulation subjects," J. Acoust. Soc. Am. **123**, 450–461.

de Cheveigné, A., and Kawahara, H. (**2002**). "YIN, a fundamental frequency estimator for speech and music," J. Acoust. Soc. Am. **111**, 1917–1930.

Dorman, M. F., Spahr, A. J., Loizou, P. C., Dana, C. J., and Schmidt, J. S. (**2005**). "Acoustic simulations of combined electric and acoustic hearing (EAS)," Ear Hear. **26**, 371–380.

Duchnowski, P., Lum, D. S., Krause, J. C., Sexton, M. G., Bratakos, M. S., and Braida, L. D. (**2000**). "Development of speechreading supplements based on automatic speech recognition," IEEE Trans. Biomed. Eng. **47**, 487–496.

Faulkner, A., Ball, V., Rosen, S., Moore, B. C., and Fourcin, A. (**1992**). "Speech pattern hearing aids for the profoundly hearing impaired: Speech perception and auditory abilities," J. Acoust. Soc. Am. **91**, 2136–2155.

Fry, D. B. (**1955**). "Duration and intensity as physical correlates of linguistic stress," J. Acoust. Soc. Am. **27**, 765–768.

Gantz, B. J., and Turner, C. (**2004**). "Combining acoustic and electrical speech processing: Iowa/ Nucleus hybrid implant," Acta Oto-Laryngol. **124**, 344–347.

Gantz, B. J., and Turner, C. W. (**2003**). "Combining acoustic and electrical hearing," Laryngoscope **113**, 1726–1730.

Gantz, B. J., Turner, C., Gfeller, K. E., and Lowder, M. W. (**2005**). "Preservation of hearing in cochlear implant surgery: Advantages of combined electrical and acoustical speech processing," Laryngoscope **115**, 796–802.

Grant, K. W., and Walden, B. E. (**1995**). "Predicting auditory-visual speech recognition in hearing-impaired listeners," The 13th International Congress of Phonetic Science, Stockholm, August.

Hillenbrand, J., Getty, L. A., Clark, M. J., and Wheeler, K. (**1995**). "Acoustic characteristics of American English vowels," J. Acoust. Soc. Am. **97**, 3099–3111.

Holt, L. L., Lotto, A. J., and Kluender, K. R. (**2001**). "Influence of fundamental frequency on stop-consonant voicing perception: A case of learned covariation or auditory enhancement?" J. Acoust. Soc. Am. **109**, 764–774.

IEEE (**1969**). "IEEE recommended practice for speech quality measurements," IEEE Trans. Audio Electroacoust. **17**, 225–246.

Kong, Y., and Carlyon, R. P. (**2007**). "Improved speech recognition in noise in simulated binaurally combined acoustic and electric stimulation," J. Acoust. Soc. Am. **121**, 3717–3727.

Kong, Y., Stickney, G. S., and Zeng, F. (**2005**). "Speech and melody recognition in binaurally combined acoustic and electric hearing," J. Acoust. Soc. Am. **117**, 1351–1361.

Laures J. S. and WeismerG. (**1999**). "The effects of a flattened fundamental frequency on intelligibility at the sentence level," J. Speech Lang. Hear. Res. **42**, 1148–1156.

Qin, M. K., and Oxenham, A. J. (**2003**). "Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers," J. Acoust. Soc. Am. **114**, 446–454.

Qin, M. K., and Oxenham, A. J. (**2006**). "Effects of introducing unprocessed low-frequency information on the reception of envelope-vocoder processed speech," J. Acoust. Soc. Am. **119**, 2417–2426.

Rosen, S., and Ball, V. (**1986**). "Speech perception with the Vienna extracochlear single-channel implant: A comparison of two approaches to speech coding.," Br. J. Audiol. **20**, 61–83.

Spahr, A. J., and Dorman, M. F. (**2004**). "Performance of subjects fit with the advanced bionics CII and nucleus 3G cochlear implant devices," Arch. Otolaryngol. Head Neck Surg. **130**, 624–628.

Spitzer, S. M., Liss, J. M., and Mattys, S. L. (**2007**). "Acoustic cues to lexical segmentation: A study of resynthesized speech," J. Acoust. Soc. Am. **122**, 3678–3687.

Stickney, G. S., Zeng, F., Litovsky, R., and Assmann, P. (**2004**). "Cochlear implant speech recognition with speech maskers," J. Acoust. Soc. Am. **116**, 1081–1091.

Turner, C. W., Gantz, B. J., Vidal, C., Behrens, A., and Henry, B. A. (**2004**). "Speech recognition in noise for cochlear implant listeners: Benefits of residual acoustic hearing," J. Acoust. Soc. Am. **115**, 1729–1735.

von Ilberg, C., Kiefer, J., Tillein, J., Pfenningdorff, T., Hartmann, R., Stürzebecher, E., and Klinke, R. (**1999**). "Electric-acoustic stimulation of the auditory system. New technology for severe hearing loss," ORL **61**, 334–340.

Anatomic development of the oral and pharyngeal portions of the vocal tract: An imaging study[a)]

Houri K. Vorperian[b)]
Waisman Center, University of Wisconsin-Madison, 1500 Highland Avenue, Room 481, Madison, Wisconsin 53705

Shubing Wang
Department of Statistics, University of Wisconsin-Madison, 1300 Highland Avenue, Madison, Wisconsin 53705

Moo K. Chung
Department of Biostatistics & Medical Informatics, University of Wisconsin-Madison, 1500 Highland Avenue, Madison, Wisconsin 53705

E. Michael Schimek, Reid B. Durtschi, and Ray D. Kent
Waisman Center, University of Wisconsin-Madison, 1500 Highland Avenue, Madison, Wisconsin 53705

Andrew J. Ziegert and Lindell R. Gentry
Department of Radiology, University of Wisconsin Hospital and Clinics, 600 Highland Avenue, E1-311 Clinical Science Center, Madison, Wisconsin 53792

(Received 1 February 2008; revised 10 September 2008; accepted 23 December 2008)

The growth of the vocal tract (VT) is known to be non-uniform insofar as there are regional differences in anatomic maturation. This study presents quantitative anatomic data on the growth of the oral and pharyngeal portions of the VT from 605 imaging studies for individuals between birth and 19 years. The oral (horizontal) portion of the VT was segmented into lip-thickness, anterior-cavity-length, oropharyngeal-width, and VT-oral, and the pharyngeal (vertical) portion of the VT into posterior-cavity-length, and nasopharyngeal-length. The data were analyzed to determine growth trend, growth rate, and growth type (neural or somatic). Findings indicate differences in the growth trend of segments/variables analyzed, with significant sex differences for all variables except anterior-cavity-length. While the growth trend of some variables displays prepubertal sex differences at specific age ranges, the importance of such localized differences appears to be masked by overall growth rate differences between males and females. Finally, assessment of growth curve type indicates that most VT structures follow a combined/hybrid (somatic and neural) growth curve with structures in the vertical plane having a predominantly somatic growth pattern. These data on the non-uniform growth of the vocal tract reveal anatomic differences that contribute to documented acoustic differences in prepubertal speech production.
 [DOI: 10.1121/1.3075589]

PACS number(s): 43.70.Aj, 43.70.Ep [AL] Pages: 1666–1678

I. INTRODUCTION

The development of speech in children is based in part on the maturation of the macroanatomy of the vocal tract (VT), including increases in size, typically expressed as vocal tract length (VTL). During development from infancy to adulthood, the length of the VT increases more than twofold, from approximately 7 to 8 cm in infants to 15 to 18 cm in adult females and males, respectively. Such growth has been characterized to be non-uniform in that the oral and pharyngeal portions of the VT are thought to undergo different growth patterns (Fant, 1960, 1975; Kent and Vorperian, 1995; Fitch and Giedd, 1999). There is longstanding interest in the relative or relational growth of the anterior (oral) portion of the VT, which is in the horizontal plane, versus the posterior (pharyngeal) portion of the VT, which is in the vertical plane. This interest has been due, in part, to understanding the acoustic changes that result from changes in the VT, particularly the differences in the formant frequencies between males and females where the differences cannot be explained by a simple scale factor inversely proportional to the overall VTL (Fant, 1975). Another reason of interest in the differential growth of the oral and pharyngeal portions of the VT has been from an evolutionary perspective. For example, the hypothesis that the elongation of the pharyngeal portion of the VT contributed to the emergence of speech in humans (Lieberman, 1975), and the hypothesis that the permanent shaping of the VT into two tubes—a horizontal oral tube and a vertical pharyngeal tube—which permit the pro-

[a)]Portions of this paper were presented in 2007 at the 154th meeting of the Acoustical Society of America in New Orleans, LA.
[b)]Electronic mail: vorperian@waisman.wisc.edu

duction of quantal vowels (Stevens, 1989), evolved gradually with increased vocalization complexity and frequency (Fitch, 2000, 2002).

The inverse relation between VTL and formant frequencies is well established. As VTL increases during the course of development, formant frequencies decrease (Fant, 1960). By puberty, there are significant differences in VTL between males and females (Fitch and Giedd, 1999). However, the acoustic differences between males and females are non-uniform and thus cannot be explained solely by differences in overall VTL. Fant (1960, 1975), using radiographic data, noted the longer pharynx in adult men compared to women and children, and using a two-tube simplified model (front tube/oral cavity length and back tube/pharyngeal cavity length), concluded that such anatomic differences in the oral versus pharyngeal portions of the VT can account for the observed differences in vowel formant frequencies between males and females. Often, the resonant characteristics of children and females are grouped together with the contention that they are similar. This assumption discounts the documented acoustic differences between males and females by the age of 4 years (e.g., Perry *et al.*, 2001), as well as the documented developmental sex differences in the first, second, and third formant frequencies (Vorperian and Kent, 2007). Furthermore, acoustic studies indicate that formant frequencies do not decrease during the first 2 years of life (Robb *et al.*, 1997; Gilbert *et al.*, 1997; Kent and Murray, 1982), a finding that appears to be inconsistent with acoustic theory at first glance since there are documented increases in VTL during this period (Vorperian *et al.*, 1999, 2005). However, it is evident in Fant's writings (Fant, 1975) that although he used simple tube models to make physiologic-acoustic interpretations, he specified the importance of dimensions other than tube length—specifically laryngeal cavity—and indicated the need for more detailed anatomical studies and calculations. Thus, ultimately it is necessary to have a thorough multi-dimensional understanding of the anatomic development of the acoustic resonator, or the VT, between males and females to help establish anatomic-acoustic correlates during the course of development.

From an evolutionary perspective, the achievement of a length ratio of 1 between the pharyngeal and oral portions of the VT has been postulated to be an anatomic advantage for the emergence of speech (Lieberman, 1975; Lieberman *et al.* 1992). However, work on articulatory models has contended otherwise and highlighted the importance of auditory feedback and neurocognition (Callan *et al.* 2000, Menard *et al.*, 2004, 2007; Boe *et al.*, 2007). Irrespective, interest in evolution persists and there is a continuing need for a thorough understanding of the developmental anatomic changes in the VT. Although there have been a select number of radiographic and imaging studies on the anatomic development of the VT (Arens *et al.*, 2002, Fitch and Giedd, 1999; Lieberman *et al.*, 2001; Vorperian *et al.* 1999, 2005), there is a paucity of detailed quantitative data on sex-specific anatomic development of the VT and its component oral and pharyngeal cavities. As noted above, such information is important for understanding the biologic basis of speech development, and the complex anatomic-acoustic interactions or formant-cavity affiliations. Furthermore, such information would be useful in advancing non-uniform scaling factors for VT or speaker normalization (Vorperian and Kent, 2007). From an anatomic perspective, development of the VT and its constituent cavities can be understood in terms of the growth of the hard and soft tissues that give form to the acoustic conduit. The complex structures of the human head have diverse embryologic structures and tissues of origin (Sadler, 2006; Larsen, 2001; Sperber, 1973) and can be grouped into different schedules of growth and maturation. Scammon (1930) described three general growth schedules of the head and neck region: neural (brain and cranium), somatic (hard and soft tissues of the face), and lymphatic (tonsils and adenoid). This heterogeneity of growth pattern is a major factor to be considered in accounting for the development of the VT.

A primary goal of this study is to characterize the anatomic growth trend and growth rate of the VT and its oral and pharyngeal portions. Also, since different biological structures have sex-specific differences in growth schedule or growth curve type such as the male and female growth charts used clinically to assess the growth of head circumference and body stature (height and weight), a secondary goal of this study is to numerically quantify the sex-specific growth curves of the VTL and its oral and pharyngeal portions as neural or somatic, following Scammon (1930). Distinguishing differences between the neural and somatic growth curves lies in growth trend/rate and percent growth. Structures with a neural growth curve display a very rapid growth following birth to achieve about 80% of its adult size during early childhood, followed by a slower steady growth until adulthood. Head circumference, a measurement that is mostly in the horizontal plane, follows such a growth curve. The somatic growth curve also displays a very rapid growth following birth, but size achieved during early childhood is barely 25%–40% of adult size. This early phase is followed by a regular and slow growth until maturity except for a brief period of rapid growth during puberty. Body height and facial growth, measurements in the vertical plane, follow this type of growth curve. Vorperian *et al.* (2005) related the growth curve type of VT structures to the anatomic orientation of the various structures. They reported that structures in the horizontal plane, such as the hard palate, appear to follow a neural growth curve, structures in the vertical plane, such as laryngeal descent, appear to follow a somatic growth curve, and structures oriented in both planes, such as tongue length and VTL, appear to have a hybrid, or a combined or intermediate neural and somatic growth curve. Although Lieberman and McCarthy (1999) and Lieberman *et al.* (2001) also reported differences in the growth type of the horizontal versus the vertical portions of the VT, they concluded that the growth of the VT has a predominantly skeletal or somatic growth curve.

II. METHODS

A. Subjects

Using imaging studies performed for medical reasons that are considered not to affect growth and development, a total of 605 head or neck imaging studies (307 MRI and 298

TABLE I. Summary of F-test for gender effect. The first two rows specify the number of male and female measurements available and included in the analysis from imaging studies per variable, and number of outliers per variable. The remaining rows reflect the results of the *F*-test for global sex differences of fits per variable and include the dfs, the *F*-values, and the *p*-values of the *F*-tests ($\times 10^{-04}$).

	VTL	VT-V	PCL	NPhL	VT-H	LTh	ACL	OPhW	VT-O
n males/ outliers	274/4	277 /5	278/4	277/3	316/4	311/8	278/6	269/7	308 /4
n females/ outliers	222/7	226 /2	224/5	223/3	263 /3	270/2	224/3	222/3	261 /1
df	4,476	4,487	4,484	4,485	4,563	4,562	4,484	4,472	4,555
F-value	40.38	27.13	34.29	4.93	6.38	11.74	0.03	2.53	3.29
p-value	$<1.0\times10^{-12}$	$<1.0\times10^{-12}$	$<1.0\times10^{-12}$	6.619×10^{-04}	5.004×10^{-05}	3.622×10^{-09}	9.980×10^{-01}	3.972×10^{-02}	1.106×10^{-02}
Significant	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes

CT) were selected for making measurements where the VT structures could be visualized. The imaging studies were from 327 males and 278 females between the ages birth to 19 years. While developing/acquiring this imaging database, significant effort was directed to select cases representative of the age range with an equivalent number of males and females per age/year. The weights of the majority of the cases were at the 50th percentile reference growth curves for boys and girls, with all cases falling between the 25th and 95th percentile growth curves as per the National Center of Health Statistics growth charts (Centers for Disease Control and Prevention (CDC), 2000).

B. Procedures

1. Image acquisition

The medical imaging studies used for making measurements included both MRI and CT cases with subjects in the supine position. The method for MRI image acquisition has been described previously (Vorperian *et al.*, 1999, 2005). The CT studies of the entire neck and face were obtained using General Electric helical CT scanners. Most young pediatric patients were sedated using either chloral hydrate 50 mg/kg administered orally, or Propofol, Midazolam, Atropine, or Fentanyl administered intramuscularly (1 mg/kg), prior to entering the scanner. Once in the scanner, the facial structures of all subjects were placed centrally in the head coil using the laser lights of the GE scanners. The CT scans were obtained using axial 1.25 mm thick slices. The images were obtained from the thoracic inlet, inferiorly, to the top of the orbits, superiorly with a 15–30 cm field of view. The field of view for young pediatric subjects ranged from 15 to 25 cm while those of older pediatric or adult subjects ranged from 25 to 30 cm. The images were reconstructed with a matrix size of 512×512. In-plane image resolution is given by dividing the field of view by the matrix size. Resolution ranged from 0.29 to 0.48 mm for pediatric patients and from 0.48 to 0.58 mm for adult patients. The axial CT scan data were reconstructed using two different algorithms (standard and bone plus) to provide two image sets, one optimized for soft tissue detail (standard algorithm) and one optimized for bone detail (bone algorithm). The axial images were then used to generate multiplanar reformatted images in the sagittal and coronal planes with a 2–3 mm slice thickness. The images were initially stored on a McKesson Horizon Rad Station PACS system. Next, the images were set anonymous using a General Electric Advantage Windows workstation. Then, the entire study was saved in DICOM format for image analysis and data acquisition.

2. Data acquisition

Data acquisition entailed making measurements of the variables defined below from the midsagittal plane. Midsagittal slice selection for CT studies entailed the use of both the standard and bone algorithms of the same slice to meet the same criteria as used and described previously for MRI studies (Vorperian *et al.*, 1999, 2005) and entails the distinct visualization of cerebral sulci extending to the corpus callosum, also the visualization of the fourth ventricle, the full length of the cerebral aqueduct of Sylvius, the pituitary gland, part of the optic chiasm, the brainstem, and the cervical cord. Of note, image reconstruction using the software EFILM (by Merge eFilm) was implemented if the slice was not in true midline. Anatomic landmarks for making measurements were placed on the midsagittal bone algorithm slice by two researchers independently while visualizing both the bone and standard algorithms of the selected midsagittal slice. The two sets of landmarks were compared, discrepancies resolved using the radiologist's medical expertise, and a final "master" set of landmarks was used for making measurements (Chung *et al.*, 2008). Given the developmental nature of this study, the use of this landmark placement protocol was necessary as it improved measurement accuracy between 82% and 100% (average 98%) as measured by reduction in error variability. Of note, despite the careful selection of imaging studies where VT structures could be visualized, occasionally, not all anatomic landmarks could be clearly seen. Rather than excluding the entire study in such instances, all the measurements that could be secured using placed landmarks were secured and included in data analysis. The number of imaging studies/cases per variable are listed in Table I. Measurements were made using the image measurement software SIGMASCAN PRO by SYSTAT (formerly SPSS and Jandel Scientific) which was calibrated for each case/slice using the hash scale mark on the CT image/slice.

3. Variables

The nine variables used in this study are illustrated in Fig. 1 and are defined below. The data were acquired either via direct distance measurements (centimeters) of the vari-

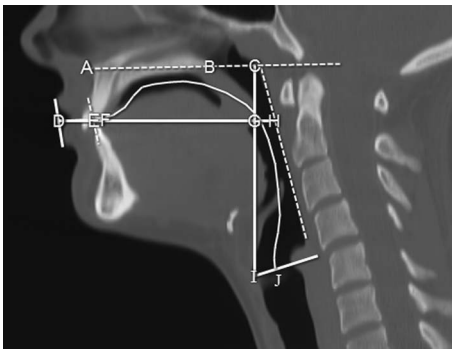


FIG. 1. Midsagittal CT image displaying the anatomic landmarks used for making measurements. Measurements include VTL, the curvilinear line extending from points *D*-to-*J*. VT-V vertical distance from points *I*-to-*C* and consisting of two segments PCL (points *I*-to-*G*) and NPhL (points *G*-to-*C*). VT-H horizontal distance from points *D*-to-*H*, consisting of three line segments: LTh (points *D*-to-*E*), ACL (points *F*-to-*G*), and OPhW (points *G*-to-*H*). Also, the segment VT-O (points *E*-to-*H*).

ables from the midsagittal slice or calculated from those direct measurements. The variables included VTL, the curvilinear distance along the midline of the tract starting at the glottis (level of true vocal folds) to the intersection with a line drawn tangentially to the lips (curvilinear distance from points *J* to *D* in Fig. 1). Variables in the vertical plane included vocal tract-vertical (VT-V), the vertical distance from the glottis to the palatal plane (the ANS-PNS plane which extends from the anterior nasal spine to the posterior nasal spine; vertical distance from point *I* to *C* in Fig. 1). This VT-V distance consisted of two segments: the posterior-cavity length (PCL), the vertical distance of a line drawn from the glottis to the intersection with the end of the oral or anterior-cavity length (ACL) (distance *I*-to-*G* in Fig. 1); and the nasopharyngeal-length (NPhL), VT-V minus PCL (distance *G*-*C* in Fig. 1). Also, variables in the horizontal plane included vocal tract-horizontal (VT-H), the horizontal distance from a line tangential to lips to the posterior pharyngeal wall (horizontal distance *D*-to-*H* in Fig. 1). This VT-H distance consisted of three segments: lip-thickness (LTh), the distance at the level of the stomion between two lines, the first of which is drawn tangential to the anterior aspect and the second to the posterior or buccal aspect of the maxillary and mandibular lips (distance *D*-to-*E* in Fig. 1); ACL, the horizontal distance of a line drawn from the lingual incisor (start of the hard palate) to the intersection with the vertical line drawn from the glottis to the *A*-to-*B* palatal plane (distance *F*-to-*G* in Fig. 1); and the oropharyngeal-width (OPhW), VT-H minus LTh minus ACL (distance *G*-to-*H* in Fig. 1). Another horizontal segment calculated included the Vocal Tract-Oral (VT-O), VT-H minus LTh (distance *E*-to-*H* in Fig. 1).

C. Statistical analysis

1. Pooling of CT and MRI data

To maximize the data available for analysis, it was desirable to include the data from both CT and MRI studies. To determine if measurements from the CT and MRI studies can be pooled, data were secured from 28 cases that had both MRI and CT studies in less than a 3 month interval and the data were compared using paired t-tests. The measurement discrepancy between CT and MRI were not significant at alpha=0.01 for the variables used in this study. Therefore, the CT and MRI data were combined for increased statistical power.

2. Analysis of growth trend, rate, and type

The data (distance measurements as a function of age) were plotted to identify growth trends and gender differences (Figs. 2–10). Following the removal of outliers from the data, as specified in Table I, two sets of analyses were done. The criterion used for outlier removal was measurements exceeding 2.576σ which gives the probability of less than 0.01 for false removal of data. The first analysis was to characterize sex-specific growth curve trends and its growth rate for each of the nine variables. Based on the model selection framework, various polynomial model fits were performed and the fourth degree model was determined to describe/fit

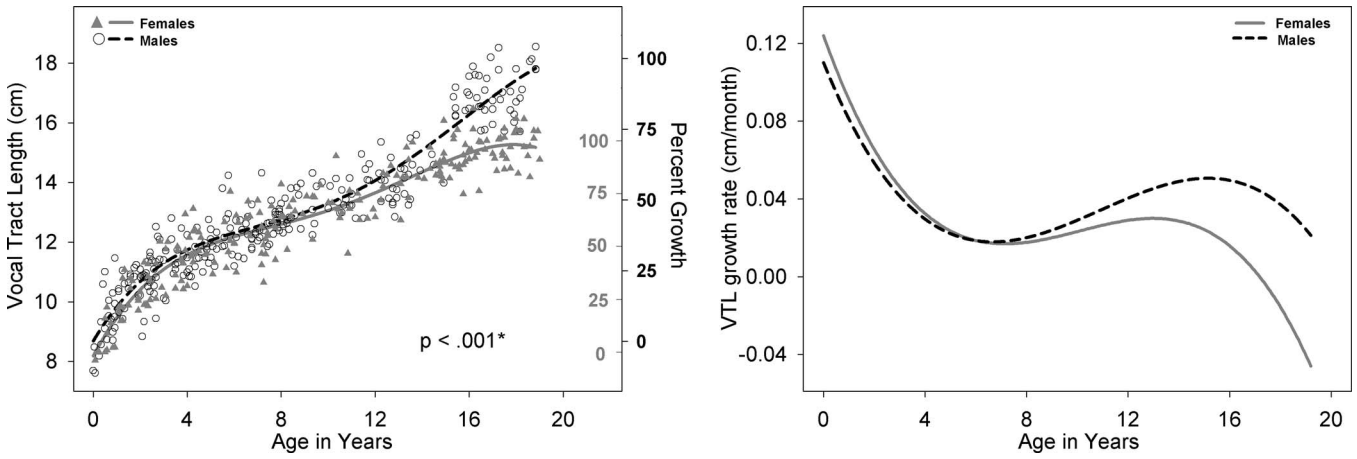


FIG. 2. Left panel: VTL development for males (black open circle) and females (shaded gray triangles) with growth curve using a fourth degree polynomial fit for males (dashed black line) and females (solid gray line). Male versus female fits are significantly different ($p<0.001$). The second *Y*-axis reflects the percent growth of adult size for males (black, outwards tick orientation) and females (gray, inwards tick orientation). VTL is defined as the curvilinear distance along the midline of the VT starting at the level of the glottis to the intersection with a line drawn tangentially to the lips. Right panel: Growth rate of VTL, derived from the polynomial fit on left (cm/months), for males (dashed black line) and females (solid gray line) as a function of age.

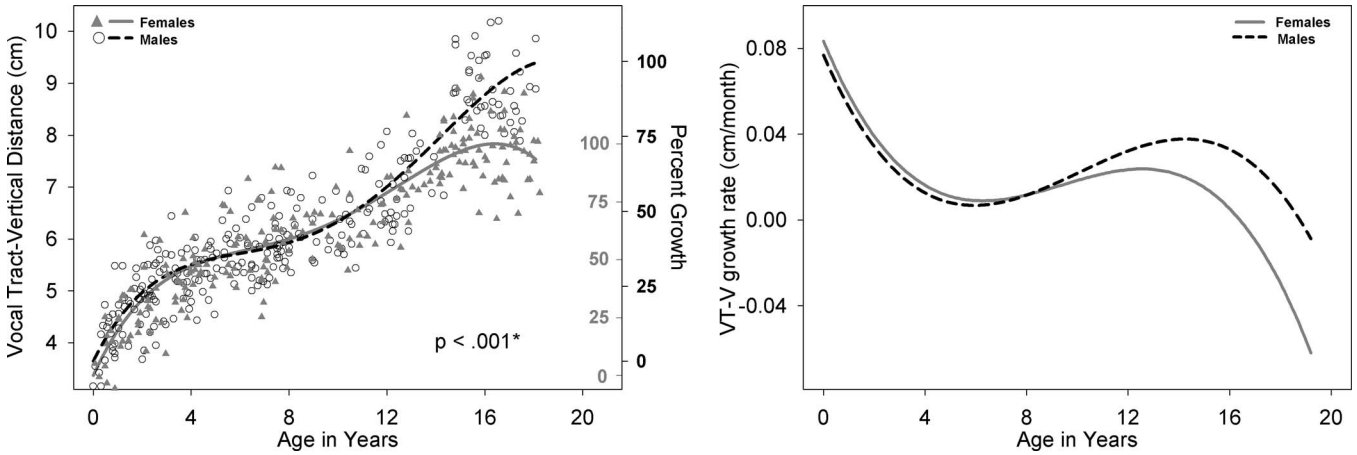


FIG. 3. Left panel: VT-V development for males (black open circle) and females (shaded gray triangles) with growth curve using a fourth degree polynomial fit for males (dashed black line) and females (solid gray line). Male versus female fits are significantly different ($p<0.001$). The second Y-axis reflects the percent growth of adult size for males (black, outwards tick orientation) and females (gray, inwards tick orientation). VT-V is defined as the vertical distance from the glottis to the palatal plane. Right panel: Growth rate of VT-V, derived from the polynomial fit on left (cm/months), for males (dashed black line) and females (solid gray line) as a function of age.

the data best. Degree 1–3 models were too simplistic to model complex growth patterns. In comparing degree 4 and 5 models, degree 4 was determined to be a better model based on checking the significance of the sum of squared residuals (Rao and Toutenburg, 1999). Sex differences of the fits were assessed using an F-test and the results are summarized in Table I. The fits are plotted in the left panel of Figs. 2–10 and the p-values for sex differences are embedded in each figure. The growth rate (cm/months), plotted in the right panel of Figs. 2–10, was computed by differentiating the estimated model. The second analysis included an assessment of growth curve type by regressing each variable's model fit to a neural (N) growth curve, a somatic (S) growth curve, and both N and S growth fits. The head circumference growth curve was used as the basic model for the neural growth curve, and the body height growth curve was used as the basic model for the somatic growth curve (Centers for Disease Control and Prevention (CDC), 2000; Nellhaus, 1968; Vorperian *et al.*, 2007). The numeric quantification results are summarized in Table II. Also, the percent growth to reach adult size is displayed on the second y-axis in the left panel of Figs. 2–10 with black outwards tick orientation for males and gray inwards tick orientation for females.

III. RESULTS

A. Growth trend and growth rate

All available measurements of the nine variables from the 605 cases between birth and age 19 years are plotted for males and females in the left panel of Figs. 2 and 10, with sex-specific fits—using the fourth degree polynomial model—and a p-value reflecting the outcome of the F-test for sex differences which were significant ($p<0.05$) for all variables except ACL. Table I includes for each of the nine variables the total number of male/female measurements available for analysis, the number of outliers per variable for each sex, the degrees of freedom (dfs), the F-value, and the p-value. The growth trend for all the variables displayed in

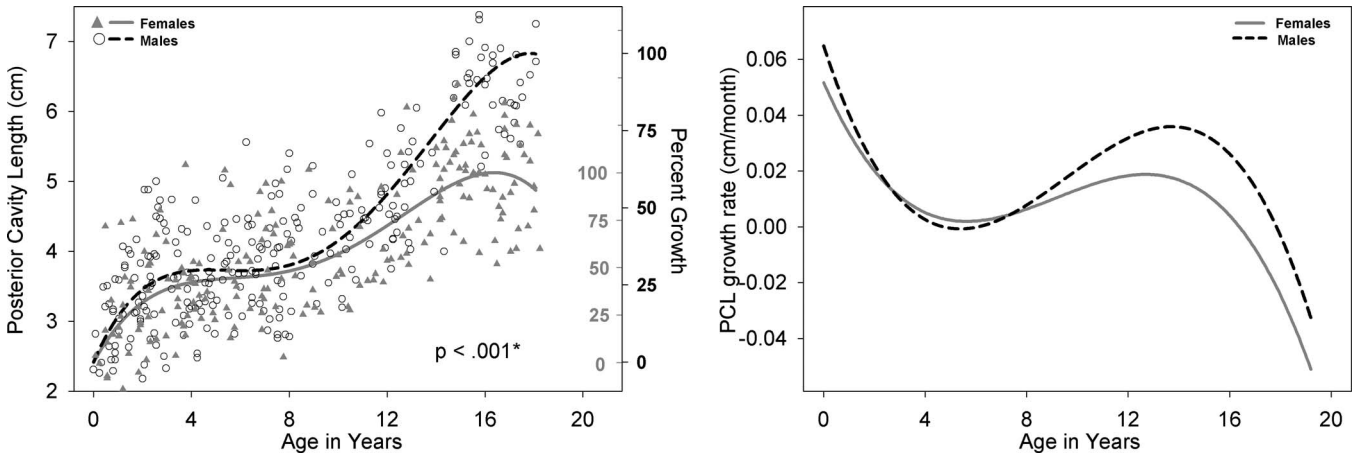


FIG. 4. Left panel: PCL development for males (black open circle) and females (shaded gray triangles) with growth curve using a fourth degree polynomial fit for males (dashed black line) and females (solid gray line). Male versus female fits are significantly different ($p<0.001$). The second Y-axis reflects the percent growth of adult size for males (black, outwards tick orientation) and females (gray, inwards tick orientation). PCL is defined as the vertical distance of a line drawn from the glottis to the intersection with the end of the oral or anterior-cavity length. Right panel: Growth rate of PCL for males (dashed black line) and females (solid gray line) as a function of age.

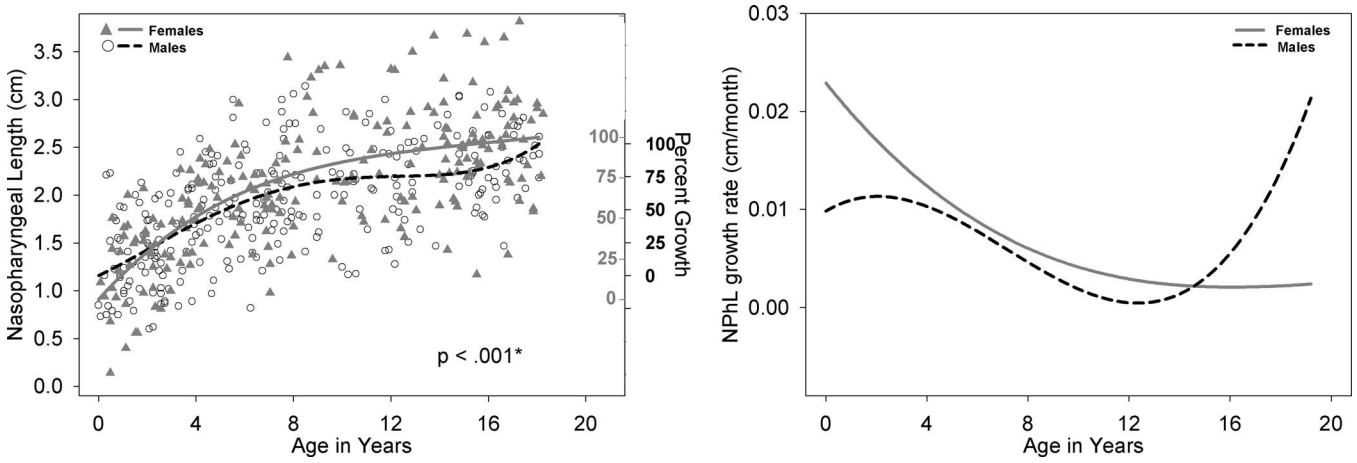


FIG. 5. Left panel: NPhL development for males (black open circle) and females (shaded gray triangles) with growth curve using a fourth degree polynomial fit for males (dashed black line) and females (solid gray line). Male versus female fits are significantly different (p=0.008). The second Y-axis reflects the percent growth for males (black, outwards tick orientation) and females (gray, inwards tick orientation). NPhL is a calculated measurement of VT-V minus PCL. Right panel: Growth rate of NPhL for males (dashed black line) and females (solid gray line) as a function of age.

Figs. 2–10 (left panel) reflects an overall nonlinear growth trend throughout the first 18 years of life, as noted by the general increases in measurements for both males and females. Of note, the negative growth noted past age 17 for select variables is mostly due to the nature of the data set, which is cross-sectional rather than longitudinal, and the insufficient number of measurements past age 17, which affect the fit at extreme ages adversely—a boundary limitation of regression fits/models. As noted above, the F-test for sex differences was significant for all variables except ACL (p =0.9). While the visual display of the fits for most structures shows large sex differences past age 12, there are some structures that show somewhat large sex differences at an earlier age [such as PCL (Fig. 4) and NPhL (Fig. 5)]. Furthermore, during early childhood, the fits reflect slight sex differences for most structures where typically females are smaller than males except for OPhW (Fig. 9) and VT-O (Fig. 10) where the female fit displays a slightly larger value than the male fit. Also, during childhood, some sex differences appear to emerge and then dissolve [e.g., OPhW (Fig. 9)]. This appears to be related to growth rate (Figs. 2–10, right panel) and also growth type, as addressed below. Figures 2–10 (left panel) also include a second Y-axis documenting the percent growth to reach the mature adult size. The growth percentages on this second y-axes indicate that some structures reach the adult size sooner [e.g., VT-H (Fig. 6)] than others [e.g., VT-V (Fig. 3)]. Such findings have implication for growth type and are further discussed below.

Figures 2–10 (right panel) show the growth rates (cm/months) for each of the nine variables, where growth rate decreases during approximately the first 8 years of life. This is then followed by an increase in growth rate during approximately ages 8–14 years, and then a decrease in growth rate after age 16 for most variables except for NPhL, OPhW, and LTh in males (Figs. 5, 7, and 9, right panel) where there is an apparent increase in growth rate after age 12 particularly for NPhL and OPhW (Figs. 5 and 9). As noted above, the polynomial fits (left panel) reflect sexual dimorphism

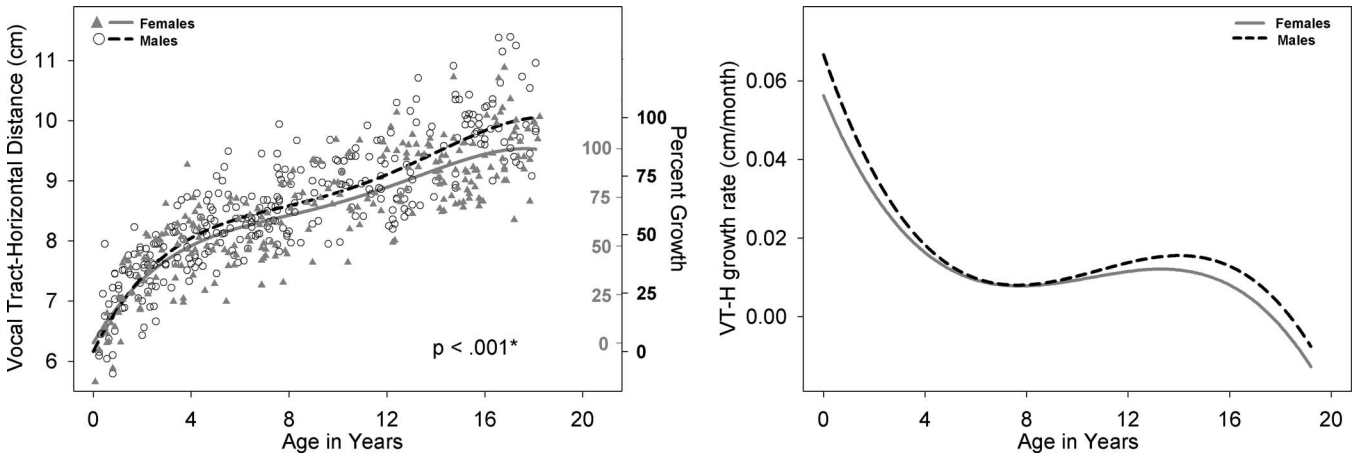


FIG. 6. Left panel: VT-H development for males (black open circle) and females (shaded gray triangles) with growth curve using a fourth degree polynomial fit for males (dashed black line) and females (solid gray line). Male versus female fits are significantly different (p<0.0001).The second Y-axis reflects the percent growth for males (black, outwards tick orientation) and females (gray, inwards tick orientation). VT-H is defined as the horizontal distance from a line tangential to lips to the posterior pharyngeal wall. Right panel: Growth rate of VT-H for males (dashed black line) and females (solid gray line) as a function of age.

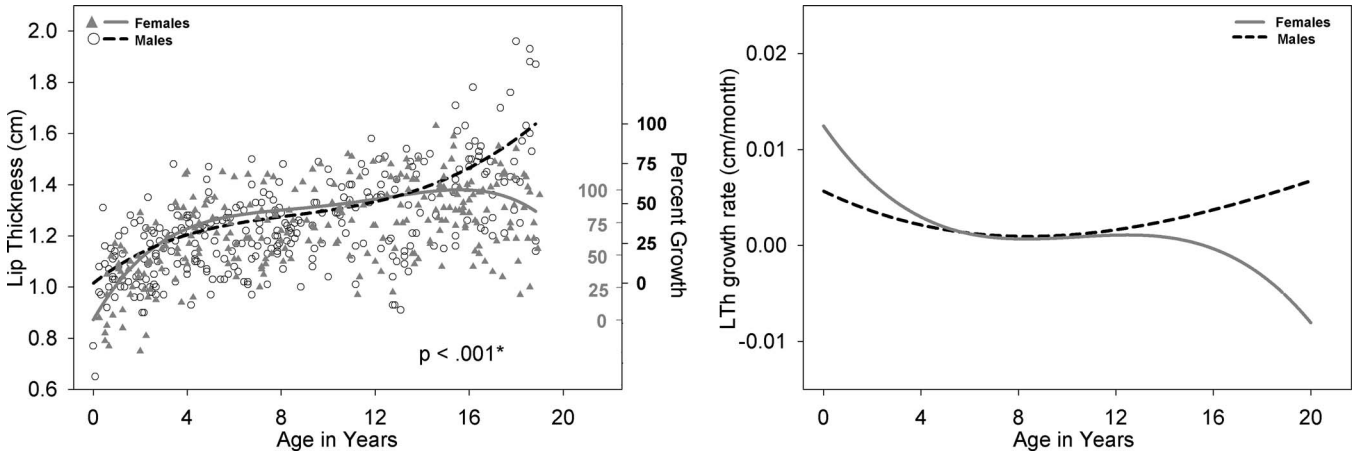


FIG. 7. Left panel: LTh development for males (black open circle) and females (shaded gray triangles) with growth curve using a fourth degree polynomial fit for males (dashed black line) and females (solid gray line). Male versus female fits are significantly different ($p<0.001$). The second Y-axis reflects the percent growth for males (black, outwards tick orientation) and females (gray, inwards tick orientation). LTh defined as the distance between two lines, the first of which is drawn tangential to the anterior aspect of the maxillary and mandibular lips and the second to the posterior or buccal aspect of the maxillary and mandibular lips. Right panel: Growth rate of LTh for males (dashed black line) and females (solid gray line) as a function of age.

emerging after age 12 for most variables; however, the fits for some variable, such as PCL (Fig. 4), NPhL (Fig. 5), and OPhW (Fig. 9) show prepubertal sex differences. Such differences are also evident in the growth rate figures of those variables, with more distinct demarcation of the age where differences in growth rate emerge. Thus, the growth rate figures may serve to detect the emergence of differences in growth trend. For example, while sex differences in VTL (Fig. 2, left panel) become apparent after age 12, the growth rate figure (Fig. 2, right panel) validates the emergence of such differences by age 8.

B. Growth type: Neural or somatic

The assessment of growth type (neural, somatic, or hybrid), as defined by Scammon (1930), has to be based on percent growth [as marked on the second Y-axes of the polynomial fit (Figs. 2–10, left panel] and Table II which quantifies numerically the regression of each fit with a neural growth curve and a somatic growth curve. Table II (last column) shows that most of the variables have a hybrid or a combined neural and somatic growth type except for LTh in females and OPhW in males. As seen in Fig. 7 (left panel), LTh growth in females approaches the mature value by age 6, an indication of a predominantly neural growth curve. Also, the numeric quantification in Table II indicates that for each variable, males and females have different growth types. For example OPhW is 61% neural for males and 75% somatic for females. However, in general, structures in the vertical plane appear to have a predominantly somatic growth pattern.

IV. DISCUSSION

A. Current findings

This study quantifies the non-uniform growth of the oral and pharyngeal portions of the VT structures in males and

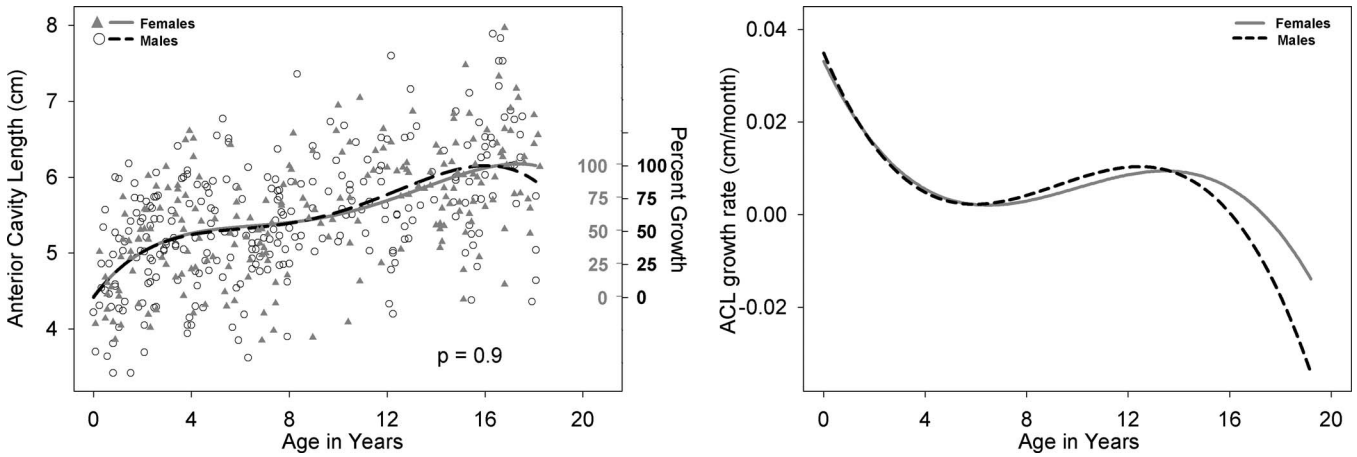


FIG. 8. Left panel: ACL development for males (black open circle) and females (shaded gray triangles) with growth curve using a fourth degree polynomial fit for males (dashed black line) and females (solid gray line). Male versus female fits are not significantly different ($p=0.9$). The second Y-axis reflects the percent growth for males (black, outwards tick orientation) and females (gray, inwards tick orientation). ACL is defined as the horizontal distance from the beginning of the hard palate to the intersection with the vertical line drawn from the glottis to the palatal plane. Right panel: Growth rate of ACL for males (dashed black line) and females (solid gray line) as a function of age

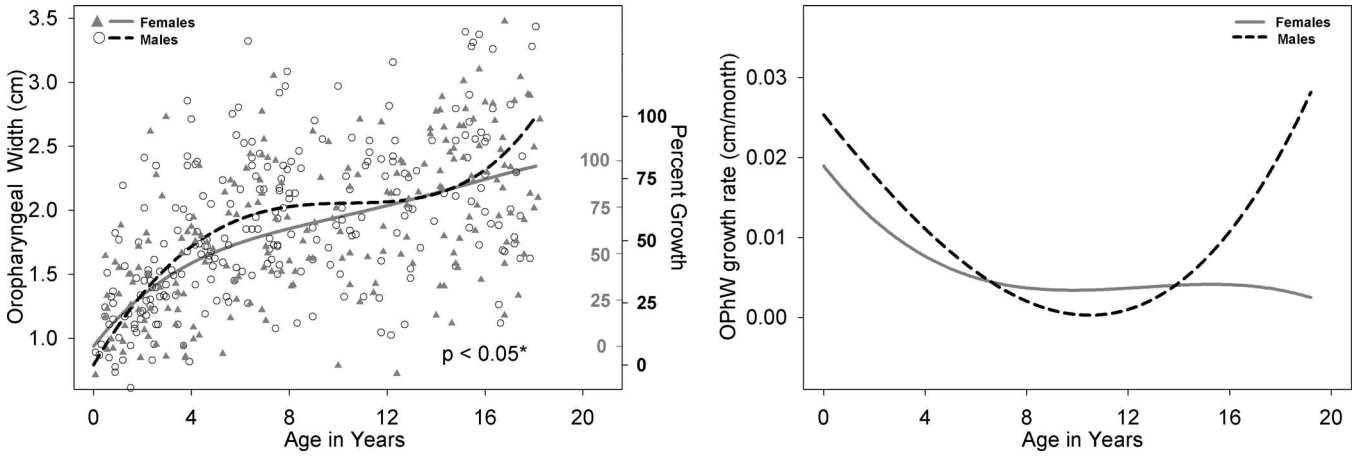


FIG. 9. Left panel: OPhW development for males (black open circle) and females (shaded gray triangles) with growth curve using a fourth degree polynomial fit for males (dashed black line) and females (solid gray line). Male versus female fits are significantly different ($p<0.01$). The second Y-axis reflects percent growth for males (black, outwards tick orientation) and females (gray, inwards tick orientation). OPhW is calculated using the measurements of VT-H minus LTh minus ACL. Right panel: Growth rate of OPhW for males (dashed black line) and females (solid gray line) as a function of age.

females during approximately the first 2 decades of life. While it has been known for decades that growth is nonuniform, the data presented here are without precedent in that they provide information on sex-specific anatomic development of VTL and of segments within the oral and pharyngeal portions of the VT. Furthermore, this quantification is detailed in that it specifies the growth trend and growth rate, as well as the growth type (neural versus somatic growth curves as defined by Scammon, 1930) for each of those various segments for males and females during the course of development. These major findings are highlighted and discussed in Secs. IV A 1 and IV A 2, followed by a discussion on implications for speech acoustics and other ramifications.

1. Growth trend and growth rate

For all nine variables, the growth trend is somewhat rapid during the first few years of life with overall growth continuing until maturity (Figs. 2–10, left panel). However, this rapid growth during the first 4–6 years of life differs for structures in the oral versus the pharyngeal portions of the VT. As seen on the second/right y-axes in Figs. 2–10, the variables in the oral region, which are in the horizontal plane, approximate the mature adult size sooner than the variables in the pharyngeal region, which are in the vertical plane. These findings were expected based on the reports by Lieberman and McCarthy (1999), Lieberman *et al.* (2001), and Vorperian *et al.* (2005) on growth types, and the implications are discussed further below in Sec. IV A 2. Also, such differences in the growth schedule of structures in the oral/horizontal versus pharyngeal/vertical portions of the VT are to be expected given their diverse embryologic origins. Specifically, the embryologic structure/tissue type for the oral cavity is the stomodeum/neural crest-ectoderm, whereas for the pharynx, it is the foregut/endoderm and splanchnic mesoderm (Sadler, 2006). Again, this issue is addressed further in Sec. IV A 2. As for the negative growth trend past age 17, as noted in Sec. III, this is not an accurate representation

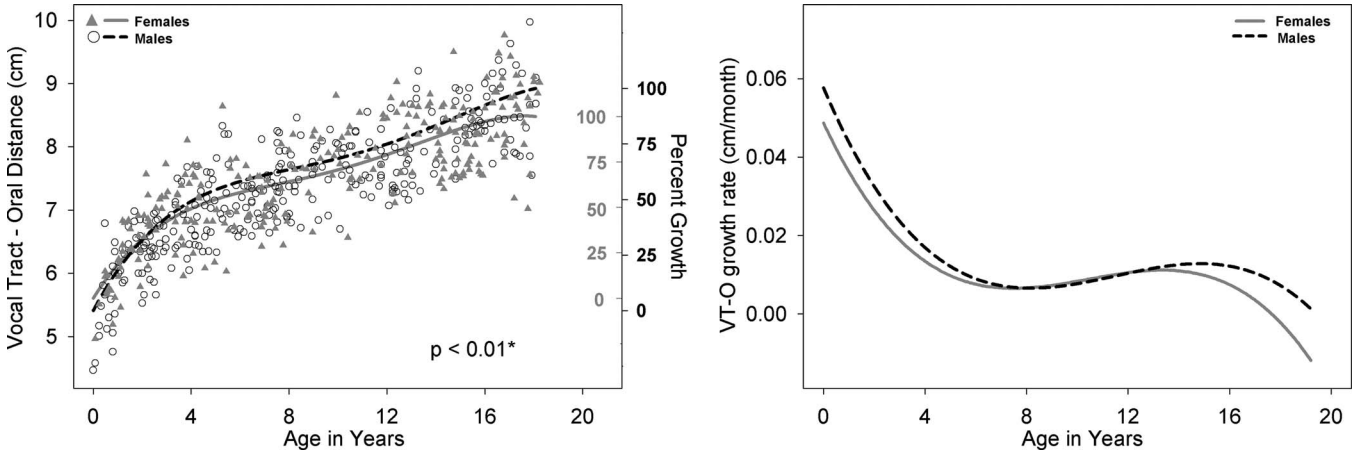


FIG. 10. Left panel: VT-O development for males (black open circle) and females (shaded gray triangles) with growth curve using a fourth degree polynomial fit for males (dashed black line) and females (solid gray line). Male versus female fits are significantly different ($p=0.03$). The second Y-axis reflects percent growth for males (black, outwards tick orientation) and females (gray, inwards tick orientation). VT-O is a calculated using the measurements VT-H minus LTh. Right panel: Growth rate of VT-O for males (dashed black line) and females (solid gray line) as a function of age.

TABLE II. Numeric quantification in percent of the regression of sex-specific polynomial model fit per variable with the neural and somatic growth curves.

Variable	Sex	% somatic	% neural	Numeric quantification
VTL	Female	88	12	Somatic/neural
	Male	100	0	Somatic
VT-V	Female	98	2	Somatic
	Male	99	1	Somatic
PCL	Female	100	0	Somatic
	Male	97	3	Somatic
NPhL	Female	89	11	Somatic/neural
	Male	91	9	Somatic/neural
VT-H	Female	60	40	Somatic/neural
	Male	76	24	Somatic/neural
LTh	Female	15	85	Neural/somatic
	Male	90	10	Somatic/neural
ACL	Female	61	39	Somatic/neural
	Male	90	10	Somatic/neural
OPhW	Female	75	25	Somatic/neural
	Male	39	61	Neural/somatic
VT-O	Female	61	39	Somatic/neural
	Male	67	33	Somatic/neural

of growth trend but a limitation of the model/regression fit at the extreme age due to the nature of the cross-sectional data set that has limited data points past age 17.

There are significant sex differences in the growth trend of eight of the nine variables studied (ACL not significant), with the majority of the variables displaying distinct sexual dimorphism after age 12 (Figs. 2–10, left panel). However, as noted in the Sec. III, there are some variables that display prepubertal sexual dimorphism at much earlier ages such as VT-H (Fig. 6) where differences are evident by age 4 and persist until adulthood, and other variables such as VTL (Fig. 2), PCL (Fig. 4), NPhL (Fig. 5), and OPhW (Fig. 9) where sex differences are either localized to specific ages or fluctuate during the course of development and appear to be related to growth rate differences between males and females. This is discussed further in the following paragraph.

Since growth rate differences between males and females are evident (Figs. 2–10, right panel), ideally the interpretation of sex differences in growth trend fits (Figs. 2–10, left panel) should be made in conjunction with growth rate curves because differences in growth rate between males and females can easily mask differences in growth trend. The variables NPhL (Fig. 5) and OPhW (Fig. 9) demonstrate this point clearly. Such an observation indicates that to assess whether prepubertal sex differences are being masked by differences in growth rate, it is necessary to implement more localized or smaller age range comparisons between males and females rather than comparisons across all ages (birth to 19) as was done in this study. In addition, it seems that such localized age group comparison between males and females should ideally use groups smaller than a 6 year span since Vorperian *et al.* (2005) did not identify sexual dimorphism of any VT structures between the ages of birth to approximately 6 years. However, Lieberman *et al.* (2001) have noted sex differences in the growth of oropharyngeal-width (the distance from the posterior pharyngeal wall to the posterior margin of oral cavity) with growth being slightly larger in males between the ages of 1.75 and 4.75 years. It is reasonable to expect the identification of prepubertal sexual dimorphism in VT structures given that the growth charts (height, weight, and head circumference) used clinically are sex-specific (Centers for Disease Control and Prevention (CDC), 2000) and the knowledge that there is a strong correlation between VTL and body size (both height and weight; Bennett, 1981; Fitch and Giedd, 1999). This issue is discussed further under Sec. IV B.

Growth rate, past age 16 (Figs. 2–10, right panel), decreases or levels off for all structures except for the variables NPhL and OPhW in males where growth rate increases. This warrants assessment of whether growth in the oro-nasopharyngeal portion persists beyond approximately the first 2 decades of life. This point is addressed further under Sec. IV B.

2. Growth type: Neural or somatic

The distinguishing difference between the neural and somatic growth curves lies not only in growth trend but also percent growth (Scammon, 1930). As expected and discussed above, structures in the oral region reached maturity sooner than structures in the pharyngeal region implying a neural growth type for VT structures in the oral/horizontal region and somatic growth type for VT structures in the pharyngeal/vertical region. Numeric quantification of growth type for the nine variables studied, as summarized in Table II, supports the following conclusions. Most of the nine variables studied appear to follow either a predominantly somatic growth type or a hybrid/combined somatic and neural growth curves for both males and females. Structures in the vertical plane or the pharyngeal region follow a predominantly somatic growth type for both males and females. The numeric quantification confirms findings reported by Lieberman and McCarthy (1999), Lieberman *et al.* (2001), and Vorperian *et al.* (2005). Structures in the horizontal plane or the oral region, however, follow a hybrid somatic/neural or neural/somatic growth type. The numeric quantification does not match the expectations of a predominantly neural growth type except for the variable LTh in females and OPhW in males. Interestingly, this latter finding on OPhW is consistent with the findings of Lieberman *et al.* (2001) of the growth of oropharyngeal-width being slightly larger in males between the ages of 1.75 and 4.75 years. The variables in the oral/horizontal region also seem to display sex differences in growth type; particularly distinct sex differences are noted for the variables LTh and OPhW. For example, the variable OPhW is predominantly neural (61%) in males with percent growth of adult size at about 70% by age 6, whereas it is predominantly somatic (75%) in females with percent growth of adult size at about 50% by age 6. As noted above,

such sex-specific differences in growth type have also been reported by Lieberman *et al.* (2001) and further highlight the need to take into account growth rate differences when making interpretations of growth pattern and growth type. VTL, which is a variable in both planes, follows a predominantly somatic growth curve in females (88%) and a purely somatic growth curve in males (100%). Thus, multiple factors are at play and ultimately determine the growth type of VT structures. While structure orientation was a useful general guide, it is necessary to consider other factors, such as differences in the growth rate between males and females and the embryologic origin of the various VT structures—such as the palate, mandible, and tongue—that border the oral and pharyngeal cavities.

B. Acoustic implications

The growth trend, growth rate and growth type quantification from medical imaging studies, as specified above, are part of the biological basis of speech development. The noted sex-specific differences in the growth of the oral and pharyngeal portions of the VT can have profound acoustic implications. One important aspect is relating anatomic growth patterns to developmental changes in speech acoustics, including male/female differences. A potential limitation for the comparison of the two modalities is that the anatomic data are from imaging studies with subjects in the supine position at rest, whereas all developmental speech acoustic data are from participants in the upright position. The gravitational effect on airway patency is well documented when subjects are in supine during quiet breathing where the pharyngeal region volume is reportedly reduced in supine as opposed to upright (e.g., Eckmann *et al.*, 1996). Also, differences in tongue behavior during speech production in upright and supine have been reported by Stone *et al.* (2007) using ultrasound imaging, as well as differences in both the soft tissues and the rigid structures (mandible and larynx) during vowel production in upright and supine have been reported by Kitamura *et al.* (2005) using open-type MRI. However, Stone *et al.* (2007) did not identify any significant phoneme effects, and no differences in the first and second formant values. Thus, it is reasonable to conclude that albeit postural differences between the two modalities, it is a worthwhile venture to hypothesize developmental anatomic-acoustic correlates given the uniqueness of the developmental anatomic data set of VT structures and its quantification that are presented in this paper.

Acoustic theory states that as the VT lengthens with age, formant frequencies decrease (Fant, 1960). Current findings on VTL growth trend (Fig. 2, left panel) show a nonlinear growth pattern with significant sex differences. Although the test of sex differences utilized here is a global test (i.e., takes into account all ages), the polynomial model growth fits show sex differences emerging at about age 12, confirming previous reports on significant sexual dimorphism in VTL after age 11 years (Fitch and Giedd, 1999) or 13.75 years (Lieberman *et al.*, 2001). The fits also show a rapid increase in VTL during early childhood that is about 2 cm during the first 2 years of life, a finding that is consistent with previous findings from a much smaller pool of imaging studies (Vorperian *et al.*, 2005). Such anatomic findings are not consistent with two reports on acoustic data. The first acoustic observation is that formant frequencies remain unchanged, i.e., do not decrease, during the first 2 years of life (Buhr, 1980; Gilbert *et al.*, 1997; Kent and Murray, 1982; Robb *et al.*, 1997). The second acoustic observation is that acoustic differences between males and females are present by age 4 (Perry *et al.*, 2001). Thus, the quest is to identify other developmental anatomic findings of VT structures that might explain those two acoustic findings.

Regarding the first point that there are no reported decreases in formant frequencies despite increase in VTL, current findings indicate that there is rapid growth of OPhW which implicates volumetric changes/increases in the pharyngeal region. Thus, it is possible to postulate that there is an interplay between VTL and volume. As Fant (1975) has noted, VT dimensions to consider in physiologic-acoustic interpretations in addition to oral and pharyngeal tube lengths include laryngeal cavity dimensions. Additional research to quantify developmental changes in length versus volume in the VT, particularly in the hypopharyngeal and laryngopharyngeal regions during the first two years of life, is warranted.

Regarding the second point of acoustic differences being present by age 4 (Perry *et al.*, 2001), although the VTL growth fits show slight differences between males and females until sexual dimorphism becomes apparent around age 12, the anatomic differences do not appear to be related to the reported acoustic differences between males and females that are present by age 4. However, it seems reasonable to hypothesize that such male/female acoustic differences are related to differences in the development of the oral and pharyngeal portions of the VT. Visual inspection of the growth fits of all the other VT variables examined in this paper (Figs. 3–10, left panel) indicate anatomic sex differences of select variables, some localized to specific ages that warrant a closer or localized examination in the future. Specifically, the variables LTh, PCL, and NPhL during the first 2 years of life; also, the variables OPhW and VT-O, from about age 4 years to 12 years, which are larger in males than in females, warrant detailed analysis of anatomic differences between males and females.

While it is premature to establish anatomic-acoustic correlates from the data presented here, additional localized assessments of those variables in the midsagittal plane would not only assist in deciphering anatomic-acoustic correlates but also guide future research efforts in studying multidimensional growth of the nasopharyngeal region (Vorperian *et al.* 2005; Vorperian and Kent, 2007). The vowel acoustic space development data summarized by Vorperian and Kent (2007) indicate that the F1–F3 vowel quadrilateral dispersions is greater than F1–F2 dispersion and that they appear to be sensitive to age and speaker sex differences. According to Fant's (1975) two-tube model, the pharyngeal cavity length is affiliated with the second formant, and the oral cavity length is affiliated with the third formant. Indeed, if the noted differences in VT-H growth, specifically the segments OPhW and VT-O, are confirmed to be significantly different be-

tween males and females, then this can explain, at least in part, the documented prepubertal acoustic differences between males and females where there are no significant differences in VTL, i.e., before age 11 (Fitch and Giedd, 1999). Of note, although the two-tube model is simplistic and ignores cross modes in the transfer function of the VT, it is nonetheless a model with a good first approximation and a reasonable approach to begin establishing hypotheses on anatomic-acoustic correlates that may be tested using various more complex modeling approaches (e.g., Motoki, 2002; Menard *et al.*, 2004, 2007; Boe *et al.*, 2007).

A final acoustic observation is the noted decrease in formant frequencies in the aging population which has been attributed to the lengthening of the VT (Benjamin, 1997; Endres *et al.*, 1971; Linville and Fisher, 1985). Xue and Hao's (2003) findings, using acoustic reflection technology, reflected changes in VT volume but not VTL for both genders; also they noted increases in oral cavity length and volume, again for both genders. The assessment of anatomic change in VT structures using imaging studies may help address whether growth persists past age 18 where most VT structures are presumed to have reached their mature size. In particular, it will be worthwhile to study the growth trend of structures in the oro-naso-pharyngeal region since current findings reflect increases in the growth rate of NPhL and OPhW in males past the age of presumed maturity.

C. Other ramifications

The quantification on the development of the oral and pharyngeal portions of the VT presented here can be used in a multitude of ways. For one, it can be used in developmental articulatory models including sex-specific developmental models to test predictions on anatomic-acoustic correlates as specified above. To date, modeling efforts have estimated or approximated the growth of the VTL, oral cavity length, and pharyngeal cavity length (e.g., Goldstein, 1980; Martland *et al.*, 1996; Callan *et al.* 2000; Menard *et al.* 2004, 2007; Boe *et al.* 2007). However, most of those modeling efforts are not sex-specific, and some are based on assumptions of growth types that appear not to be valid based on current findings. For example, the predictions of Martland *et al.* (1996) are based on the assumption that the growth of the oral and the pharyngeal tracts follow the neural and somatic growth types, respectively. However, present findings indicate that while the assumption of somatic growth type of the pharyngeal region is mostly valid, the assumption on the growth of the oral region is not since present findings show the growth of this region to follow a combined somatic/neural growth type with apparent sex differences (females, 60/40% and males, 76/24% males). One implication from having detailed anatomic quantification on the non-uniform growth of the VT is that the data can be used in establishing scaling factors for normalization (age and sex differences in the acoustic properties of speech), a long standing issue in acoustic phonetics and speech technology.

Another ramification is promoting our understanding of the anatomic bases of motor adjustment during development, including speech development. To date, the noted high degree of variability, from acoustic and physiologic studies, in children's performance has been linked to the maturation of the neuromuscular control or speech motor control (Eguchi and Hirsh, 1969; Tingley and Allen, 1975, Smith *et al.*, 1983; Sharkey and Folkins, 1985; Smith, 1994; Kent, 1976, 1992; Kent and Forner, 1980; Lee *et al.*, 1999; Walsh and Smith, 2002; Wohlert and Smith, 2002). However, it is important not to discount the role of anatomic change in neuromuscular control. For example, this study reflects a rapid growth rate in the pharyngeal region during approximately the ages 8–14 particularly in males. It is within this age range, specifically ages 10–14, that there are reports on high levels of childhood asphyxiation by food, particularly in males, (Baker *et al.*, 1992). Thus, it is reasonable to relate rapid anatomic growth to limited neuromuscular control.

Furthermore, as Boe *et al.* (2007) pointed out, knowledge of the size of the oral and pharyngeal cavities throughout ontogeny not only help understand speech acquisition processes in infants but also assist in understanding phylogeny and the evolution of speech. More recently, Fitch (2000, 2002) noted that many mammals dynamically create a two-tube VT during loud vocalizations where the larynx is lowered into the pharyngeal cavity resulting in a long and well-defined pharyngeal tube. He therefore hypothesized that the evolution of the human speech apparatus with a descended larynx arose gradually with increased vocalization complexity and frequency. Undoubtedly, such a viewpoint places more emphasis on understanding the growth of the oral cavity and specially the pharyngeal cavity, and less emphasis on the ratio of the horizontal and vertical parts of the VT.

V. CONCLUSION

This study quantifies the non-uniform growth of the VT in terms of both regional differences (oral versus pharyngeal portions of the VT) and sex differences in the growth of those regions. The growth quantification of nine variables in the oral and pharyngeal regions during approximately the first 2 decades of life in males and females indicate that structures in the oral region reach the mature size sooner than those in the pharyngeal region. Structures in the oral region follow a combined somatic and neural growth curve, whereas structures in the pharyngeal region follow a predominantly somatic growth curve as defined by Scammon (1930). While findings confirm global sex differences for most structures, the growth trend and growth rate figures reflect not only postpubertal sexual dimorphisms for most structures but also prepubertal sex differences at particular ages for select structures. This warrants localized assessment of sex differences as a means to explore the anatomic basis of the observed acoustic differences between males and females during early childhood.

ACKNOWLEDGMENTS

This work was supported in part by NIH Research Grant Nos. R03 DC4362 (Anatomic Development of the Vocal Tract: MRI Procedures) and R01 DC6282 (MRI and CT Studies of the Developing Vocal Tract), from the National Institute of Deafness and other Communicative Disorders

(NIDCD). Also, by a core Grant No. P-30 HD03352 to the Waisman Center from the National Institute of Child Health and Human Development (NICHHD). We thank Celia Choih for assistance with placing the anatomic landmarks and making the necessary measurements and Katelyn J. Kassulke for assistance with data entry and figures.

Arens, R., McDonough, J., Corbin, A., Hernandez, M., Maislin, G., Schwab, R., and Pack, A.(**2002**). "Linear dimensions of the upper airway structure during development: Assessment by magnetic resonance imaging," Am. J. Respir. Crit. Care Med. **165**, 117–122.

Baker, S. P., O'Neill, B., Ginsburg, M. J., and Li, G. (**1992**). *The Injury Fact Book*, 2nd ed. (Oxford University Press, New York).

Benjamin, B. J. (**1997**). "Speech production of normally aging adults," Semin. Speech Lang. **18**, 135–141.

Bennett, S. (**1981**). "Vowel formant frequency characteristics of preadolescent males and females," J. Acoust. Soc. Am. **69**, 231–238.

Boe, L., Heim, J., Honda, K., Maeda, S., Badin, P., and Abry, C. (**2007**). "The vocal tract of newborn humans and Neanderthals: Acoustic capabilities and consequences for the debate on the origin of language. A reply to Liberman (2007a)," J. Phonetics **35**, 564–581.

Buhr, R. D. (**1980**). "The emergence of vowels in an infant," J. Speech Lang. Hear. Res. **23**, 73–94.

Callan, D. E., Kent, R. D., Guenther, F. H., and Vorperian, H. K. (**2000**). "An auditory-feedback-based neural network model of speech production that is robust to developmental changes in the size and shape of the articulatory system," J. Speech Lang. Hear. Res. **43**, 721–736.

Centers for Disease Control and Prevention (CDC) (**2000**). National Center for Health Statistics Clinical Growth Charts, http://www.cdc.gov/growthcharts/ (Last viewed November, 2006).

Chung, D., Chung, M. K., Durtschi, R. B., Gentry, L. R., and Vorperian, H. K. (**2008**). "Measurement consistency from magnetic resonance images," Acad. Radiol. **15**, 1322–1330.

Eckmann, D. M., Glassenberg, R., and Gavriely, N. (**1996**). "Acoustic reflectometry and endotracheal intubation," Anesth. Analg. (Baltimore) **83**, 1084–1089.

Eguchi, S., and Hirsh, I. J. (**1969**). "Development of speech sounds in children," Acta Oto-Laryngol. **257**, 1–51.

Endres, W., Bamback, W., and Flosser, G. (**1971**). "Voice spectrograms as a function of age, voice disguise, and voice imitation," J. Acoust. Soc. Am. **49**, 1842–1848.

Fant, G. (**1975**). "A note on vocal tract size factors and non-uniform F-pattern scalings," Speech Transm. Lab. Q. Prog. Status Rep. **4**, 22–30.

Fant, G. (**1960**). *Acoustic Theory of Speech Production* (Mouton, The Hague).

Fitch, W. T. (**2002**). "Comparative vocal production and the evolution of speech: Reinterpreting the descent of the larynx," in *The Transition to Language*, edited by A. Wray (Oxford University Press, Oxford), pp. 21–45.

Fitch, W. T. (**2000**). "The evolution of speech: A comparative review," Trends Cogn. Sci. **4**, 258–267.

Fitch, W. T., and Giedd, J. (**1999**). "Morphology and development of the human vocal tract: A study using magnetic resonance imaging," J. Acoust. Soc. Am. **106**, 1511–1522.

Gilbert, H. R. Robb, M. P., and Chen, Y. (**1997**). "Formant frequency development: 15 to 36 months," J. Voice **11**, 260–266.

Goldstein, U. G. (**1980**). "An articulatory model for the vocal tracts of growing children," Ph.D. thesis, MIT, Cambridge, MA.

Kent, R. D. (**1992**). "The biology of phonological development," *Phonological Development: Models, Research, Implications*, edited by C. A. Ferguson, L. Menn, and C. Stoel-Gammon (York, Timonium, MD), pp. 65–90.

Kent, R. D. (**1976**). "Anatomical and neuromuscular maturation of the speech mechanism: Evidence from acoustic studies," J. Speech Hear. Res. **19**, 421–447.

Kent, R. D., and Forner, L. L. (**1980**). "Speech segment durations in sentence recitations by children and adults," J. Phonetics **12**, 157–68.

Kent, R. D., and Murray, A. D. (**1982**). "Acoustic features of infant vocalic utterances at 3, 6, and 9 months," J. Acoust. Soc. Am. **72**, 353–65.

Kent, R. D., and Vorperian, H. K. (**1995**). "Anatomic development of the craniofacial-oral-laryngeal systems: A review," J. Med. Speech-Lang. Pathol. **3**, 145–190.

Kitamura, T., Takemoto, H., Honda, K., Shimada, Y., Fujimoto, I., Syakudo, Y., Masaki, S., Kuroda, K., Oku-uchi, N., and Senda, M. (**2005**). "Difference in vocal tract shape between upright and supine postures: Observation by an open-type MRI scanner," Acoust. Sci. & Tech. **26**, 465–468.

Larsen, W. (**2001**). *Human Embryology* (Churchill Livingstone, Philadelphia, PA).

Lee, S., Potamianos, A., and Narayanan, S. (**1999**). "Acoustics of children's speech: Developmental changes of temporal and spectral parameters," J. Acoust. Soc. Am. **105**, 1455–1468.

Lieberman, P. (**1975**). *On the Origins of Language* (Macmillan, New York).

Lieberman, D. E., McCarthy, R. C., Hiiemae, K. M., and Palmer, J. B. (**2001**). "Ontogeny of postnatal hyoid and larynx descent in humans," Arch. Oral Biol. **46**, 117–128.

Lieberman, D. E., and McCarthy, R. C. (**1999**). "The ontogeny of cranial base angulation in humans versus chimpanzees and its implications for reconstructing pharyngeal dimensions," J. Hum. Evol. **36**, 487–517.

Lieberman, P., Laitman, J. T., Reidenberg, J. S., and Gannon, P. S. (**1992**). "The anatomy, physiology, acoustics, and perception of speech: Essential elements in analysis of the evolution of human speech," J. Hum. Evol. **23**, 447–467.

Linville, S. E., and Fisher, H. B. (**1985**). "Acoustic characteristics of perceived versus actual vocal age in controlled phonation by adult females," J. Acoust. Soc. Am. **78**, 40–8.

Martland, P., Whiteside, S. P., Beet, S. W., and Baghai-Ravary, L. (**1996**). "Estimating child and adolescent formant frequency values from adult data," Proceedings from ICSLP '96, *The Fourth International Conference on Spoken Language*, IEEE, Philadelphia, PA, pp. 626–629.

Menard, L., Schwartz, J., and Boe, J. (**2004**). "Role of vocal tract morphology in speech development: Perceptual targets and sensorimotor maps for synthesized French vowels from birth to adulthood," J. Speech Lang. Hear. Res. **47**, 1059–1080.

Menard, L., Schwartz, J., Boe, L., and Aubin, J. (**2007**). "Articulatroy-acoustic relationships during vocal tract growth for French vowels: Analysis of real data and simulations with an articulatory model," J. Phonetics **35**, 1–19.

Motoki, K. (**2002**). "Three-dimensional acoustic field in vocal-tract," Acoust. Sci. Technol. **23**, 207–212.

Nellhaus, G. (**1968**). "Head circumference from birth to eighteen years: Practical composite international and interracial graphs," Pediatrics **41**, 106–114.

Perry, T. L., Ohde, R. N., and Ashmead, D. H. (**2001**). "The acoustic bases for gender identification from children's voices," J. Acoust. Soc. Am. **109**, 2988–98.

Rao, C. R., and Toutenburg, H. (**1999**). *Linear Models: Least Squares and Alternatives* (Springer-Verlag, New York).

Robb, M. P., Chen, Y., and Gilbert, H. (**1997**). "Developmental aspects of formant frequency and bandwidth in infants and toddlers, " Folia Phoniatr Logop **49**, 88–95.

Sadler, T. W. (**2006**). *Medical Embryology* (Lippincott Williams & Wilkins, Philadelphia).

Scammon, R. E. (**1930**). "The measurement of the body in childhood," edited by J. A. Harris, C. M. Jackson, D. G. Patterson, and R. E. Scammon, *The Measurement of Man* (University of Minnesota Press, Minneapolis), pp. 173–215.

Sharkey, S. G., and Folkins, J. W. (**1985**). "Variability of lip and jaw movements in children and adults: Implications for the development of speech motor control," J. Speech Hear. Res. **28**, 8–15.

Smith, B. L. (**1994**). "Effects of experimental manipulations and intrinsic contrasts on relationships between duration and temporal variability in children's and adult's speech," J. Phonetics **22**, 155–175.

Smith, B. L., Sugarman, M. D., and Long, S. H. (**1983**). "Experimental manipulation of speaking rate for studying temporal variability in children's speech," J. Acoust. Soc. Am. **74**, 744–749.

Sperber, G. H. (**1973**). *Craniofacial Embryology* (Henry Ling Ltd., Dorchester, Great Britain).

Stevens, K. N. (**1989**). "On the quantal nature of speech," J. Phonetics **17**, 3–45.

Stone, M., Stock, G., Bunin, K., Kumar, K., Epstein, M., Kambhamettu, C., Li, M., Parthasarathy, V., and Prince, J. (**2007**). "Comparison of speech production in upright and supine position," J. Acoust. Soc. Am. **122**, 532–541.

Tingley, B. M., and Allen, G. D. (**1975**). "Development of speech timing control in children," Child Dev. **46**, 186–194.

Vorperian, H. K., Durtschi, R. B., Wang, S., Chung, M. K., Zeigert, A. J., and Gentry, L. R. (**2007**). "Estimating head circumference from imaging studies: An improved method," Acad. Radiol. **14**, 1102–1107.

Vorperian, H. K., and Kent, R. D. (**2007**). "Vowel acoustic space development in children: A synthesis of acoustic and anatomic data," J. Speech Lang. Hear. Res. **50**, 1510–1545.
Vorperian, H. K., Kent, R. D., Lindstrom, M. J., Kalina, C. M., Gentry, L. R., and Yandell, B. S. (**2005**). "Development of vocal tract length during early childhood: A magnetic resonance imaging study," J. Acoust. Soc. Am. **117**, 338–350.
Vorperian, H. K., Kent, R. D., Gentry, L. R., and Yandell, B. S. (**1999**). "Magnetic resonance imaging procedures to study the concurrent anatomic development of the vocal tract structures: Preliminary results," Int. J. Pediatr. Otorhinolaryngol. **49**, 197–206.
Walsh, B., and Smith, A. (**2002**). "Articulatory movements in adolescents: Evidence for protracted development of speech motor control processes," J. Speech Lang. Hear. Res. **45**, 1119–1133.
Wohlert, A. B., and Smith, A. (**2002**). "Developmental change in variability of lip muscle activity during speech," J. Speech Lang. Hear. Res. **45**, 1077–1087.
Xue, S. A., and Hao, G. J. (**2003**). "Changes in the human vocal tract due to aging and the acoustic correlates of speech production: A pilot study," J. Speech Lang. Hear. Res. **46**, 689–701.

Does harmonicity explain children's cue weighting of fricative-vowel syllables?[a)]

Susan Nittrouer[b)] and Joanna H. Lowenstein
Department of Otolaryngology-Head & Neck Surgery, Ohio State University, Cramblett Suite 4A, 456 West 10th Avenue, Columbus, Ohio 43210

(Received 18 June 2008; revised 3 December 2008; accepted 4 December 2008)

When labeling syllable-initial fricatives, children have been found to weight formant transitions more and fricative-noise spectra less than adults, prompting the suggestion that children attend more to the slow vocal-tract movements that create syllabic structure than to the rapid gestures more closely aligned with individual phonetic segments. That explanation fits well with linguistic theories, but an alternative explanation emerges from auditory science: Perhaps children attend to formant transitions because they are found in voiced signal portions, and so formants share a common harmonic structure. This work tested that hypothesis by using two kinds of stimuli lacking harmonicity: sine-wave and whispered speech. Adults and children under 7 years of age were asked to label fricative-vowel syllables in each of those conditions, as well as natural speech. Results showed that children did not change their weighting strategies from those used with natural speech when listening to sine-wave stimuli, but weighted formant transitions less when listening to whispered stimuli. These findings showed that it is not the harmonicity principle that explains children's preference for formant transitions in phonetic decisions. It is further suggested that children are unable to recover formant structure when those formants are not spectrally prominent and/or are noisy. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3056561]

PACS number(s): 43.71.An, 43.71.Ft [MSS] Pages: 1679–1692

I. INTRODUCTION

The primary focus of the study of human speech perception has traditionally been on the question of which signal properties listeners use in making phonetic judgments. An early example of this focus was a study published by Harris (1958) demonstrating that adult, native speakers of English base their decisions about the sibilant fricatives /s/ and /ʃ/ primarily on the spectral structure of the noise itself, but base their decisions about the weak fricatives /f/ and /θ/ on the pattern of formant transitions in the voiced syllable portion. That experiment was done using natural, spliced speech samples, but much of the work on human speech perception has been conducted with synthetic speech. Typically a single acoustic property was manipulated along a continuum spanning settings that evoke one phonetic judgment or another, assessing how those settings affected the choice of which phoneme was heard. The properties manipulated in such experiments came to be known as "acoustic cues" (Repp, 1982).

Although generally informative, one flaw in this approach was that acoustic evidence clearly showed, even then, that the qualities of many discrete regions of the signal varied as a function of surrounding phonetic structure. For example, the noise spectrum of syllable-initial /s/ or /ʃ/ differs depending on the vowel that follows (Kunisaki and Fujisaki, 1977; Heinz and Stevens, 1961). It was reasonable that the effects of this sort of variation on phonetic decisions should be examined, and so experimental paradigms were adjusted to incorporate a second acoustic cue, manipulated in a dichotomous manner, that is, set to be appropriate for the phoneme represented at one end or the other of the acoustic continuum formed by manipulating the first cue. For example, Mann and Repp (1980) found that when voiced formant transitions were set to be appropriate for either a preceding alveolar or palatal constriction, the /ʃ/-/s/ phoneme boundary on an acoustic continuum was shifted. This sort of interaction was termed cue trading or trading relations. These sorts of results showed that the relation between acoustic cues and phonetic judgments is more complicated than at first envisioned.

Later Nittrouer and Studdert-Kennedy (1987) set out to explore whether or not children (3–7 years old) exhibit the sort of cue trading reported by Mann and Repp (1980). For that purpose fricative-vowel syllables similar to those of Mann and Repp (1980) were constructed with synthetic noises spanning a spectral /ʃ/-to-/s/ continuum, and vocalic portions taken from natural speech samples with formant transitions appropriate for either an initial /ʃ/ or /s/. Within this design the spectral structure of the fricative noise was presumed to be the "primary" cue, defined as the one that would be given the greatest perceptual weight in the fricative decision. That designation followed from the work of Harris (1958) showing that indeed adult listeners primarily base their sibilant decisions on characteristics of the noise itself. It was presumed that the role of formant transitions in this decision would be secondary, and likely learned through years of experience hearing the patterns of formant change arising from the vocal tract moving from a fricative target to a vowel

[a)]Portions of this work were presented at the 154th Meeting of the Acoustical Society of America, New Orleans, LA, November 2007.
[b)]Electronic mail: nittrouer.1@osu.edu

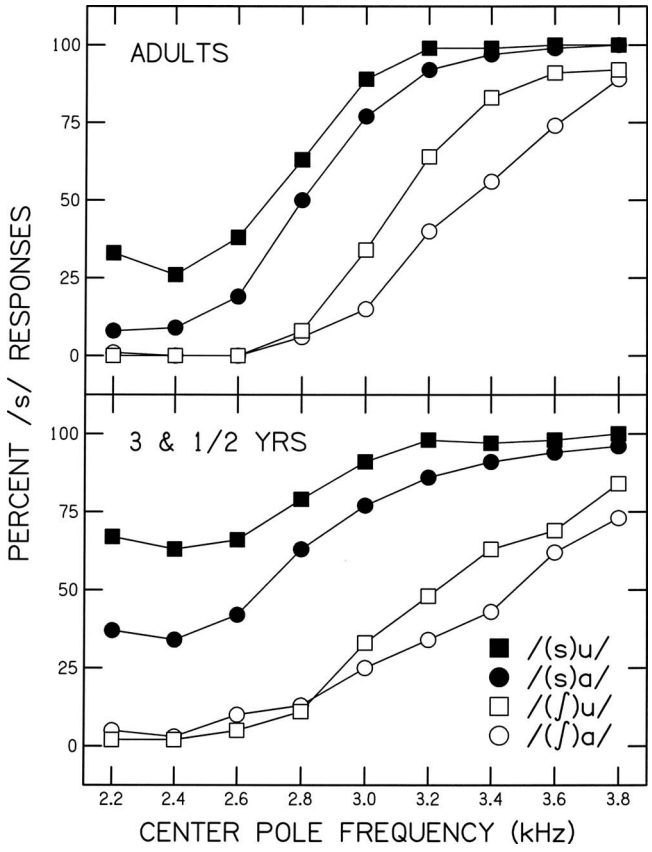


FIG. 1. Labeling functions for adults (top) and $3\frac{1}{2}$-year-olds (bottom) from Nittrouer (1992). Center frequencies of the synthetic fricative noises are shown on the abscissa, and percents responses are shown on the ordinate. A separate function is plotted for each syllable type.

target (Stevens, 1975). The specific prediction was that the youngest children would show a greatly attenuated trading of cues, with a developmental increase in the magnitude of the trade. In fact, precisely the opposite pattern of results was obtained: The youngest listeners showed the greatest influence of formant transitions on fricative labeling, with a developmental decrease in effect. These findings were supported by subsequent research, which demonstrated that a developmental change occurs in the relative amount of attention paid to these acoustic cues for children between roughly 3 and 7 years of age (Mayo *et al.*, 2003; Nittrouer, 1992; Nittrouer and Miller, 1997a, 1997b). Figure 1, from Nittrouer (1992), illustrates this trend: Children's labeling functions are more separated as a function of whether formant transitions are appropriate for a preceding /s/ or /ʃ/ closure. This pattern indicates that children's responses were more strongly based on vocalic formant transitions than were those of adults. At the same time, children's functions are shallower than those of adults, indicating that they paid less attention to the spectra of fricative noises in making their decisions. This developmental change in the amount of attention, or weight, given to separate acoustic cues came to be known as the "developmental weighting shift" (e.g., Nittrouer, 1996; Nittrouer *et al.* 1993; Nittrouer and Miller, 1997a, 1997b).

The finding that children weight acoustic cues differently from adults in their linguistic decision-making complemented emerging evidence that speakers, and so listeners, of different languages weight acoustic cues differently (e.g., Beddor and Strange, 1982; Crowther and Mann, 1992, 1994; MacKain *et al.*, 1981). The way that the acoustic speech signal is shaped by gestures of the vocal tract in order to provide for the recognition of phonetic structure varies across languages (e.g., Crowther and Mann, 1992; Flege and Port, 1981; Lisker and Abramson, 1964). A primary chore of the child learning a language is to discover which properties of the signal provide linguistically significant variation, and so must be attended to, and which properties contain variation that is irrelevant to deriving meaning, and so may safely be ignored. It is explicitly this language-related discovery process that the developmental weighting shift seeks to capture. But while this notion may explain why weighting strategies change through childhood, an unresolved question concerns why children's initial strategies are the way they are. Why would initial weighting strategies be so dependent on formant transitions within the voiced portion of the signal? The answer to this question has theoretical and clinical implications.

Several related explanations have been offered for why children pay particular attention to formant transitions in their phonetic decisions. In general it has been speculated that these relatively slow modulations in spectral structure (i.e., formant transitions) attract children's attention early in development precisely because they are dynamic. Research into the development of visual perception shows that children attend more to moving properties than to static properties in object recognition (Ruff, 1982; Spelke *et al.*, 1989). It is reasonable that they would do the same in auditory perception. Another possible explanation for children's early weighting strategies has a purely linguistic basis: It is commonly asserted that young children use more global processing strategies than adults to recognize words and syllables (e.g., Cole and Perfetti, 1980; Charles-Luce and Luce, 1990, 1995; Jusczyk, 1992; Nittrouer, 1996; Walley *et al.*, 1986). Although exactly what the term "global processing strategies" is referring to has not always been specified in these accounts, it is reasonable to suggest that the term references children's enhanced attention to signal properties that are syllabic in nature, rather than specific to individual phonemes. The idea that children pay special attention to formant transitions is compatible with the notion. Finally, it has been suggested that these formant transitions reflect the general movements of the vocal tract. In production, children appear to master general articulatory placements and patterns of movement between these targets before they learn how to make precise constriction shapes (e.g., Goodell and Studdert-Kennedy, 1993; MacNeilage and Davis, 1991; McGowan and Nittrouer, 1988). In sum, several hypotheses have been offered to explain children's preferential weighting of formant transitions.

Yet another hypothesis for why children focus so strongly on formant transitions emerges from the field of general auditory perception. To appreciate this hypothesis, it is necessary to first ask ourselves why variation across formant frequencies would provide the listener with any information at all. There is no reason to expect three disparate

spectral components, as these formant frequencies are, to cohere, providing a single perceptual impression. So, why do listeners hear these separate but co-occurring patterns as one entity, providing information about the structure of the signal? Again, different explanations have been offered to answer that question, with one set of explanations involving strictly auditory principles. This set of explanations is best exemplified by the work of Bregman (1990) on Auditory Scene Analysis (ASA). These principles suggest that separate spectral components cohere if they have synchronous onsets and offsets, or if they have the same harmonic structure, a principle known as harmonicity. Because formant transitions in fricative-vowel syllables come from the voiced portion of the syllable they share a harmonic structure, and so the principle of harmonicity may explain the coherence of formants in speech perception.

Another kind of explanation for why formant transitions may cohere perceptually is that they all arise from a single vocal tract, a fact that is apparent to the listener. This position is best explained and supported by Remez *et al.* (1994) who methodically tested each principle of ASA and serially rejected each as an explanation for why the various components of the speech signal cohere. Of course, those experiments were done with adult listeners. In this current study, we entertained the possibility that children, being less experienced with speech, may require that signals strictly adhere to ASA principles if they are to integrate separate spectral portions and use them for speech perception. Perhaps children's greater weighting of formant transitions in fricative-vowel syllables, compared to that of adults, arises from nothing more than the fact that these signal portions adhere to ASA principles. In the work reported here, the specific possibility examined was that children may attend strongly to formant transitions in voiced signal portions because they have a preference for signals that adhere to the harmonicity principle.

To test this possibility, stimuli were needed that retained formant transitions but eliminated harmonic structure. This requirement necessitated that speech signals be processed in some way, which is always a risky business. While there are many ways to modify a speech signal to eliminate harmonicity, each creates concomitant acoustic changes that may affect perception in unpredictable ways. For that reason, two processing algorithms were used in this study: Both disrupted the principle of harmonicity, but each did so in different ways. Sine-wave speech was used because it lacks harmonic structure, yet preserves the dynamic information provided by formant transitions. Whispered speech was used because it also lacks harmonic structure and preserves formant transitions. It is, in addition, familiar to young children, and thus not as "artificial" as sine-wave speech might be viewed as being.

Sine-wave speech uses three sinusoidal tones to replicate the first three formants of speech. Since the stimuli consist of sine waves, there are no harmonic relations between the components. There is ample evidence that adult listeners can understand sine-wave replicas of sentences when they have been instructed that what they are listening to is speech (Barker and Cooke, 1999; Remez *et al.* 1981; Remez *et al.*, 1994), but results for children are mixed. A recent study by Vouloumanos and Werker (2007) reported that 1- to 4-day-old infants preferred natural speech stimuli to sine-wave analogs of the same stimuli. Older children, however, have been shown to comprehend sine-wave sentences as well as adults (Nittrouer *et al.*, 2009). In that latter experiment adults and 7-year-olds were asked to repeat four-word sentences presented as sine-wave analogs or as 4- or 8-channel noise vocoded analogs. The 7-year-olds performed similarly to adults for the sine-wave stimuli, but much more poorly for the noise vocoded stimuli. Serniclaes *et al.* (2001) presented sine-wave analogs of stimuli from a /bɑ/-/dɑ/ continuum to dyslexic and normal-reading 13-year-olds as a discrimination task. The stimuli were first presented as nonspeech whistles, and then presented as speech. When the sine-wave syllables were presented as speech, both the dyslexics and normal readers showed a phoneme boundary effect that was absent when the sine-wave syllables were presented as nonspeech whistles. Thus the evidence indicates that by 7 years of age, at the latest, children treat sine-wave analogs as speech.

Whispering is something that most children are familiar with. In whispered speech, turbulent noise produced by air flow through partially adducted vocal folds is shaped by the supralaryngeal vocal tract, resulting in speech that is very understandable, but that lacks vocal fold vibrations (and thus a fundamental). Magnetic resonance imaging and endoscopic studies of whispered vowels provide evidence that during whispering the supraglottal structures shift downwards to inhibit vocal fold vibration (Matsuda and Kasuya, 1999; Tsunoda *et al.*, 1997). Acoustic studies of whispered speech have shown that formant frequencies, particularly F1, tend to be higher in whispered than in voiced vowels (Kallail and Emanuel, 1984a, 1984b). Formant bandwidths appear broader in whispered than in voiced speech, as well. In terms of perception, Tartter (1989) found that untrained adults could accurately label consonant-vowel syllables (which consisted of the vowel [ɑ] preceded by voiced and voiceless labial, alveolar, and velar stops, nasals, liquids, and fricatives) 64% of the time, with the primary confusions being for normally voiced consonants. Whispered vowels in [hVd] context were identified correctly 82% of the time (Tartter, 1991). In sum we would expect recognition of unvoiced consonant-vowel syllables to be highly accurate for whispered speech.

Although sine-wave analogs and whispered speech both provide ways of examining speech perception when the harmonicity principle is violated, they differ from each other in ways that should provide useful insights beyond the central question of this study. For example, sine waves have effective bandwidths of 1 Hz, and so the spacing between formants is much greater than in natural speech: Sine-wave formants are quite prominent. Whispered speech, on the other hand, has broadened bandwidths, and so formants are not as prominent as in natural speech. Investigators have worried in the past that the broader auditory filters associated with hearing loss might make it hard for listeners with hearing loss to resolve individual formants, and so to understand speech (Summerfield *et al.*, 1981; Turner and Holte, 1987; Turner and Van Tasell, 1984). Although that work with adults

showed that it is generally not a problem, the possibility exists that children might have a stronger need for more spectrally prominent formants. Evidence for this situation would be obtained if children showed no change from natural speech in their weighting of formant transitions when presented with sine-wave stimuli, but had difficulty with whispered speech.

Another difference between the two kinds of signals used in this study was that when whispered vocalic portions were appended to synthetic fricative noises, the result was signals consisting completely of noise. Previous studies examining perceptual weighting strategies for fricative-vowel stimuli not only found that children give more weight to formant transitions than adults in their fricative decisions but that there is a reciprocal decrease in the weight given to fricative-noise spectra (Nittrouer and Miller, 1997a, 1997b; Nittrouer, 1992; Nittrouer and Studdert-Kennedy, 1987). It may simply be that children are not as skilled at using information in noisy signal portions as adults are.

We stand to improve our understanding of the nature of perceptual development in general by exploring the questions asked by this work. In addition, understanding the basis of children's perceptual preference for dynamic signal components in speech perception may affect what we do in fitting auditory prostheses for children with hearing loss. Pediatric audiologists worry that fitting children with hearing aids using the same algorithms as those used with adults is not appropriate (e.g., Pittman *et al.*, 2003), and, in particular, they worry about preserving children's abilities to recognize sibilants because of their importance in grammatical marking (Stelmachowicz *et al.*, 2007). Therefore, it is appropriate to ask why children show the perceptual weighting strategies that they do, and what conditions support or harm their abilities to use these strategies. Results showing either that children require formants to be spectrally prominent or that they avoid noisy signal components would provide useful information for the future design and fitting of auditory prostheses.

In summary, the two experiments in this study were conducted to test the hypothesis that children greatly weight formant transitions because they share harmonic structure. Two types of stimuli that lack harmonic structure were used: sine-wave and whispered speech. If children were found to pay less attention to formant transitions with both kinds of nonharmonic stimuli, support would be provided for the hypothesis that children attend especially to signal portions that adhere to ASA principles. On the other hand, if children were found to perform similarly with natural speech and both kinds of nonharmonic stimuli, that finding would provide evidence that children attend to these signal portions for other reasons, likely the language-related reasons described above. If children demonstrated modifications in their perceptual weighting strategies for one, but not the other, processed signal, we would have to conclude that it was some concomitant change in signal structure that accounted for the perceptual change, rather than the lack of harmonicity.

II. EXPERIMENT 1: VOICED AND SINE-WAVE VOCALIC PORTIONS

The primary goal of this experiment was to test the hypothesis that children's strong weighting of formant transitions in phonetic judgments of syllable-initial fricatives is accounted for by the fact that voiced syllable portions adhere to the harmonicity principle, whereas noise syllable portions do not. Support for this hypothesis would be obtained if children decreased the weight assigned to formant transitions when the harmonicity principle was disrupted by creating sine-wave analogs of the voiced syllable portions.

A. METHOD

1. Subjects

Twenty adults between the ages of 18 and 40 years, 20 7-year-olds, 22 5-year-olds, and 23 3-year-olds participated. The mean age of 7-year-olds was 7 years, 1 months (7;1), with a range from 6;11 to 7;5. The mean age of 5-year-olds was 5;1, with a range from 4;11 to 5;5. The mean age of 3-year-olds was 3;9, with a range of 3;5 to 3;11. None of the listeners (or their parents, in the case of children) reported any history of hearing or speech disorder, and they all seemed typical to the experimenter doing the testing. Nonetheless, objective criteria were used to document that they were indeed typical. In particular, all listeners passed hearing screenings consisting of the pure tones of 0.5, 1, 2, 4, and 6 kHz presented at 25 dB HL to each ear separately. Children were given the Goldman Fristoe 2 Test of Articulation (Goldman and Fristoe, 2000) to screen for speech problems and were required to score at or better than the 30th percentile for their age. An additional constraint on these scores was that no child could exhibit errors in producing /ʃ/ or /s/. All children were free from significant histories of otitis media, defined as six or more episodes during the first three years of life. Adults were given the reading subtest of the Wide Range Achievement Test-Revised (Jastak and Wilkinson, 1984) as a way to screen for language problems and were required to demonstrate at least an 11th-grade reading level.

2. Equipment and materials

All testing took place in a soundproof booth, with the computer that controlled the experiment in an adjacent room. Hearing was screened with a Welch Allyn TM262 audiometer using TDH-39 headphones. Stimuli and stories were stored on a computer and presented through a Creative Labs Soundblaster card, a Samson headphone amplifier, and AKG-K141 headphones. Two drawings (on 8×8 in.^{2} cards) were used to represent each response label in each experiment, such as "Sue" (a girl) and a "shoe," or "Sa" (a space alien) and "Sha" (a king). Gameboards with ten steps were also used with children; they moved a marker to the next number on the board after each block of test stimuli. Cartoon pictures were used as reinforcement and were presented on a color monitor after each block of stimuli. A bell sounded while the pictures were being shown and served as additional reinforcement.

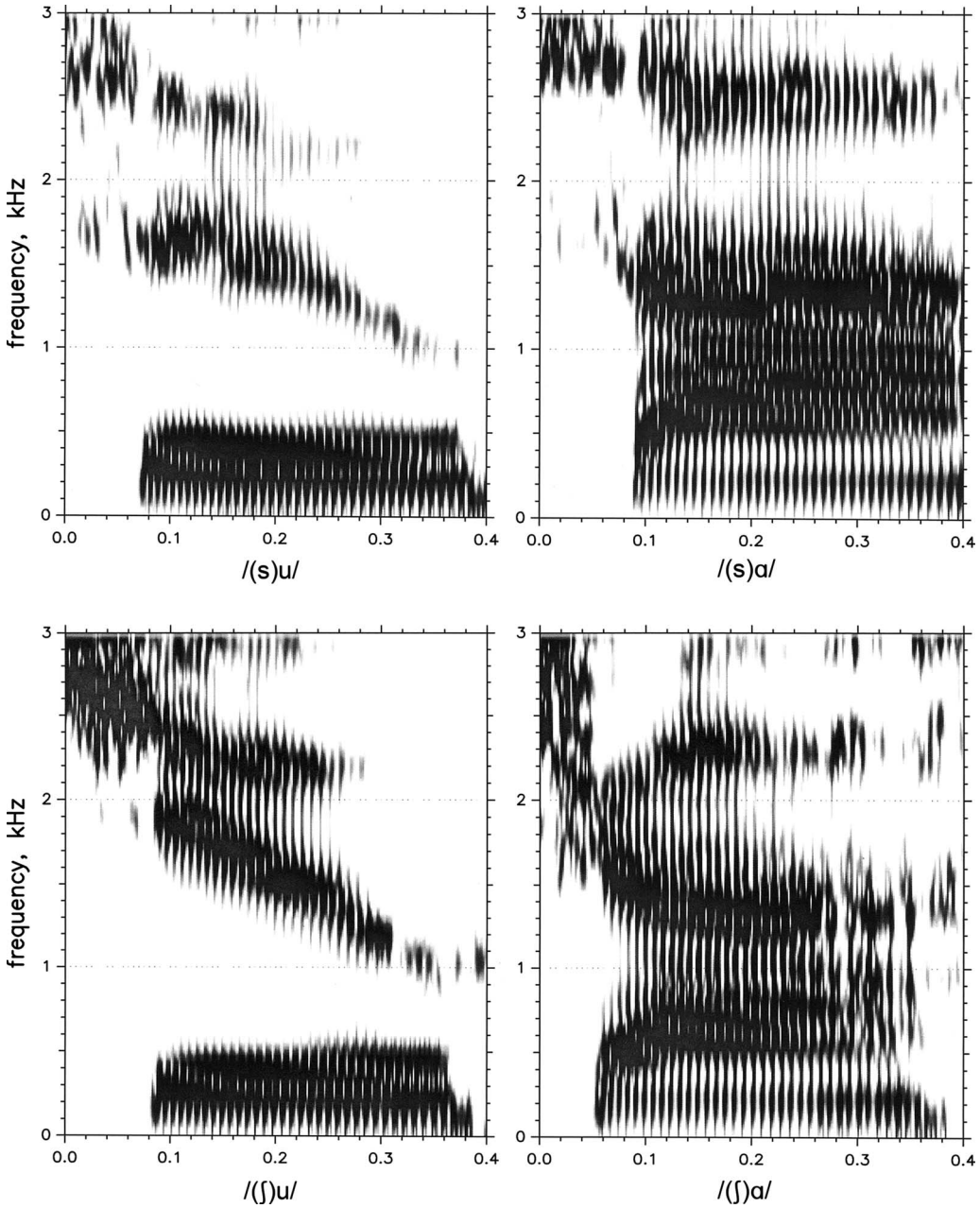


FIG. 2. Spectrograms displaying examples of the natural vocalic portions used in the VOICED stimuli of experiments 1 and 2. The fricative label in parentheses indicates the fricative context from which the vocalic portion was separated.

3. Stimuli

Thirty-six separate stimuli were created for each of the VOICED and SINE-WAVE conditions: nine fricative noises ranging along a continuum from one that is readily identified as /ʃ/ to one that is readily identified as /s/, combined with each of four vocalic portions. These vocalic portions differed in vowel quality (/ɑ/ or /u/) and in whether formant transitions at the start of the portion were appropriate for a preceding /ʃ/ or /s/. Design of stimuli was based on stimuli that have been used previously in several experiments (Nittrouer 1992, 1996; Nittrouer and Miller, 1997a), and so we had general notions of what to expect in responses from children and adults. The VOICED stimuli in this experiment were exactly those used in the earlier experiments. In these stimuli, the fricative portions were single-pole, synthetic noises, with the center frequency ranging from 2.2 to 3.8 kHz in nine 200 Hz steps. The vocalic portions of these VOICED stimuli were taken from a male speaker saying /su/, /sa/, /ʃu/, and /ʃɑ/. The recordings were sampled at 22.05 kHz and low-pass filtered below 11.025 kHz. Three tokens of each syllable were used, and Fig. 2 shows spectrograms of one of each kind of vocalic portion. Table I displays mean values for some phonetically relevant properties of the vocalic portions. For all /u/ portions, F1 remained stable throughout the vocalic portion, and F2 fell in frequency through the entire portion to a mean ending frequency of 930 Hz. For all /ɑ/ vocalic portions, F1 rose in frequency over roughly the first 60 ms to a mean steady-state frequency of 703 Hz, and F2 fell in frequency over the first 60–90 ms to a mean steady-state frequency of 1309 Hz. For all stimuli,

TABLE I. Means across the three tokens for the phonetically relevant acoustic properties of the vocalic portions of the VOICED stimuli used in experiments 1 and 2.

Portion	f_0 (Hz)	Dur (ms)	F1-onset (Hz)	F2-onset (Hz)	F3-onset (Hz)
(s)ɑ	93	333	457	1365	2457
(ʃ)ɑ	95	337	379	1532	2367
(s)u	99	347	320	1520	2496
(ʃ)u	97	348	308	1706	2288

F3 fell in frequency over roughly the first 100 ms to a mean steady-state value of 2320 Hz for /u/ and 2365 Hz for /ɑ/.

Natural vocalic portions were separated from the fricative noises using a waveform editor, and recombined with each of the synthetic fricative noises, yielding 36 kinds of VOICED stimuli (9 noises × 4 vocalic portions), with 3 tokens of each. The center panel of Fig. 3 shows a spectrogram of a representative VOICED /(ʃ)ɑ/ stimulus with the 3.0 kHz noise. RMS amplitude was equal for the noise and vocalic syllable portions.

The SINE-WAVE stimuli used the same synthetic fricative noises as the VOICED stimuli. The vocalic portions were created based on recordings of natural /su/, /sɑ/, /ʃu/, and /ʃɑ/ from the male speaker. Three sine waves were used to approximate the first three formants of each vowel and were adjusted until they most closely resembled the natural vowels in Fig. 2. The SINE-WAVE stimuli were all 250 ms in length and were generated at a sampling rate of 22.05 kHz and low-pass filtered at 11.025 kHz using TONE software (Tice and Carrell, 1997). The sine-wave vowels were combined with each of the synthetic noises, resulting in 36 SINE-WAVE stimuli (9 noises × 4 sine-wave vowels). The left panel of Fig. 3 shows a spectrogram of a representative SINE-WAVE /(ʃ)ɑ/ stimulus with the 3.0 kHz noise. RMS amplitude was equal for the noise and sine-wave vocalic portions.

4. Procedures

The screening tasks were presented first, followed by the labeling task. Stimuli with each vowel (/u/ or /ɑ/) and stimulus type (VOICED or SINE-WAVE) were presented separately, making four conditions. Adults, 7-year-olds, and 5-year-olds participated in two test sessions. For these listeners order of presentation of conditions was randomized across listeners, with the restrictions that conditions with the same vowel or stimulus type were not presented in the same session and that order of presentation of the VOICED and SINE-WAVE conditions be different across sessions. For example, if the first condition, randomly chosen, consisted of VOICED /u/ stimuli, the second condition presented at that session would necessarily consist of SINE-WAVE /ɑ/ stimuli. At the next session the listener would be presented with SINE-WAVE /u/, followed by VOICED /ɑ/. The same ordering procedures were followed for 3-year-olds, except that they were spread out over four test sessions because pilot testing had revealed that 3-year-olds were more likely to complete testing if it was spread out like that.

During testing, stimuli were presented ten times each, in blocks of 18. Because there were three tokens of each VOICED stimulus, the program randomly selected one of the tokens to present during the first block, and then repeated this random selection during the next block without replacement. After the first three blocks, this process was repeated until testing was done. The listener's task during testing was

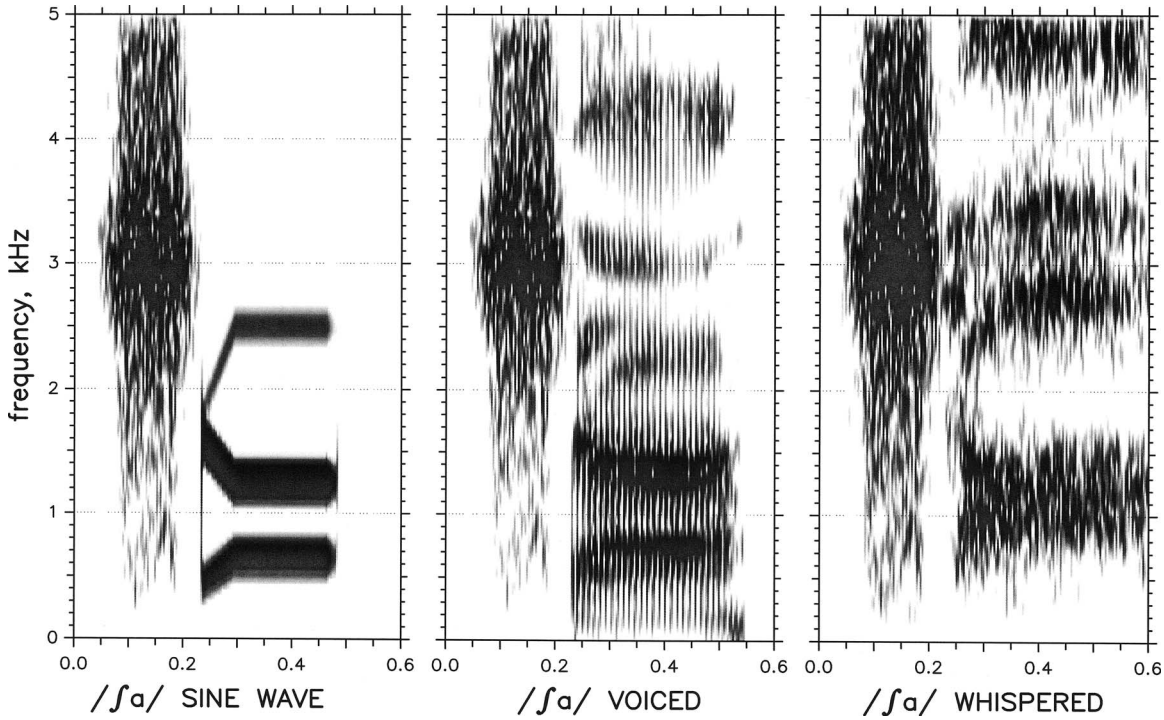


FIG. 3. Spectrograms displaying examples of the SINE-WAVE, VOICED, and WHISPERED /ʃɑ/ stimuli from experiments 1 and 2.

to point to the picture representing their response choice, and say the word that corresponded to that choice.

Children were introduced to the response labels (sa, sha, sue, or shoe) before testing with brief, digitized stories accompanied by pictures. Each story was presented once with natural speech for the VOICED condition. For the SINE-WAVE condition, each story was presented twice, once with natural speech and once with sine-wave speech. These stories served both to familiarize children with the response labels and to provide experience listening to sine-wave speech. During pilot testing it was found that presenting these stories to adults did not affect their responding. Consequently, we did not play them to adults.

All listeners heard 12 practice items (6 each of /sV/ and /ʃV/) before testing, using the best exemplars of each category (i.e., the syllable consisting of the lowest-frequency fricative noise combined with a vocalic portion with /ʃ/ formant transitions or the syllable consisting of the highest-frequency fricative noise combined with a vocalic portion with /s/ formant transitions). Each listener had to label 11 of the 12 best exemplars correctly to proceed to testing. This criterion ensured that all listeners could perform the required task, and knew the response labels. 3-year-olds had one additional familiarization task: Before hearing the practice items, they were presented with completely natural tokens of each syllable and had to label 11 of 12 correctly to proceed to the next practice. This extra step ensured that 3-year-olds recognized natural exemplars of these stimuli correctly. When it came to testing, listeners were required to give at least 80% accurate responses to the best exemplars, thereby ensuring that data were included only from listeners who maintained attention to the task.

To summarize, slightly different procedural steps were followed for each group: Adults received practice with the best exemplars, followed by testing. Children who were 5- or 7-years-old heard stories, followed by practice with the best exemplars, and then testing. The 3-year-olds heard the stories, followed by practice with natural tokens, then practice with the best exemplars, and finally testing.

For analysis purposes, partial correlation coefficients were computed between the proportion of "s" responses and the acoustic properties of formant transitions (appropriate for /s/ or /ʃ/) and fricative noise (step on the continuum), across vowels. These correlation coefficients served as weighting coefficients, explaining the proportion of variance in response labels associated with formant transitions and with fricative-noise spectra (Nittrouer, 2002).

B. Results

1. VOICED condition

Five 3-year-olds were dismissed because they did not meet the criterion of getting 11 of the 12 syllables correct during practice with completely natural tokens. All of the remaining 18 3-year-olds were able to complete the VOICED /u/ condition with 80% accurate responses on best exemplars or better, but only 12 of them were able to complete the VOICED /ɑ/ condition.

Figure 4 displays the labeling functions for adults, 7-year-olds, 5-year-olds, and 3-year-olds for the VOICED condition. Adults and 7-year-olds performed similarly. 3-year-olds and 5-year-olds showed greater separation in functions based on whether the vocalic portion originally started with /ʃ/ or /s/, and generally shallower functions than adults or 7-year-olds. These two age-related differences in response patterns reflect greater weighting of formant transitions and lesser weighting of fricative-noise spectra, respectively, by the younger listeners (Nittrouer, 2002). When Fig. 4 is compared to Fig. 1 it is seen that response patterns for adults and 3-year-olds in this experiment are very similar to what has been found in previous studies (Nittrouer, 1992, 1996; Nittrouer and Miller, 1997a, 1997b).

Table II shows mean weighting coefficients for the VOICED condition. The response patterns observed in Fig. 4 are apparent in these coefficients: 3- and 5-year-olds weighted transitions more than adults or 7-year-olds, and weighted the fricative noises less. When these weighting coefficients are compared to weighting coefficients derived from responses of listeners hearing the same stimuli in the past (e.g., Nittrouer and Miller, 1997b), very similar results are obtained: Weighting coefficients of formant transitions for adults, 7-year-olds, and $4\frac{1}{2}$-year-olds were, respectively, 0.32, 0.36, and 0.46; for fricative noises weighting coefficient were 0.82, 0.77, and 0.68 (Nittrouer, 2002). Thus, we are assured that the listeners in this experiment are similar to those in earlier experiments.

One-way analyses of variance (ANOVAs) with age as the main effect and *post hoc* comparisons were done on weighting coefficients for formant transitions and fricative noises obtained in this experiment. Results are shown in Table III. Here, and in the remainder of this paper, precise statistical results are given only when $\rho < 0.10$. Otherwise, results are reported as not significant (ns). These results show significant overall age effects for both formant transitions and fricative noises. The *post hoc* *t*-tests confirm that 3-year-olds and 5-year-olds weighted formant transitions more and fricative noises less than 7-year-olds or adults. It is also observed that the four age groups form a dichotomy in response patterns, such that the two youngest groups did not differ in weighting strategies from each other, and the two oldest groups did not differ from each other.

2. SINE-WAVE condition

Three-year-olds found the SINE-WAVE condition very difficult: Only 8 out of 18 of them (44%) responded to the SINE-WAVE practice items and best exemplars during testing at levels accurate enough to have their data included in analyses. When less than half of the listeners in a group are

TABLE II. Mean weighting coefficients in the VOICED condition of experiment 1. Standard deviations are in parentheses.

	Adults	7-year-olds	5-year-olds	3-year-olds
Transition	0.30 (0.17)	0.32 (0.15)	0.48 (0.19)	0.50 (0.15)
Noise	0.82 (0.11)	0.81 (0.11)	0.71 (0.14)	0.66 (0.13)

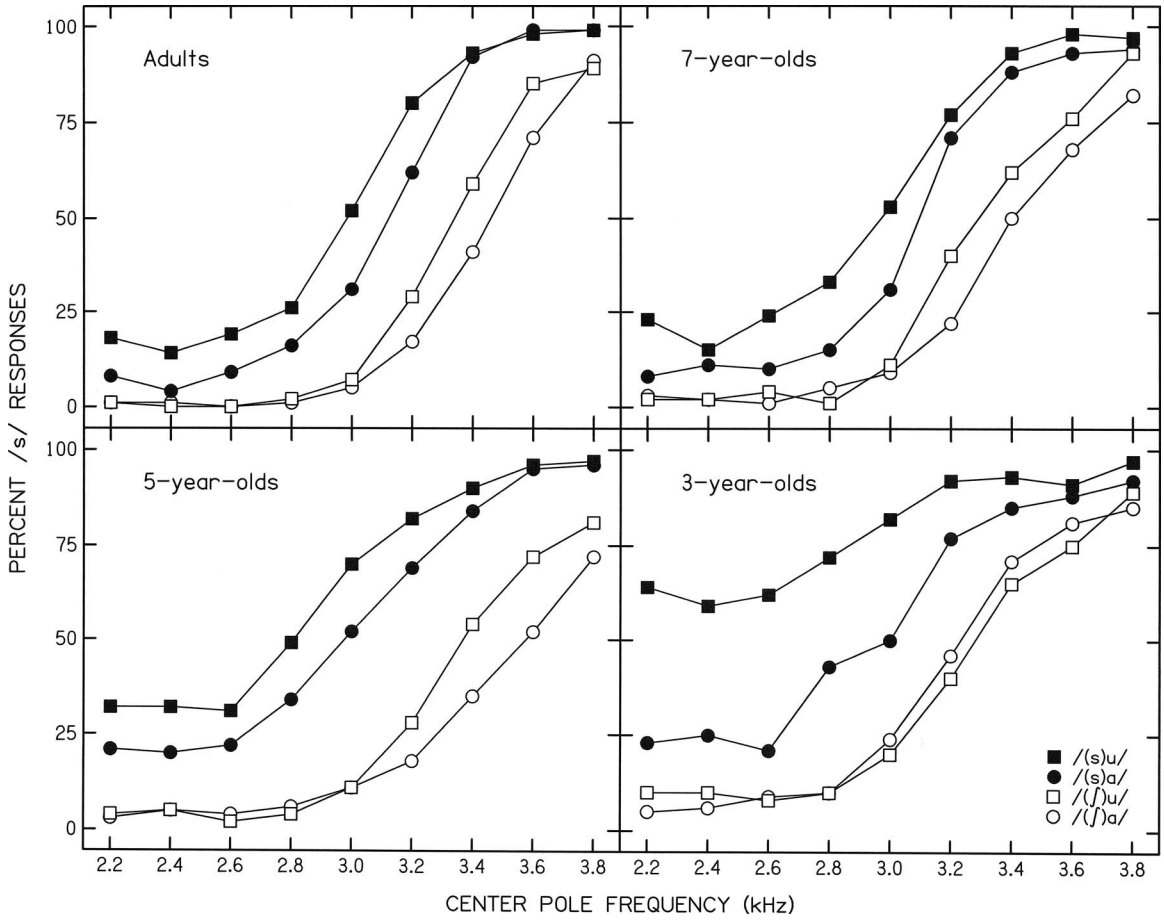


FIG. 4. Labeling functions for the VOICED condition of experiment 1 for adults, 7-year-olds, 5-year-olds, and 3-year-olds. See legend for Fig. 1 for more details.

able to perform a task, their performance on that task cannot be taken as representative of the group as a whole. Therefore, data from the 3-year-olds were not analyzed for this condition. On the other hand, 80% of 7-year-olds and 75% of 5-year-olds were able to complete the SINE-WAVE condition. Data from these listeners were analyzed.

Table IV shows mean weighting coefficients for the SINE-WAVE condition. In comparing the numbers shown on Table IV to those on Table II, the greatest difference appears for adults: They seem to have weighted the fricative noises similarly, but to have weighted formant transitions less in this condition than in the VOICED condition. To a lesser extent it appears as if 5-year-olds weighted fricative noises less in this SINE-WAVE condition than in the VOICED condition.

ANOVAS, with *post hoc t*-tests, were done on weighting coefficients for formant transitions and fricative noises separately, and results are shown in Table V. There was a significant age effect for both transitions and fricative noises. The *post hoc t*-tests confirm that 5-year-olds continued to weight formant transitions more and fricative noises less in their

TABLE III. Results of ANOVAs and *post hoc t* tests done on weighting coefficients in the VOICED condition of experiment 1. Bonferroni significance levels indicate that the contrast is significant at the level shown when corrections are made for multiple tests.

	df	*F* or *t*	*p*	Bonferroni significance
		Transition		
Age effect	3,76	8.11	<0.001	
Adults vs 7-year-olds	76	0.38	ns	
Adults vs 5-year-olds	76	3.53	<0.001	0.01
Adults vs 3-year-olds	76	3.80	<0.001	0.01
7- vs 5-year-olds	76	3.14	0.002	0.05
7- vs 3-year-olds	76	3.46	0.001	0.01
5- vs 3-year-olds	76	0.46	ns	
		Noise		
Age effect	3,76	7.61	<0.001	
Adults vs 7-year-olds	76	−0.21	ns	
Adults vs 5-year-olds	76	−2.79	0.007	0.05
Adults vs 3-year-olds	76	−3.98	<0.001	0.001
7- vs 5-year-olds	76	−2.58	0.02	0.10
7- vs 3-year-olds	76	−3.78	<0.001	0.01
5- vs 3-year-olds	76	−1.35	ns	

TABLE IV. Mean weighting coefficients in the SINE-WAVE condition of experiment 1. Standard deviations are in parentheses.

	Adults	7-year-olds	5-year-olds
Transition	0.21 (0.10)	0.31 (0.16)	0.50 (0.19)
Noise	0.85 (0.06)	0.80 (0.12)	0.65 (0.20)

TABLE V. Results of ANOVAs and *post hoc* *t*-tests done on weighting coefficients in the SINE-WAVE condition of experiment 1. Bonferroni significance levels indicate that the contrast is significant at the level shown when corrections are made for multiple tests.

	df	*F* or *t*	*p*	Bonferroni significance
		Transition		
Age effect	2,59	20	<0.001	
Adults vs 7-year-olds	59	2.16	ns	
Adults vs 5-year-olds	59	6.21	<0.001	0.001
7- vs 5-year-olds	59	4.00	<0.001	0.001
		Noise		
Age effect	2,59	12.50	<0.001	
Adults vs 7-year-olds	59	−1.03	ns	
Adults vs 5-year-olds	59	−4.73	<0.001	0.001
7- vs 5-year-olds	59	−3.68	<0.001	0.01

fricative decisions than 7-year-olds or adults. These older two groups performed similarly to each other, as they had for the VOICED stimuli.

3. Across-conditions comparison

Of course, the major focus of this experiment was to examine whether listeners, particularly children, modify their weighting strategies when stimuli lack harmonicity. For that purpose, weighting coefficients obtained for the VOICED and SINE-WAVE stimuli were compared for each group separately, using matched *t*-tests. Results are shown in Table VI. These tests confirm that 7-year-olds assigned similar weighting coefficients to both properties in each condition. The tests further confirm that adults weighted formant transitions less in the SINE-WAVE than in the VOICED condition. Results for 5-year-olds revealed that they weighted fricative noises slightly less in the SINE-WAVE than in the VOICED condition, but the difference fell just short of statistical significance.

C. Discussion

The purpose of this experiment was to test the hypothesis that listeners would change their perceptual weighting strategies when sine-wave replicas replaced natural vocalic portions, consequently eliminating the harmonicity of the stimuli. In particular, we wanted to examine whether children would show diminished weighting of formant transitions. Clearly that was not found: 7-year-olds' weighting of the two acoustic cues examined were consistent across testing with these stimulus types, and 5-year-olds decreased the weight they assigned to the fricative noises only (not to the formant transitions) in the SINE-WAVE condition, compared to the VOICED condition. Consequently, no support was garnered for the hypothesis that children weight formant transitions strongly in their phonetic decisions regarding syllable-initial fricatives because they adhere to the harmonicity principle. In fact, it is interesting that adults were the only listeners to decrease the amount of weight they gave to formant transitions in making these decisions when the vocalic portions were sine waves.

The results reported here were obtained in spite of the fact that sine-wave stimuli are very unnatural in quality. Regardless, children showed similar weighting of formant transitions for both sets of stimuli.

TABLE VI. Results of matched *t*-tests done on within-group weighting coefficients across the VOICED and SINE-WAVE conditions for experiment 1.

	df	*t*	*p*
	Transition		
Adults	19	2.16	0.04
7-year-olds	19	0.16	ns
5-year-olds	21	−0.80	ns
	Noise		
Adults	19	−1.18	ns
7-year-olds	19	0.16	ns
5-year-olds	21	1.89	0.07

III. EXPERIMENT 2: VOICED AND WHISPERED VOCALIC PORTIONS

As with experiment 1, the purpose of this second experiment was to test the hypothesis that children's strong weighting of formant transitions in judgments of syllable-initial fricatives is explained by the fact that voiced syllable portions, where formant transitions reside, adhere to the harmonicity principle. Unlike experiment 1, removing harmonicity from vocalic syllable portions in this second experiment meant creating syllables that consisted entirely of noise. This was accomplished by presenting stimuli with whispered vocalic portions. This modification to the speech signal meant that the stimuli were more natural in quality than were the sine waves of experiment 1, but it also meant that formant bandwidths were broader and stimuli consisted entirely of noise.

A. Method

1. Subjects

Twenty adults between the ages of 18 and 40 years, 20 7-year-olds, and 20 5-year-olds participated. The mean age of 7-year-olds was 7;3, with a range from 7;1 to 7;5. The mean age of 5-year-olds was 5;2, with a range from 4;11 to 5;5. All listeners met the same criteria as in experiment 1. 3-year-olds were not tested in this second experiment both because they had performed similarly to 5-year-olds in the VOICED condition of experiment 1 and because they did so poorly with the stimuli that were not VOICED in that experiment. It seemed neither necessary nor advisable to include them.

2. Equipment and stimuli

The same equipment and VOICED stimuli were used as in experiment 1.

The WHISPERED stimuli used the same synthetic fricative noises as the VOICED and SINE-WAVE stimuli. In these stimuli, however, natural whispered vocalic portions were combined with those synthetic noises. These portions were taken from a male speaker whispering /su/, /sɑ/, /ʃu/, and /ʃɑ/. The recordings were sampled at 22.05 kHz and low-

TABLE VII. Means across the three tokens for the phonetically relevant acoustic properties of the vocalic portions of the WHISPERED stimuli used in experiment 2.

Portion	Dur (ms)	F1-onset (Hz)	F2-onset (Hz)	F3-onset (Hz)
(s)ɑ	375	718	1617	2770
(ʃ)ɑ	405	574	1680	2656
(s)u	387	517	1769	2727
(ʃ)u	341	445	2010	2627

pass filtered below 11.025 kHz. Three tokens of each vocalic portion were used. Table VII presents mean values for duration and formant-frequency onsets for each whispered vocalic portion. For /(ʃ)u/ and /(s)u/, F1 remained stable throughout the vocalic portion, F2 fell through the entire portion to a mean ending frequency of 1572 Hz, and F3 fell over roughly the first 100 ms to a mean steady-state frequency of 2469 Hz. For /(ʃ)ɑ/ and /(s)ɑ/, F1 rose over the first 60–90 ms to a steady-state frequency of 969 Hz, and F2 fell to a mean steady-state frequency of 1314 Hz. For /(s)ɑ/, F3 remained fairly stable through the entire vocalic portion; for /ʃɑ/, F3 rose over the first 100 ms to a mean steady-state frequency of 2756 Hz. Compared to the voiced vocalic portions, formant frequencies were generally higher in these whispered portions, which matches the findings of (Kallail and Emmanuel 1984a, 1984b). A WHISPERED /(ʃ)ɑ/ stimulus with a 3.0 kHz fricative noise is shown on the right-hand side of Fig. 3. It is apparent from this spectrogram that the formants were much less prominent in the whispered stimuli, compared to the sine wave and natural stimuli. Furthermore, in this token, where the frequency of F1 is high, F1 does not appear attenuated. However, in /u/ tokens, where F1 frequency is lower, F1 is greatly attenuated. This is shown in Fig. 5, where a whispered /(ʃ)u/ is displayed next to a voiced /(ʃ)u/. This figure also more clearly shows that a subglottal

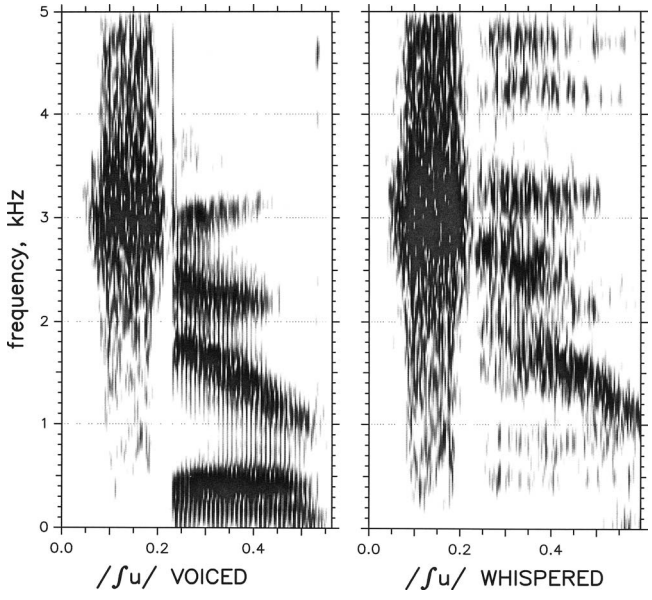


FIG. 5. Spectrograms displaying representative tokens of voiced and whispered /ʃu/.

TABLE VIII. Mean weighting coefficients in the VOICED condition of experiment 2. Standard deviations are in parentheses.

	Adults	7-year-olds	5-year-olds
Transition	0.37 (0.19)	0.37 (0.17)	0.51 (0.12)
Noise	0.78 (0.12)	0.78 (0.14)	0.65 (0.14)

resonance was present at roughly 1500 Hz in the whispered tokens, making the measurement of F2 frequency tricky.

One issue that arose in creating the WHISPERED stimuli concerned the amplitude ratio of the fricative noise and vocalic syllable portions. In natural speech, whispered vocalic portions are much lower in amplitude than the fricative noises. When we tried to combine fricative and whispered portions with equal amplitudes, as we had done for the other stimuli, the two syllable portions did not cohere perceptually. Instead, it was necessary to combine them with the whispered vocalic portion 6 dB lower than the fricative-noise portion.

3. Procedures

The same procedures were used as in experiment 1.

B. Results

1. VOICED condition

Table VIII shows mean weighting coefficients for the VOICED condition. As in experiment 1, there appear to be differences in cue weightings for 5-year-olds compared to listeners in the two older groups: It appears that 5-year-olds weighted formant transitions more and fricative noises less than adults or 7-year-olds did. To examine these apparent findings, one-way ANOVAs, with age as the main effect, were done on weighting coefficients for transitions and fricative noises, separately, for results from the VOICED condition in experiment 2. Results are shown in Table IX. There was a significant main effect of age for both formant transitions and fricative noises. *Post hoc t* tests confirmed that

TABLE IX. Results of ANOVAs and *post hoc t*-tests done on weighting coefficients in the VOICED condition of experiment 2. Bonferroni significance levels indicate that the contrast is significant at the level shown when corrections are made for multiple tests.

	df	*F* or *t*	*p*	Bonferroni significance
		Transition		
Age effect	2,57	4.85	0.011	
Adults vs 7-year-olds	57	0.02	ns	
Adults vs 5-year-olds	57	2.71	0.009	0.05
7- vs 5-year-olds	57	2.69	0.009	0.05
		Noise		
Age effect	2,57	6.23	0.004	
Adults vs 7-year-olds	57	−0.01	ns	
Adults vs 5-year-olds	57	−3.06	0.003	0.01
7- vs 5-year-olds	57	−3.05	0.004	0.05

TABLE X. Mean weighting coefficients in the WHISPERED condition of experiment 2. Standard deviations are in parentheses.

	Adults	7-year-olds	5-year-olds
Transition	0.35 (0.26)	0.21 (0.18)	0.35 (0.22)
Noise	0.75 (0.15)	0.81 (0.12)	0.69 (0.18)

5-year-olds weighted formant transitions more and fricative noises less than did 7-year-olds or adults.

Next we examined if there were any differences between the VOICED conditions in experiments 1 and 2. To accomplish that goal two-group *t*-tests were performed for each age group separately. There was no significant difference in weighting of formant transitions or of fricative noises across experiments for any age group. Consequently we may conclude that the listeners in this experiment were similar to those in experiment 1 in terms of their perceptual weighting strategies.

2. WHISPERED condition

75% of 7-year-olds and 50% of 5-year-olds were able to complete the WHISPERED condition with 80% accurate responses on best exemplars during testing. These percentages were lower than the 80% of 7-year-olds and 75% of 5-year-olds who were able to complete the SINE-WAVE condition in experiment 1. Apparently whispered speech was more difficult than sine-wave speech for children to recognize.

Table X presents mean weighting coefficients for the WHISPERED condition. When these coefficients are compared to those in Table VIII, for the VOICED condition, it is apparent that adults performed similarly across the two conditions. Both 7- and 5-year-olds, however, showed greatly diminished weighting of the formant transitions; the weights they assigned to fricative noises appear similar across conditions.

ANOVAs were done on weighting coefficients for transitions and fricative noises separately, and results are shown in Table XI. The main effect of age for transition was significant. *Post hoc t*-tests revealed that for the first time 7-year-olds performed differently from adults, having greatly decreased the weights they assigned to formant transitions in these WHISPERED stimuli. Now 7-year-olds weighted formant transitions less than adults. 5-year-olds similarly decreased the weights they assigned to formant transitions, but in this case that change brought them inline with the weights used by adults.

Although no great changes in weighting coefficients for fricative noises were observed, the small shifts across groups led to changes from previous conditions in statistical outcomes. In particular, adults' slight decrease and 5-year-olds' slight increase in weighting coefficients meant that there was no statistically significant difference between these groups. As before, there was a significant difference between 7- and 5-year-olds.

TABLE XI. Results of ANOVAs and *post hoc t*-tests done on weighting coefficients in the WHISPERED condition of experiment 2. Bonferroni significance levels indicate that the contrast is significant at the level shown when corrections are made for multiple tests.

	df	*F* or *t*	*p*	Bonferroni significance
		Transition		
Age effect	2,57	3.29	0.044	
Adults vs 7 year-olds	57	−2.21	0.032	0.10
Adults vs 5-year-olds	57	0.03	ns	
7- vs 5-year-olds	57	2.24	0.029	0.10
		Noise		
Age effect	2,57	4.21	0.02	
Adults vs 7-year-olds	57	1.50	ns	
Adults vs 5-year-olds	57	−1.40	ns	
7- vs 5-year-olds	57	−2.90	0.005	0.05

3. Across-conditions comparison

As in experiment 1, the most relevant analysis for the main hypothesis being tested was the one for each group separately, across stimulus conditions. Matched *t*-tests were done comparing performance of each group across the VOICED and WHISPERED conditions in this second experiment, and the results are shown in Table XII. Here we see a significant effect of stimulus type for the weights assigned to formant transitions for both groups of children, but not for adults. In fact, the magnitude of change in those weights across conditions is remarkably similar for 7- and 5-year-olds. For example, if we compute effect sizes for the changes in weights using Cohen's *d* (Cohen, 1988), we find $d=0.91$ for 7-year-olds and $d=0.90$ for 5-year-olds. Clearly these young listeners had difficulty recovering and/or using these noisy formant transitions in their phonetic decisions.

The slight changes in weights assigned to fricative noises across conditions were not statistically significant for any group.

C. Discussion

The purpose of this experiment was to test the hypothesis that listeners would change their perceptual weighting strategies when whispered vocalic portions replaced natural vocalic portions, consequently eliminating the harmonicity of the stimuli. In particular, we wanted to examine whether

TABLE XII. Results of matched *t*-tests done on within-group weighting coefficients across the VOICED and WHISPERED conditions for experiment 2.

	df	*t*	*p*
	Transition		
Adults	19	0.36	ns
7-year-olds	19	3.62	0.002
5-year-olds	19	4.31	<0.001
	Noise		
Adults	19	0.93	ns
7-year-olds	19	−0.97	ns
5-year-olds	19	−1.21	ns

children would show diminished weighting of formant transitions when these whispered vocalic portions were used.

Adults performed similarly in the VOICED and WHISPERED conditions, while children performed differently in the two conditions. In stark contrast to what was found in experiment 1, where harmonicity was disrupted by using sine-wave speech, children were found to greatly diminish the weights they assigned to formant transitions in this experiment with whispered speech. Of course, a lack of harmonicity could not readily explain this observed shift in children's weighting strategies: Stimuli in experiment 1 similarly lacked harmonicity, and children's weighting strategies were not affected. That being the case, the question arose as to what could explain these observed changes in perceptual weighting.

This second experiment showed that children greatly reduced the weights they assigned to formant transitions when whispered vocalic portions were used in stimulus construction. Either children were unable to recover the patterns of global spectral change in these noisy signals or they chose not to tune their perceptual attention to these signal components. Whatever the reason might be, whispered speech is very natural, so the difficulty could not have been due to a lack of naturalness. We suspect that instead the difficulty was due to the broadened formant structure of whispered speech. As Figs. 3 and 5 illustrate, whispering results in formants that are less prominent than in voiced speech; energy is spread more evenly across the spectrum. This lack of prominence for formants could make it difficult for children to recover the dynamic spectral patterns associated with changing vocal-tract cavity shapes and sizes. In some sense, listening to whispered speech is like listening to speech through the widened auditory filters of hearing loss. Turner and Van Tassell (1984) reported that adult listeners did not encounter difficulty listening to speech under such conditions. Our results for adults with normal hearing listening to whispered speech match those results for adults with hearing loss: They did not modify their weighting strategies for the whispered stimuli. In fact, it is interesting that adults in this current study actually decreased their weighting of formant transitions when sine-wave replicas of the vocalic portions were used in stimulus construction. This finding suggests that adults might be hindered slightly by processing algorithms that emphasize the spectral prominences of formants. A study by Hillenbrand and Houde (2002) in which adult listeners showed slightly poorer consonant and vowel recognition for spectrally enhanced speech signals supports that suggestion, although a different study by Summerfield *et al.* (1981) did not find that adults were hindered in their speech recognition when formants were made more spectrally prominent.

But regardless of whether adults are hindered or not by algorithms that enhance the spectral prominences (i.e., formants) of speech, it is clear that they are not hurt by broadened bandwidths. Our results for children, on the other hand, suggest quite a different situation: Children did encounter difficulty when the formants were not spectrally prominent, as occurred with the whispered stimuli. This finding matches other results coming from this laboratory. In particular, Nittrouer *et al.* (2009) found that 7-year-old children were poorer at recognizing noise vocoded sentences than adults, but were similarly skilled at recognizing the same sentences when they were presented as sine waves. Even more so than whispered stimuli, noise vocoded speech broadens the bandwidths of the spectral prominences. If children with normal hearing and typical speech development demonstrate these constraints in their processing of such signals, it is reasonable to expect that children with hearing loss would also have difficulty processing signals with reduced spectral prominences. Perhaps children with hearing loss would benefit from algorithms that track formants in the signal and shape the spectral output to make those formants more prominent.

Of course, signals in this second experiment, as well as those in Nittrouer *et al.* (2009), shared another attribute. In both cases they were constructed entirely of noise. That means that possibly the shift away from weighting formant transitions strongly demonstrated by this experiment and by the poor sentence recognition found in Nittrouer *et al.* (2009) for noisy stimuli might be attributable to children failing to attend to noise components in speech signals. The current study cannot really separate the contributions made to children's results of broadened bandwidths and noisy signals. Future investigations will need to examine that distinction.

IV. GENERAL DISCUSSION

Previous research has revealed that adults and children assign different weights to the various components of the acoustic speech signal. Adults, it seems, are quite attentive to the spectrotemporal details of the signal. These are properties that arise from precise timing relations among articulatory actions, or exact constriction shapes. In the production of a fricative, for example, many variables come into play: the place of the constriction in the vocal tract must be accurate; the degree of constriction between articulatory structures must be tight enough that turbulence noise is generated, but not so tight that stopping is introduced; the length and shape of the constriction must be right; and any secondary cavities that help mold the acoustic output, such as a sublingual cavity in the case of English /ʃ/, must be created (Perkell *et al.*, 1979). Learning to produce speech with all these parameters properly implemented takes time, as apparently does learning to pay attention to their acoustic consequences so that phonetic structure may be accurately and efficiently recovered.

Unlike adults, research into children's speech perception and word recognition has uncovered consistent evidence that children pay particular attention to global spectral patterns within the acoustic signal (i.e., ongoing change in formant frequencies), rather than to the details that capture adults' attention. Previously it was hypothesized that this perceptual strategy is preferred early in life because these global, recurrent patterns of structure help children isolate linguistically significant stretches, such as words (e.g., Jusczyk, 1993; Nittrouer *et al.*, 2009; Studdert-Kennedy, 1987). However, careful scientific inquiry dictates that all alternative explanations be explored before accepting just one. It may be that the enhanced weighting observed in children's speech perception

for global changes in formant frequencies have to do with nonlinguistic considerations. In the study reported here we explored the possibility that children's preferential weighting of formant transitions in decisions of syllable-initial fricatives may actually be explained by the fact that the portion of formant transitions involved in stimulus design in early experiments was always voiced, while the fricative noises were always just that—noise. Perhaps children attend to signal portions that adhere to principles of ASA, especially the harmonicity principle. At least that was the hypothesis tested.

Results of these experiments led us to reject that hypothesis: When the principle of harmonicity was disrupted in the stimuli by using sine-wave replicas of the vocalic portions, children assigned the same weights to formant transitions as they did for stimuli with natural vocalic portions. Consequently, the assertion that children rely on the slowly modulating formants in the speech signal, which arise from fluid changes in vocal-tract cavity shapes and sizes, remains supportable. But that was not the end of this story. It was also found that children experience difficulty attending to formant transitions when those transitions are not spectrally prominent and consist only of noise. It was suggested that if the fact that the whispered stimuli lacked spectrally prominent formants explained the lion's share of difficulty encountered by children it would mean that children have more difficulty than adults perceiving speech under the condition of sensorineural hearing loss.

In summary, the two experiments reported here were designed to investigate whether listeners, particularly children, would show decreased weighting of formant transitions when vocalic portions failed to provide common harmonic structure across the formants. Children were not found to decrease their weightings of formant transitions in the first experiment, and the suggestion was made that results of the second experiment may be explained by difficulty on the part of children in recovering formant transitions when formants are not spectrally prominent and/or consist only of noise. Future research will need to test the veracity of that suggestion.

ACKNOWLEDGMENTS

This work was supported by research Grant No. R01 DC-00633 from the National Institute on Deafness and Other Communication Disorders, the National Institutes of Health, to S.N.

Barker, J., and Cooke, M. P. (**1999**). "Is the sine-wave speech cocktail party worth attending?," Speech Commun. **27**, 159–174.

Beddor, P. S., and Strange, W. (**1982**). "Cross-language study of perception of the oral-nasal distinction," J. Acoust. Soc. Am. **71**, 1551–1561.

Bregman, A. S. (**1990**). *Auditory Scene Analysis* (MIT, Cambridge, MA).

Charles-Luce, J., and Luce, P. A. (**1990**). "Similarity neighbourhoods of words in young children's lexicons," J. Child Lang **17**, 205–215.

Charles-Luce, J., and Luce, P. A. (**1995**). "An examination of similarity neighbourhoods in young children's receptive vocabularies," J. Child Lang **22**, 727–735.

Cohen, J. (**1988**). *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. (Erlbaum, Hillsdale, NJ).

Cole, R. A., and Perfetti, C. A. (**1980**). "Listening for mispronunciations in a children's story: The use of context by children and adults," J. Verbal Learn. Verbal Behav. **19**, 297–315.

Crowther, C. S., and Mann, V. (**1992**). "Native language factors affecting use of vocalic cues to final consonant voicing in English," J. Acoust. Soc. Am. **92**, 711–722.

Crowther, C. S., and Mann, V. (**1994**). "Use of vocalic cues to consonant voicing and native language background: The influence of experimental design," Percept. Psychophys. **55**, 513–525.

Flege, J. E., and Port, R. (**1981**). "Cross-language phonetic interference: Arabic to English," Lang Speech **24**, 125–146.

Goldman, R., and Fristoe, M. (**2000**). *Goldman Fristoe 2: Test of Articulation* (American Guidance Service, Inc., Circle Pines, MN).

Goodell, E. W., and Studdert-Kennedy, M. (**1993**). "Acoustic evidence for the development of gestural coordination in the speech of 2-year-olds: A longitudinal study," J. Speech Hear. Res. **36**, 707–727.

Harris, K. S. (**1958**). "Cues for the discrimination of American English fricatives in spoken syllables," Lang Speech **1**, 1–7.

Heinz, J. M., and Stevens, K. N. (**1961**). "On the properties of voiceless fricative consonants," J. Acoust. Soc. Am. **33**, 589–593.

Hillenbrand, J. M., and Houde, R. A. (**2002**). "Speech synthesis using damped sinusoids," J. Speech Lang. Hear. Res. **45**, 639–650.

Jastak, S., and Wilkinson, G. S. (**1984**). *The Wide Range Achievement Test-Revised* (Jastak Associates, Wilmington, DE).

Jusczyk, P. W. (**1982**). "Auditory versus phonetic coding of speech signals during infancy," in *Perspectives on Mental Representation: Experimental and Theoretical Studies of Cognitive Processes and Capacities*, edited by J. Mehler, M. F. Garrett, and E. C. T. Walker (Erlbaum, Hillsdale, NJ), pp. 361–387.

Jusczyk, P. W. (**1993**). "From general to language-specific capacities: The WRAPSA model of how speech perception develops," J. Phonetics **21**, 3–28.

Kallail, K. J., and Emanuel, F. W. (**1984a**). "An acoustic comparison of isolated whispered and phonated vowel samples produced by adult male subjects," J. Phonetics **12**, 175–186.

Kallail, K. J., and Emanuel, F. W. (**1984b**). "Formant-frequency differences between isolated whispered and phonated vowel samples produced by adult female subjects," J. Speech Hear. Res. **27**, 245–251.

Kunisaki, O., and Fujisaki, H. (**1977**). "On the influence of context upon perception of voiceless fricative consonants," Ann. Bulletin of the RILP **11**, 85–91.

Lisker, L., and Abramson, A. S. (**1964**). "A cross-language study of voicing in initial stops: Acoustical measurements," Word **20**, 384–422.

MacKain, K. S., Best, C. T., and Strange, W. (**1981**). "Categorical perception of English /r/ and /l/ by Japanese bilinguals," Appl. Psycholinguist. **2**, 369–390.

MacNeilage, P. F., and Davis, B. L. (**1991**). "Acquisition of speech production: Frames, then content," in *Attention & Performance XIII*, edited by M. Jeanerod (Erlbaum, New York), pp. 453–476.

Mann, V. A., and Repp, B. H. (**1981**). "Influence of preceding fricative on stop consonant perception," J. Acoust. Soc. Am. **69**, 548–558.

Matsuda, M., and Kasuya, H. (**1999**). "Acoustic Nature of the Whisper," Eurospeech '99, pp. 133–136.

Mayo, C., Scobbie, J. M., Hewlett, N., and Waters, D. (**2003**). "The influence of phonemic awareness development on acoustic cue weighting strategies in children's speech perception," J. Speech Lang. Hear. Res. **46**, 1184–1196.

McGowan, R. S., and Nittrouer, S. (**1988**). "Differences in fricative production between children and adults: Evidence from an acoustic analysis of /sh/ and /s/," J. Acoust. Soc. Am. **83**, 229–236.

Nittrouer, S. (**1992**). "Age-related differences in perceptual effects of formant transitions within syllables and across syllable boundaries," J. Phonetics **20**, 351–382.

Nittrouer, S. (**1996**). "The discriminability and perceptual weighting of some acoustic cues to speech perception by three-year-olds," J. Speech Hear. Res. **39**, 278–297.

Nittrouer, S. (**2002**). "Learning to perceive speech: How fricative perception changes, and how it stays the same," J. Acoust. Soc. Am. **112**, 711–719.

Nittrouer, S., and Miller, M. E. (**1997a**). "Predicting developmental shifts in perceptual weighting schemes," J. Acoust. Soc. Am. **101**, 2253–2266.

Nittrouer, S., and Miller, M. E. (**1997b**). "Developmental weighting shifts for noise components of fricative-vowel syllables," J. Acoust. Soc. Am. **102**, 572–580.

Nittrouer, S., and Studdert-Kennedy, M. (**1987**). "The role of coarticulatory effects in the perception of fricatives by children and adults," J. Speech Hear. Res. **30**, 319–329.

Nittrouer, S., Lowenstein, J. H., and Packer, R. (**2009**). "Children discover

the spectral skeletons in their native language before the amplitude envelopes," J. Exp. Psychol. in press.
Nittrouer, S., Manning, C., and Meyer, G. (**1993**). "The perceptual weighting of acoustic cues changes with linguistic experience," J. Acoust. Soc. Am. **94**, S1865.
Perkell, J. S., Boyce, S. E., and Stevens, K. N. (**1979**). "Articulatory and acoustic correlates of the [sš] distinction," J. Acoust. Soc. Am. **65**, S24.
Pittman, A. L., Stelmachowicz, P. G., Lewis, D. E., and Hoover, B. M. (**2003**). "Spectral characteristics of speech at the ear: Implications for amplification in children," J. Speech Lang. Hear. Res. **46**, 649–657.
Remez, R. E., Rubin, P. E., Berns, S. M., Pardo, J. S., and Lang, J. M. (**1994**). "On the perceptual organization of speech," Psychol. Rev. **101**, 129–156.
Remez, R. E., Rubin, P. E., Pisoni, D. B., and Carrell, T. D. (**1981**). "Speech perception without traditional speech cues," Science **212**, 947–949.
Repp, B. H. (**1982**). "Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception," Psychol. Bull. **92**, 81–110.
Ruff, H. A. (**1982**). "Effect of object movement on infants' detection of object structure," Dev. Psychol. **18**, 462–472.
Serniclaes, W., Sprenger-Charolles, L., Carre, R., and Demonet, J. F. (**2001**). "Perceptual discrimination of speech sounds in developmental dyslexia," J. Speech Lang. Hear. Res. **44**, 384–399.
Spelke, E., von Hofsten, C., and Kestenbaum, R. (**1989**). "Object perception in infancy: Interaction of spatial and kinetic information for object boundaries," Dev. Psychol. **25**, 185–196.
Stelmachowicz, P. G., Lewis, D. E., Choi, S., and Hoover, B. (**2007**). "Effect of stimulus bandwidth on auditory skills in normal-hearing and hearing-impaired children," Ear Hear. **28**, 483–494.
Stevens, K. N. (**1975**). "The potential role of property detectors in the perception of consonants," in *Auditory Analysis and Perception of Speech*, edited by G. Fant and M. A. A. Tatham (Academic, New York), pp. 303–330.
Studdert-Kennedy, M. (**1987**). "The phoneme as a perceptuomotor structure," in *Language Perception and Production: Relationships Between Listening, Speaking, Reading, and Writing*, edited by A. Allport, D. G. MacKay, W. Prinz, and E. Scheerer (Academic, Orlando), pp. 67–84.
Summerfield, Q., Tyler, R., Foster, J., Wood, E., and Bailey, P. J. (**1981**). "Failure of formant bandwidth narrowing to improve speech reception in sensorineural impairment," J. Acoust. Soc. Am. **70**, S108–S109.
Tartter, V. C. (**1989**). "What's in a whisper?," J. Acoust. Soc. Am. **86**, 1678–1683.
Tartter, V. C. (**1991**). "Identifiability of vowels and speakers from whispered syllables," Percept. Psychophys. **49**, 365–372.
Tice, B., and Carrell, T. (**1997**) TONE v.1.5b (software).
Tsunoda, K., Ohta, Y., Soda, Y., Niimi, S., and Hirose, H. (**1997**). "Laryngeal adjustment in whispering magnetic resonance imaging study," Ann. Otol. Rhinol. Laryngol. **106**, 41–43.
Turner, C. W., and Holte, L. A. (**1987**). "Discrimination of spectral-peak amplitude by normal and hearing-impaired subjects," J. Acoust. Soc. Am. **81**, 445–451.
Turner, C. W., and Van Tasell, D. J. (**1984**). "Sensorineural hearing loss and the discrimination of vowel-like stimuli," J. Acoust. Soc. Am. **75**, 562–565.
Vouloumanos, A., and Werker, J. F. (**2007**). "Listening to language at birth: Evidence for a bias for speech in neonates," Dev. Sci. **10**, 159–164.
Walley, A. C., Smith, L. B., and Jusczyk, P. W. (**1986**). "The role of phonemes and syllables in the perceived similarity of speech sounds for children," Mem. Cognit. **14**, 220–229.

Vowel devoicing and the perception of spoken Japanese words

Anne Cutler[a)]
Max Planck Institute for Psycholinguistics, PO Box 310, 6500 AH Nijmegen, The Netherlands and MARCS Auditory Laboratories, University of Western Sydney, Penrith South DC NSW 1797, Australia

Takashi Otake[b)]
Max Planck Institute for Psycholinguistics, PO Box 310, 6500 AH Nijmegen, The Netherlands and E-Listening Laboratory, Tokorozawa, 359 0021, Japan

James M. McQueen[c)]
Max Planck Institute for Psycholinguistics, PO Box 310, 6500 AH Nijmegen, The Netherlands and Radboud University Nijmegen, Postbus 9104, 6500 HE Nijmegen, The Netherlands

(Received 6 August 2007; revised 20 November 2008; accepted 14 December 2008)

Three experiments, in which Japanese listeners detected Japanese words embedded in nonsense sequences, examined the perceptual consequences of vowel devoicing in that language. Since vowelless sequences disrupt speech segmentation [Norris *et al.* (1997). Cognit. Psychol. **34**, 191–243], devoicing is potentially problematic for perception. Words in initial position in nonsense sequences were detected more easily when followed by a sequence containing a vowel than by a vowelless segment (with or without further context), and vowelless segments that were potential devoicing environments were no easier than those not allowing devoicing. Thus *asa*, "morning," was easier in *asau* or *asazu* than in all of *asap, asapdo, asaf*, or *asafte*, despite the fact that the /f/ in the latter two is a possible realization of *fu*, with devoiced [u]. Japanese listeners thus do not treat devoicing contexts as if they always contain vowels. Words in final position in nonsense sequences, however, produced a different pattern: here, preceding vowelless contexts allowing devoicing impeded word detection less strongly (so, *sake* was detected less accurately, but not less rapidly, in *nyaksake*—possibly arising from *nyakusake*—than in *nyagusake*). This is consistent with listeners treating consonant sequences as potential realizations of parts of existing lexical candidates wherever possible. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3075556]

PACS number(s): 43.71.Es [PEI] Pages: 1693–1703

I. INTRODUCTION

In spoken Japanese, the high vowels /i/ and /u/ are devoiced when they follow a voiceless consonant and precede either another voiceless consonant or (in the case of fricatives and affricates) a pause, and do not bear accent (Vance, 1987). Thus *sashimi* "sashimi" (raw fish dish) and *sashiki* "cutting" begin with the same two-mora sequence; in writing, it is the same in both words. But in *sashiki* the /i/ of the medial mora, occurring before /k/, is devoiced, even in careful speech. The /i/ in *sashimi*, occurring before /m/, never devoices. Although devoicing is not obligatory, analyses of the Corpus of Spontaneous Japanese (Maekawa, 2003) show that it is highly probable (over 98% in some environments; Kondo, 2005; Maekawa and Kikuchi, 2005).

The effect of this devoicing is the creation of sequences of consonants not separated by the periodic articulation normally associated with vowels. Otherwise, though, Japanese phonology drastically restricts the occurrence of consonant sequences. Japanese has no consonant clusters, and allows only a very restricted range of simple syllable codas: nasals (*Hondo; Unzen*) or geminate consonants (*Hokkaido*; *Sapporo*—these are all place names). Loan-words that contain consonant sequences or un-Japanese codas are adapted by vowel insertion; e.g., *glove* becomes *gurabu* and *express* becomes *ekisupuresu*. This process of vowel insertion is so fundamental to Japanese phonology that in perceptual tasks Japanese listeners respond to nonsense VCCV strings such as *ebzo* as if they were VCVCV *ebuzo* (Dupoux *et al.*, 1999).

Japanese is not unusual in preferring consonants and vowels to alternate; such a preference appears across languages, and has a good perceptual foundation. It has long been known that a following vowel facilitates consonant identification (Liberman *et al.*, 1954; van Son and Pols, 1995). In line with this, the deletion of vowels, even where it occurs regularly in casual speech, makes words harder to recognize: for example, lexical decision responses can be slower for words with deleted vowels (e.g., *s'maine* for *semaine*; Racine and Grosjean, 2000). In contrast, insertion of a vowel into a consonant cluster (e.g., *fillum* for *film*) makes recognition easier, in part because the consonants in the cluster indeed become easier to identify if separated (van Donselaar *et al.*, 1999).

Vowel devoicing in Japanese could thus be perceptually disadvantageous. Further, it could complicate the parsing of continuous speech into its component words. All speech input is potentially consistent with alternative interpretations; *legacy* contains *leg* embedded within it, but *leg* itself con-

[a)]Present address: Max Planck Institute for Psycholinguistics, 6500 AH Nijmegen, The Netherlands. Electronic mail: anne.cutler@mpi.nl
[b)]Electronic mail: otake@e-listeninglab.com
[c)]Electronic mail: james.mcqueen@mpi.nl

tains *egg*, and whenever we hear the longer word spoken, we also receive input consistent with the shorter word or words. There are many powerful and efficient techniques which listeners can apply to reduce interference from unwanted embedded words that are accidentally present in the speech stream (and so might be activated by speech input). In the above example, for instance, interference from *egg* in recognition of *leg* is negligible. The process which achieves this is called the possible word constraint (PWC) (Norris *et al.*, 1997); it exploits the widespread rule that a vowel alone can be a word, but a consonant in general cannot.

The effect of the PWC can be seen in word-spotting (Cutler and Norris, 1988; McQueen, 1996), a psycholinguistic task for investigating segmentation of speech. In word-spotting, listeners detect any real word embedded in spoken nonwords. The task has exposed language-particular segmentation effects such as use of stress information in English (Cutler and Norris, 1988), of syllables in French (Dumay *et al.*, 2002), and of vowel harmony in Finnish (Suomi *et al.*, 1997), it has shown effects of native-language sequence constraints on segmentation in a second language (Weber and Cutler, 2006), it has confirmed that word onsets contribute relatively more in spoken-word recognition than offsets (McQueen, 1998), and it has revealed listeners' sensitivity to the likelihood of a word boundary in a given string of phonemes (McQueen, 1998; van der Lugt, 2001; Warner *et al.*, 2005).

It is very difficult to spot a word if accepting it leaves a vowelless residue of the input. Thus *egg* in *fegg* or *sugar* in *sugarth* are detected less easily than *egg* in *maffegg* or *sugar* in *sugarthig* (Norris *et al.*, 1997). The residues *maff* and *thig* are syllables, so although they are, in fact, not words, they might have been; *f* and *th*, however, as single consonants, are impossible word candidates in English. This finding appears in English (Norris *et al.*, 1997, 2001), in Dutch (McQueen and Cutler, 1998), in Japanese (McQueen *et al.*, 2001), in Sesotho (Cutler *et al.*, 2002a), in French (Spinelli *et al.*, 2003), and in Cantonese (Yip, 2004). These languages vary widely in the surface constraints on what may be a syllable of the language and what may be a stand-alone word, but the PWC difference between possible and impossible residues remains effectively constant. The PWC has thus been held (Norris *et al.*, 2001; Cutler *et al.*, 2002a) to express a universal constraint on syllabic viability: Across languages, vowels alone can be syllables, but consonants cannot. Syllables can be words; thus, because consonants alone cannot be syllables, they also cannot be words, though vowels can be (*eye*, *awe*). Applying the PWC radically reduces the effects of embedding in speech (Cutler *et al.*, 2002b), which makes it potentially very useful in everyday listening.

The Japanese experiment of McQueen *et al.* (2001) allowed a comparison between possible residues (with a vowel) and impossible residues (without a vowel) with the length in number of segments controlled; residues were a single vowel versus a single consonant. This was possible because Japanese allows sequences of vowels. Thus spotting of *uni* "sea urchin" was compared in *puni* versus *iuni*, and of *hiru* "noon" in *hiruk* versus *hirua*. Spotting the word was always hardest when the single-consonant residue remained, just as in the other languages. This shows that the findings from other languages could not have been due to length differences between impossible versus possible residues, and that Japanese listeners, like listeners with other native languages, are sensitive to the different viabilities of vowels versus consonants as residues in speech segmentation. Although vowels can be inserted into non-Japanese input (loanwords or nonwords such as those presented by Dupoux *et al.*, 1999), detection of real Japanese words is seriously hindered when the adjacent context contains no vowel.

The sequences created by the process of devoicing in Japanese thus have the potential to make speech segmentation and word recognition harder. Recall, however, that devoicing generally occurs only for /i/ and /u/ and only in voiceless contexts. It is possible that a single voiceless consonant that is heard often in devoicing environments will not hinder segmentation (and disrupt embedded word detection in consequence), because it is interpreted not as a vowelless consonant but as equivalent to a (devoiced) syllable. The test of the PWC by McQueen *et al.* (2001) in Japanese deliberately excluded potential devoicing contexts. In the present study, we focus specifically on these cases.

It is important to test both preceding contexts (which disrupt word onsets) and following contexts (which disrupt word offsets). In the PWC studies (Norris *et al.*, 1997; McQueen *et al.*, 2001), as in other studies (e.g., McQueen, 1998), preceding contexts exercised the strongest effects. Thus we here test both context positions. We begin with following contexts: in Experiment 1 we compare impossible single-consonant codas, which in the study of McQueen *et al.* (2001) made word-spotting difficult, with single-vowel contexts, which made word-spotting easy, and both of these with single-consonant devoicing environments. For instance, we compare how difficult it is to spot *asa* "morning" in *asap*, *asau*, or *asaf*; the stop /p/ could not occur at the end of a Japanese word, but /f/ is a voiceless fricative and so could occur followed by an underlying devoiced vowel. This comparison allows us to determine whether the devoiced case will pattern more like an impossible consonant or more like a possible vowel context.

Note that in the impossible coda case, what is being presented to listeners is an illegal sequence of the language. Phonotactic legality may affect both how sequences are perceived by listeners and how they are uttered by speakers. It was important, therefore, to rule out such speaker-related effects in our listening experiment. We did this by conducting two versions of the experiment, in one of which (Experiment 1B) the final consonants were produced as (illegal) codas, while in the other (Experiment 1C) they were produced as (legal) onsets of CV syllables, from which the final vowel was then digitally removed prior to the experiment. If speaker-related legality affects listener responses to the final consonants, Experiments 1B and 1C will return different response patterns. Since this comparison meant that the tokens of the embedded target words differed across contexts, a control lexical decision experiment first tested whether the versions differed in how recognizable they were (Experiment 1A).

TABLE I. Materials for Experiments 1 and 2.

		Experiment 1			Experiment 2		
Target noun	English gloss	V context	Possible devoicing C context	Impossible devoicing C context	CV context	Possible devoicing CCV context	Impossible devoicing CCV context
ase	sweat	asea	ases(u)	asep(u)	aseka	aseska	asepge
ani	brother	aniu	anits(u)	anit(u)	anigu	anitsse	anitmo
uso	lie	usoa	usoch(i)	usop(u)	usota	usochsi	usopzu
chizu	map	chizua	chizuf(u)	chizut(u)	chizuya	chizufte	chizutba
haru	spring	harua	haruts(u)	harup(u)	harupa	harutspu	harupbu
yuzu	citron	yuzua	yuzuch(i)	yuzut(u)	yuzupa	yuzushpi	yuzutge
saru	monkey	sarua	saruts(u)	sarup(u)	saruza	sarutsse	sarupze
motsu	giblets	motsua	motsuch(i)	motsug(u)	motsugu	motsuchka	motsugda
nasu	eggplant	nasua	nasuch(i)	nasub(u)	nasuza	nasushte	nasubda
hiru	noon	hirua	hiruch(i)	hiruk(u)	hiruha	hiruchhe	hirukbe
aki	autumn	akia	akich(i)	akib(u)	akiha	akichka	akibzo
ibo	wart	iboi	iboch(i)	ibop(u)	ibogi	ibochta	ibopzo
uzu	whirlpool	uzua	uzus(u)	uzut(u)	uzupa	uzuspe	uzutme
kinu	silk	kinua	kinuf(u)	kinup(u)	kinuza	kinufko	kinupgo
kuzu	trash	kuzua	kuzus(u)	kuzut(u)	kuzuga	kuzuska	kuzutge
naya	shed	nayau	nayaf(u)	nayat(u)	nayapu	nayafpe	nayatma
mitsu	honey	mitsua	mitsuch(i)	mitsud(u)	mitsupa	mitsufpo	mitsudba
kazu	number	kazua	kazus(u)	kazup(u)	kazuha	kazusko	kazupgo
rusu	absence	rusua	rusus(u)	rusug(u)	rusuha	rususke	rusugde
yoru	night	yorua	yoruts(u)	yorup(u)	yoruza	yorucha	yorupba
asa	morning	asau	asaf(u)	asap(u)	asazu	asafte	asapdo
usu	mortar	usua	usuf(u)	usud(u)	usuza	usufsha	usudba
uni	sea urchin	unia	units(u)	unip(u)	unika	unitska	unipge
eki	station	ekia	ekich(i)	ekid(u)	ekipa	ekichpi	ekidbi
matsu	pine tree	matsua	matsush(i)	matsud(u)	matsuha	matsuchta	matsudbi
moya	mist	moyau	moyach(i)	moyap(u)	moyazu	moyachse	moyapzu
gasu	gas	gasua	gasuch(i)	gasub(u)	gasupa	gasushpu	gasubgo
mesu	surgical knife	mesua	mesuch(i)	mesud(u)	mesuza	mesutsso	mesudzo
risu	squirrel	risua	risuf(u)	risud(u)	risupa	risufte	risudge
tsuru	crane	tsurua	tsuruch(i)	tsurut(u)	tsuruna	tsurushta	tsurutbe

II. EXPERIMENT 1

A. Method

1. Materials and design

Thirty vowel-initial two-mora target words (e.g., *asa*) were selected. Most had been used in the study of McQueen *et al.* (2001), where they had been chosen to contain as few embedded words as possible. These words were placed in five following contexts: (i) a single vowel (e.g., *asau*), (ii) a possible-devoicing consonant (e.g., *asaf*), (iii) the same consonant followed by a high vowel ([u] or [i], e.g., *asafu*), (iv) an impossible-devoicing consonant (e.g., *asap*), and (v) the same consonant followed by [u] (e.g., *asapu*). The devoicing contexts were fricatives or affricates; the impossible contexts were voiced stops, /t/ or /p/ (in native Japanese words, /t/ cannot precede /i/ or /u/ and a single /p/ cannot occur intervocalically). Appendix A lists the phonemes of Japanese, and Table I lists all target-bearing items. We further constructed 84 fillers. Four were words with following CV contexts which should be easy to spot (e.g., *biruta* with *biru* "building"). Eighty were not words and contained no embedded bimoraic words; 30 of these were matched to target-bearing items, with 10 each preceding a vowel (e.g., *dozao*), a possible-devoicing consonant (e.g., *zanuf*) or an impossible consonant (e.g., *bugep*). The remaining 50 fillers were all bimoraic (C)VCV nonwords. There were also 12 practice items (four with targets), modeled on the experimental items.

All materials were recorded by a phonetically trained native speaker of Tokyo dialect (the second author) in a sound-damped booth to digital audio tape, sampling at 48 kHz. Stops produced in final position were released and the targets' default accent patterns were preserved within the recorded items. The materials were transferred to computer (down-sampled to 16 kHz, 16 bits), examined, and labeled and the duration of each target word was measured using the XWAVES speech editor. Target-bearing items in contexts (i), (ii), and (iv) were used in Experiment 1B. The final vowels from contexts (iii) and (v) were digitally removed, cutting at a zero-crossing at the point at which no auditory trace of the final vowel remained. The resulting consonant-final items, plus those with vowel contexts (i), were used in Experiment 1C. A timing pulse was aligned with the onset of each target-bearing item. For the control Experiment 1A, the entire context was removed from each version of each target word, and from each of the 30 fillers matched to target-bearing items, to

give nonwords such as *doza*, *zanu*, and *buge*.

For Experiment 1B three counter-balanced lists were made, with ten targets in each of the three context conditions per list (vowel, possible-devoicing consonant, impossible consonant); all targets appeared once on each list. These target-bearing items were mixed in pseudo-random order with all 84 fillers; there was always at least one filler between any pair of target-bearing items. The same lists were used in Experiment 1C (but with the different versions of the consonant-final target-bearing items). The design of Experiment 1A was similar, except that the 30 target words were rotated over five lists, with six words from each of the original five recorded contexts per list.

2. Participants

One hundred and thirty undergraduate members of Dokkyo University received course credits for participating in the study: 40 in Experiment 1A and 45 each in Experiments 1B and 1C. All were native Japanese speakers from the Tokyo area.

3. Procedure

In Experiment 1A, listeners were told that they would hear a list of words and fillers, and were asked to respond by pressing a button as rapidly as possible whenever they heard a real word, and then to say in a low voice what that word was into a microphone. In Experiments 1B and 1C, listeners were told that they would hear a list of nonsense words, some of which would contain real words embedded at their onset. Examples of bimoraic words in the three context conditions were provided. Listeners were asked to press a button as fast as possible if they spotted any real word and then to say what that word was. No prior information was provided as to the identity of the target words. In all experiments participants were tested in separate sound-attenuating carrels in a quiet room, either individually or in pairs. Prior testing ensured that listeners tested in pairs could not hear each other's spoken responses. The spoken responses were recorded. Participants were asked to press the button with their preferred hand. Each listener heard a practice list, and then one of the experimental lists (eight participants per list in Experiment 1A; 15 per list in Experiments 1B and 1C).

The experiments were run from a Sony TCD D10 DAT player and a computer running NESU experiment control software. The computer clock was started by each timing pulse and stopped by each button-press; responses were logged on the computer.

B. Results

In all experiments analyses of variance (ANOVAs) with participants ($F1$) and items ($F2$) as repeated measure were conducted on both reaction time (RT) and error data. Target durations were subtracted from the raw RTs prior to analysis, to obtain RTs from target offset.

Two control participants (Experiment 1A) who detected no words in one condition of the experiment were excluded from the analyses. No word-spotting participants (Experiments 1B and 1C) had to be excluded for this reason. Listeners' spoken responses were analyzed first. On a few trials (2.4% in Experiment 1A, 0.4% in Experiment 1B, and 0.2% in Experiment 1C), listeners misidentified target words. The button-press responses on these trials were treated as errors. Seven items (*haru*, *motsu*, *hiru*, *ibo*, *kuzu*, *naya*, *mesu*) were excluded from the final analyses because those items were missed by all participants in at least one condition in at least one of the three subexperiments. Mean RTs and error rates for each condition in Experiments 1B and 1C are shown in Tables II and III, respectively, along with the relevant control data from Experiment 1A (*N.B.* the control results for the vowel-context condition are therefore the same across tables).

TABLE II. Experiments 1A (control lexical decision) and 1B (word-spotting): Mean correct RTs (in milliseconds, from target word offset) and mean error rates (in percent) by context condition (*asa*=morning; items in the consonant-context conditions were recorded with no following context).

	Context		
	Vowel	Possible devoicing consonant	Impossible devoicing consonant
	Lexical decision (Experiment 1A)		
Mean RT	565	677	655
Mean error	22%	25%	26%
Example	asa[u]	asa[f]	asa[p]
	Word-spotting (Experiment 1B)		
Mean RT	734	800	756
Mean error	21%	26%	36%
Example	asau	asaf	asap

1. Lexical decision control (Experiment 1A)

The RT ANOVA revealed a main effect of context (i.e., the five different contexts from which the target words had been excised): $F1(4,132)=6.41$, $p<0.005$; $F2(4,88)=3.42$, $p<0.05$. The sets of three conditions corresponding to Ex-

TABLE III. Experiments 1A (control lexical decision) and 1C (word-spotting): Mean correct RTs (in milliseconds, from target word offset) and mean error rates (in percent) by context condition (*asa*=morning; items in the consonant-context conditions were recorded with following vowels, but those vowels were removed prior to the experiment).

	Context		
	Vowel	Possible devoicing consonant	Impossible devoicing consonant
	Lexical decision (Experiment 1A)		
Mean RT	565	618	606
Mean error	22%	28%	27%
Example	asa[u]	asa[f(u)]	asa[p(u)]
	Word-spotting (Experiment 1C)		
Mean RT	712	825	762
Mean error	15%	23%	31%
Example	asau	asaf(u)	asap(u)

periments 1B and 1C, respectively, were compared in pairwise t-tests. These showed the main effect of context to be due to responses in the vowel-context condition (e.g., to *asa* excised from *asau*) being faster than in the other four conditions. For the contexts tested in Experiment 1B and 1C, RTs were shorter to words from vowel contexts than to words from both possible-devoicing contexts [1B: $t1(37)=3.49$, $p<0.005$, $t2(22)=3.37$, $p<0.005$; 1C: $t1(37)=2.21$, $p<0.05$, $t2(22)=1.31$, $p>0.2$] and impossible contexts [1B: $t1(37)=4.24$, $p<0.001$, $t2(22)=2.22$, $p<0.05$; 1C: $t1(37)=1.65$, $p=0.11$, $t2(22)=2.09$, $p<0.05$], though note that the Experiment 1C results were not statistically significant across both participants and items. There were no significant differences between any consonant-context conditions. The overall mean error rate (25%) was relatively low given that these items had all been excised from context; the error ANOVAs showed no effects of context ($F1$ and $F2<1$). The RT effects, however, indicate that the target words for the word-spotting experiments were, irrespective of context, not equally easy to recognize. In all by-item ($F2$) analyses of the word-spotting data, therefore, the lexical decision data were entered in analyses of covariance (ANCOVAs) as covariates (control RTs in the RT analyses and control error rates in the error analyses). The by-participant analyses were standard ANOVAs.

2. Word spotting: Consonants recorded without following context (Experiment 1B)

In the overall analyses, there were effects of context in RTs [$F1(2,84)=3.46$, $p<0.05$; $F2(2,43)=2.18$, $p=0.125$] and, more strongly, in errors [$F1(2,84)=15.53$, $p<0.001$; $F2(2,43)=3.93$, $p<0.05$]. Pairwise comparisons of the context conditions showed that participants spotted words more rapidly in vowel contexts (e.g., *asa* in *asau*) than in possible-devoicing contexts (*asa* in *asaf*): $F1(1,44)=6.33$, $p<0.05$; $F2(1,21)=6.58$, $p<0.05$, and spotted words more accurately in vowel contexts than in impossible-devoicing consonantal contexts (*asa* in *asap*): $F1(1,44)=21.22$, $p<0.001$; $F2(1,21)=7.70$, $p<0.05$. No other pairwise comparisons of either RTs or errors were significant by both $F1$ and $F2$. There were thus no statistically reliable differences between the two-consonant-context conditions, and word-spotting in both of these contexts was more difficult than in vowel contexts (observed for the possible-devoicing contexts primarily in RTs, for the impossible-devoicing contexts primarily in errors).

3. Word spotting: Consonants recorded with following context (Experiment 1C)

The overall by-participant ANOVAs and by-item ANCOVAs revealed significant effects of context in RTs [$F1(2,84)=10.30$, $p<0.001$; $F2(2,43)=3.97$, $p<0.05$] and errors [$F1(2,84)=21.30$, $p<0.001$; $F2(2,43)=4.66$, $p<0.05$]. Word-spotting performance in the vowel contexts was faster [$F1(1,44)=20.66$, $p<0.001$; $F2(1,21)=7.59$, $p<0.05$] and more accurate [$F1(1,44)=9.40$, $p<0.005$; $F2(1,21)=5.82$, $p<0.05$] than in the possible-devoicing consonantal contexts. Word-spotting performance in the vowel contexts was also more accurate [$F1(1,44)=27.15$, $p<0.001$; $F2(1,21)=11.19$, $p<0.005$] than in the impossible-devoicing consonantal contexts. No other pairwise comparison was significant by both participants and items. Although there were thus small differences across Experiments 1B and 1C, the major pattern in the data was the same in both.[1] Finally, 2×2 ANOVAs combining the consonant-context data of Experiments 1B and 1C tested this directly; for both RTs and errors, the experiment by context interaction was insignificant ($F1$ and $F2<1$).

C. Discussion

The answer to the question posed in Sec. I is thus very clear: the results for the devoicing case are like those for single-consonant contexts, and not like those for vowel contexts. As in McQueen *et al.* (2001), vowel contexts made detection of embedded words easy, and all single-consonant contexts made detection hard. Differences of acoustic goodness could not underlie the results, since such differences were factored out by covarying the control lexical decision data. The difficulty of single-consonant contexts was constant even though some of the consonants, namely, those in the possible-devoicing condition, can effectively occur in final position when the vowel they precede is devoiced; word-spotting in this condition was as hard as in the condition where the consonants were impossible codas. The way in which the consonant contexts were produced (intentionally, or by excision of a vowel) did not influence performance.

Nonetheless, since Japanese allows no obstruent codas, all the consonant-final items in Experiment 1 were, as we have noted, illegal. In Experiment 2, we provided the devoicing cases with a potentially legal environment, thus providing a clearer and much more ecologically valid test of our hypothesis. A consonant sequence consisting of voiceless consonants (such as [ʃk] in a natural utterance of *sashiki*) is the canonical environment for devoicing in natural speech. We provided such environments by extending the contexts appended to the target words: each vowel context became CV, and each consonant context became CCV. For example, *asa* now occurred in *asazu* (which we predict to be very easy), *asapdo* (which we predict to be very hard), and *asafte*. The voiceless [f] and [t] in sequence could arise from devoicing in natural speech; in Experiment 2 we can assess whether this two-consonant sequence makes word-spotting hard too, or whether the potential for devoicing renders it easy.

III. EXPERIMENT 2

A. Method

1. Materials and procedure

For the 30 two-mora target words of Experiment 1, three new following contexts were constructed: a CV mora (e.g., *asazu*), a CCV in which devoicing was possible (fricative or affricate plus voiceless C plus V, e.g., *asafte*), and a CCV in which devoicing was impossible (the second C was voiced, e.g., *asapdo*). Fillers were also constructed as for Experiment 1, with again 30 matched to target-bearing items (e.g., *dozago*, *zanufte*, *bugepga*); 50 fillers were trimoraic

TABLE IV. Experiments 2A (control lexical decision) and 2B (word-spotting): Mean correct RTs (in milliseconds, from target word offset) and mean error rates (in percent) by context condition (*asa*=morning).

	Context		
	Consonant+vowel	Possible devoicing consonant cluster+vowel	Impossible devoicing consonant cluster+vowel
	Lexical decision (Experiment 2A)		
Mean RT	606	653	653
Mean error	18%	36%	29%
Example	asa[zu]	asa[fte]	asa[pdo]
	Word-spotting (Experiment 2B)		
Mean RT	715	841	846
Mean error	20%	38%	29%
Example	asazu	asafte	asapdo

(C)VCVCV non-words. The materials were recorded and measured as in Experiment 1. Experiment 2A was a lexical decision control, with truncation applied as in Experiment 1A. In Experiment 2A, the procedure was as for Experiment 1A, and in Experiment 2B as for Experiments 1B/C.

2. Participants

Sixty-nine Dokkyo University undergraduates participated, for course credit: 24 in Experiment 2A, and 45 in Experiment 2B. None had taken part in Experiment 1.

B. Results

All participants (eight per list) in the lexical decision control experiment (Experiment 2A) were included in the analyses. Three word-spotting participants (Experiment 2B) were excluded (one listener missed all targets in one condition; the other two, excluded to balance the sets, were the most erroneous on each of the other two lists). This left 14 participants per list. The button-press responses on 17 trials (2.1%) in Experiment 2A and one trial (0.1%) in Experiment 2B were accompanied by incorrect spoken responses and thus treated as errors. The data from six items were excluded from the final analysis (*haru*, *motsu*, *hiru*, *mitsu*, *matsu*, *moya*): Five words because all participants missed them in at least one condition in one of the subexperiments, the sixth because it had been recorded incorrectly in one condition. Table IV shows mean RTs and error rates.

1. Lexical decision control (Experiment 2A)

The effect of the context from which words had been excised was significant by participants [$F1(2,42)=4.73$, $p<0.05$] but not by items [$F2(2,46)=1.34$, $p>0.2$] in the RT analysis. Errors (overall mean 28%) were again reasonable for excised words; the context effect in errors was significant in both analyses [$F1(2,42)=16.75$, $p<0.001$; $F2(2,46)=4.20$, $p<0.05$]. No pairwise comparison was significant by both participants and items for RTs. Error rates, however, were lower for words taken from CV contexts (e.g., *asa* from *asazu*) than for words from CCV contexts where devoicing was possible [*asa* from *asafte*; $t1(23)=3.91$, $p<0.005$; $t2(23)=2.94$, $p<0.01$], and, less robustly, for words from CCV contexts where devoicing was impossible [*asa* from *asapdo*; $t1(23)=2.16$, $p<0.05$; $t2(23)=1.76$, $p<0.1$]. The difference between the two CCV contexts was not significant. As in Experiment 1, these results suggest that the target words differed across contexts in how easy they were to recognize. These control data were therefore again used as covariates in by-item ANCOVAs of the word-spotting data. The by-participant analyses of the word-spotting data were again ANOVAs.

2. Word spotting (Experiment 2B)

There was a main effect of context in RTs [$F1(2,78)=17.96$, $p<0.001$; $F2(2,45)=11.44$, $p<0.001$] but not in errors [$F1(2,78)=24.76$, $p<0.001$; $F2(2,45)=1.95$, $p>0.15$]. Pairwise comparisons on the RT data showed that participants spotted words faster in CV contexts than in either type of CCV context (possible-devoicing environment: $F1(1,41)=25.14$, $p<0.001$; $F2(1,22)=18.98$, $p<0.001$; impossible environment: $F1(1,41)=28.05$, $p<0.001$; $F2(1,22)=14.81$, $p<0.005$). There was no latency difference between the two CCV context conditions ($F1$ and $F2<1$). No pairwise comparisons on the error data were significant by both participants and items. Word spotting was thus easier in CV contexts than in CCV contexts, with no effect of whether the CCV sequence was or was not a possible-devoicing environment.

C. Discussion

Again, the contexts which were potential devoicing environments made it very hard for listeners to spot the words embedded in the nonwords they heard. The results were very similar to those of Experiment 1; providing a potentially legal and hence more natural environment for devoicing to occur did not increase the acceptability of consonant sequences in Japanese speech segmentation. Both *asapdo* and *asafte* consist of three full morae (*a*, *sa*, and *do* or *te*) plus a single consonant ([p], [f]); it makes no difference that [f] before *te* could possibly have arisen from *fu*, whereas [p] before *do* cannot have arisen from *pu*. Vowelless sequences, it seems, are not treated as if they might be hiding a vowel; just as in other languages, such sequences are impossible word candidates and hence they resist being segmented from the adjacent speech stream.

Experiments 1 and 2 allow us to dispense with the possibility that Japanese listeners always treat sequences of voiceless consonants as if they contained a vowel; clearly, they do not. But results with following contexts, which potentially combine with word offsets, do not force us to conclude that devoicing will always disrupt the segmentation of normal Japanese speech. It is still necessary to assess devoicing in preceding contexts, which potentially combine with word onsets and thus exercise a more powerful effect in word segmentation. Preceding contexts are processed before targets, thus affecting target recognition differently than following contexts; their effect can be so strong that a target word is not even recognized at all (as occurred in the study

of McQueen *et al.*, 2001, for example). In our third experiment, we therefore examined devoicing sequences attached as preceding context to an embedded word.

Again we compared the devoicing sequences to other sequences which the previous findings had indicated as hard or as easy. We used two types of item: VCV words such as *asa*, preceded by clearly easy CCVCV versus hard CCVC contexts (see McQueen *et al.*, 2001), and CVCV words such as *sake* "salmon," preceded by clearly easy CCVCV contexts and potential devoicing CCVC contexts. Thus detection of *asa* was compared in *myojiasa* (easy, because the context consists of two full morae: *myo*, *ji*) versus *myochasa* (hard, because the context is a mora *myo* plus a vowelless affricate). The prohibition of devoicing before vowels (or any voiced segment) means that these vowel-initial words can never be preceded by devoicing; they are therefore the baseline against which we can compare the voiceless-initial words like *sake*. Detection of *sake* was compared in *nyagusake* (easy: *nya*, *gu*) versus *nyaksake* (hard: *nya*, [k], but potentially a rendition of *nyakusake* with devoicing). The results of McQueen *et al.* (2001) lead us to expect a large difference between easy and hard contexts for the VCV words like *asa*; the crucial question is whether there is also a large difference between easy and hard contexts for CVCV words like *sake*.

If there is an equivalently large difference in the *sake* case, we will have to conclude that devoicing indeed makes Japanese speech segmentation more difficult. The vowelless affricate in *myochasa* can never be licensed as a possible word, because devoicing cannot occur before vowels. Thus if the context effect for CVCV words is as large as for VCV words, we would have to conclude that the vowelless stop in *nyaksake* likewise cannot be a possible word.

On the other hand, if the context effect for CVCV words is significantly less than for VCV words (i.e., if *sake* is as easy or almost as easy to spot in *nyaksake* as in *nyagusake*), we may conclude that Japanese listeners interpret /ks/ (and like sequences) as containing an underlying vowel, and hence licensed as a possible word. Sequences such as /ks/ will often have been heard; the Japanese lexicon contains many words in which the first mora allows vowel devoicing, including words beginning *kusa-* (e.g., *kusabi* "wedge" and *kusari* "chain"), which will be pronounced with a devoiced first vowel, so effectively with an initial /ks/. If such experience licenses the devoiced sequence /ks/ in *nyaksake* as a possible word, it should not interfere with segmentation.

IV. EXPERIMENT 3

A. Method

1. Materials and procedure

Fifty-two high-frequency two-mora target words were selected: 26 VCV words (e.g., *asa*) and 26 CVCV words beginning with voiceless consonants (e.g., *sake*). CCVC (e.g., *bya*+C) and CCVCV (e.g., *bya*+CV) nonsense sequences were constructed as preceding contexts for both types of target. For the CVCV targets, the CCVC contexts created possible-devoicing environments between the context and target word onset (e.g., *nyaksake* is a potentially devoiced rendition of *nyakusake*). This was not the case for the CCVC contexts followed by the vowel-initial VCV targets, because devoicing cannot occur before vowels.

We further constructed 106 nonsense fillers with no embedded bimoraic words in offset position. Of these, 52 were matched to target-bearing items: 26 ending with CVCV after either a CCVC or CCVCV sequence like those in the target-bearing items (e.g., *chaksomi* and *kyagukeni*), and 26 ending with VCV sequence after a CCVC or CCVCV (e.g., *gyopagi* and *myoguige*). The remaining 54 fillers were all trimoraic CVCVCV nonwords. There were 24 practice items modeled on the experimental materials, including eight with embedded target words (four CVCV, four VCV).

The materials were recorded as in Experiments 1 and 2, by the same speaker. The accent pattern of the targets was preserved in the way the items were recorded. The items with CVCV targets in CCVC contexts were recorded with the potential underlying vowel fully devoiced (e.g., the [u] in *nyak(u)sake* was not realized). The target-bearing stimuli were then digitally cross-spliced, cutting at zero-crossings and using auditory criteria to determine the excision points. Tokens of the CVCV targets recorded in the CCVCV context (e.g., *sake* from *nyagusake)* were spliced onto a CCVC context sequence (e.g., *nyak* from *nyaksake*) and onto a CCVCV sequence from a second recording of a CCVCV item (e.g., *nyagu* from a different token of *nyagusake* than that used for the target token). A similar cross-splicing procedure was used for the VCV targets (e.g., the *asa* in the final experimental stimuli came from a recording of *myojiasa*). All target-bearing stimuli were thus cross-spliced from two recordings, and the target word realization in all contexts was constant. Six items with VCV targets could not be cross-spliced without audible discontinuities and were therefore excluded (along with six matched fillers) from the experiment. There was no detectable trace of splicing in the remaining items (all listed in Tables V and VI).

Two counter-balanced lists were constructed, with 13 CVCV and 10 VCV targets in each context condition per list, and with all 46 targets appearing once per list. As in the earlier experiments, target-bearing and filler items were presented in pseudo-random order. The procedure was as in Experiments 1B/C and 2B.

2. Participants

Thirty-two undergraduates (16 per list) from the same population took part in return for course credits. None had participated in Experiments 1 or 2.

B. Results and discussion

No participants were excluded from the analyses, but one item was. This word (*soro*, "solo," from the CVCV set) was missed by all participants in the devoiced vowel context and by all but one participant in the surface vowel context. No incorrect spoken responses were recorded. Mean word-spotting RTs and error rates are shown in Table VII. Note that since the materials in this experiment were controlled by the cross-splicing, control lexical decision data, as required for Experiments 1 and 2, are here unnecessary.

TABLE V. Materials for Experiment 3: CVCV targets.

Target	English gloss	Voiceless consonant+devoiced vowel	Voiced consonant+vowel
		Many word candidates	
kachi	value	gyats(u)kachi	gyazukachi
kado	corner	kyah(u)kado	kyabukado
kako	past	nyach(i)kako	nyajikako
kashu	singer	byas(u)kashu	byazukashu
kata	shoulder	shof(u)kata	shobukata
kare	he	byos(u)kare	byozukare
saru	monkey	gyuk(u)saru	gyugusaru
kumo	cloud	nyos(u)kumo	nyozukumo
tsuru	crane	nyosh(i)tsuru	myojitsuru
kazu	number	shas(u)kazu	shazukazu
sake	salmon	nyak(u)sake	nyagusake
kaba	hippopotamus	nyach(i)kaba	nyajikaba
kage	shadow	shos(u)kage	shozukage
kechi	stinginess	byaf(u)kechi	byabukechi
		Few word candidates	
fugu	blowfish	byas(u)fugu	byazufugu
hamu	ham	nyak(u)hamu	nyaguhamu
haru	spring	gyash(u)haru	gyajuharu
hage	baldness	chosh(u)hage	chojuhage
hada	skin	nyots(u)hada	nyozuhada
kamo	duck	myosh(u)kamo	myojukamo
sora	sky	kyof(u)sora	kyobusora
shuwa	sign language	nyok(i)shuwa	nyogishuwa
shugo	subject	pyach(i)shugo	pyajishugo
chibi	kid	ryosh(u)chibi	ryojuchibi
sobo	grandmother	myak(u)sobo	myagusobo
soro	solo	myok(u)soro	myogusoro

TABLE VI. Materials for Experiment 3: VCV targets.

Target	English gloss	Voiceless consonant+vowel	Voiced consonant+vowel
ase	sweat	gyasase	gyazuase
aka	red	myachaka	myajuaka
ani	brother	gyachani	gyajuani
eki	station	nyaheki	nyabueki
aki	autumn	shochaki	shojuaki
ama	nun	ryochama	ryojiama
aji	horse mackerel	chochaji	chojuaji
ibo	wart	nyahibo	nyabuibo
umi	sea	shochumi	shojiumi
aku	badness	nyoshaku	nyojuaku
asa	morning	myochasa	myojiasa
ato	mark	kyachato	kyajuato
ane	sister	shasane	shazuane
uni	sea urchin	byachuni	byajiuni
ego	ego	gyosego	gyozuego
oku	inside	myahoku	myabuoku
ine	rice plant	chuchine	chujuine
imi	meaning	kyochimi	kyojuimi
obi	a sash	ryuhobi	ryubuobi
ima	now	gyuchima	gyujuima

The most striking result is in the error rates. As in McQueen *et al.* (2001), Japanese listeners found it almost impossible to spot VCV target words in a CCVC context where the initial vowel of the target was not aligned with a mora boundary (e.g., *asa* in *myochasa*). Error ANOVAs revealed main effects of target type (VCV words were harder to spot than CVCV words: $F1(1,30)=107.84$, $p<0.001$; $F2(1,43)=19.00$, $p<0.001$) and context type [words were harder to spot in CCVC contexts than in CCVCV contexts: $F1(1,30)=302.30$, $p<0.001$; $F2(1,43)=134.72$, $p<0.001$], and these two factors interacted [$F1(1,30)=90.97$, $p<0.001$; $F2(1,43)=54.76$, $p<0.001$]. Pairwise comparisons showed that the context effect was significant for both types of target: CVCV words (*sake*) were harder to spot in CCVC contexts (possible devoicing environments, e.g., *nyaksake*) than in CCVCV contexts (e.g., *nyagusake*), $F1(1,30)=15.13$, $p<0.001$, $F2(1,24)=9.90$, $p<0.005$; as already noted, VCV targets (*asa*) were much harder to spot in CCVC contexts (e.g., *myochasa*) than in CCVCV contexts (e.g., *myojiasa*): $F1(1,30)=368.79$, $p<0.001$, $F2(1,19)=163.75$, $p<0.001$.

Given the high error rates for the VCV control words, the context effect in RTs was analyzed only for the CVCV words. CVCV targets (e.g., *sake*) were detected equally rapidly in the two contexts ($F1$ and $F2<1$); there was in this case no increased difficulty for devoicing (*nyak-*) over vowel (*nyagu-*) contexts. Within the CCVCV contexts, the VCV words were spotted as quickly as the CVCV words ($F1$ and $F2<1$).

The results for the VCV baseline condition show that the listeners here were behaving exactly as the listeners in the study of McQueen *et al.* (2001). Preceding context without a vowel (e.g., a single affricate) makes segmentation and hence word-spotting hard. The crucial results are for the CVCV words. Here the devoicing contexts were clearly less problematic for listeners than they had proven to be in Experiments 1 and 2. Although the error rate was raised by a vowelless context, it was not raised to the heights observed for the VCV words (or by McQueen *et al.*, 2001). And in the RTs, no delay of word-spotting as a function of context could be observed at all. This striking finding suggests that Japanese listeners are indeed sensitive to the potential presence of a devoiced vowel in voiceless obstruent sequences such as /ks/.

TABLE VII. Experiment 3: Mean correct RTs (in milliseconds, from target word offset) and mean error rates (in percent) by context condition (*sake*=salmon; *asa*=morning; no reliable estimate could be computed of the mean RT for VCV targets in CCVC contexts).

	Target type	Context: CCVC	Context: CCVCV
	CVCV		
Mean RT		739	755
Mean error		43%	29%
Example		nyaksake	nyagusake
	VCV		
Mean RT		···	750
Mean error		93%	29%
Example		myochasa	myojiasa

We conducted two further analyses of the data, neither of which explained away this finding. We first examined the phonetic structure of the devoicing contexts we had tested. Maekawa and Kikuchi's (2005) analyses of the corpus of spontaneous Japanese showed devoicing to vary as a function of the surrounding consonants. The manner of articulation of the following consonant has the strongest effect: the likelihood of the vowel being devoiced is far greater before a stop or affricate than before a fricative, both for /i/ and for /u/ and across all types of preceding consonants. According to Kondo (2005), devoicing is virtually obligatory in the most favored environments, and is only inhibited for potential sequences of devoiced syllables. If listeners are sensitive to these probabilities, they may find the most likely cases least difficult. Accordingly, we divided the CVCV items into two sets, varying in frequency of devoicing occurrence. One set contained 11 words beginning with fricatives (in Maekawa and Kikuchi's (2005) data, devoiced in about 61% of cases), while the other set had 14 targets beginning with stops and affricates (about 96% devoiced). An analysis of the error data including this factor revealed a significant context effect [$F1(1,30)=17.92$, $p<0.001$; $F2(1,23)=9.73$, $p<0.005$]: CCVC contexts made word-spotting harder than CCVCV contexts. But there was no effect of devoicing probability ($F1$ and $F2<1$) and no interaction of this factor with the context effect ($F1$ and $F2<1$). The context effect was present both where devoicing is very likely [$F1(1,30)=6.60$, $p<0.05$; $F2(1,13)=6.74$, $p=0.05$], and where it is less likely [$F1(1,30)=13.48$, $p<0.001$; $F2(1,10)=3.74$, $p<0.1$]. Thus this factor appeared not to have affected our results.

The second analysis examined the potential lexical support for devoiced vowels in our sequences. From Sugito (1995) we computed the number of words consistent with the bimoraic sequence linking CCVC context and each CVCV target word (e.g., *kusa* given the target *sake* in *nyaksake*, if the devoiced [u] were realized). The materials formed two sets (see Table V), one of 14 items where relatively many words (≥ 10) either matched the sequence or had onsets the same as the sequence (e.g., *kusa* in *nyaksake*) and one of 11 items (minus *soro*) with few words (≤ 5) matching the sequence (e.g., *shuha* given the target *haru* "spring" in the CCVC+target sequence *gyashharu*). The mean number of competitors for the former set was 13.6, for the latter 2.0. ANOVAs on the error data for the CVCV words split in this way revealed a main effect of word set size by participants but not by items [$F1(1,30)=6.52$, $p<0.05$; $F2(1,23)=1.10$, $p>0.3$]: Targets in the many-words set were spotted more accurately, across context conditions, than those in the few-words set. There was also still a main effect of context [$F1(1,30)=22.06$, $p<0.001$; $F2(1,23)=10.16$, $p<0.005$]. In the by-participant analysis only, the context effect varied with word set size: Words were harder to spot in CCVC than in CCVCV contexts, but more so when fewer lexically consistent words were available [$F1(1,30)=6.97$, $p<0.05$; $F2<1$]. As pairwise comparisons confirmed, the context effect was less robust in the many-words set [$F1(1,30)=2.50$, $p>0.1$; $F2(1,13)=4.52$, $p=0.05$] than in the few-words set [$F1(1,30)=32.91$, $p<0.001$; $F2(1,10)=5.20$, $p<0.05$]. This suggests that greater lexical support may strengthen licensing of devoicing environments, but is not the sole rationale for it.

V. GENERAL DISCUSSION

Our first conclusion must be that vowel devoicing in Japanese certainly does not make speech processing easier. The basic perceptual difficulty of consonant sequences holds as much for Japanese listeners as for listeners to any other language: consonants are easier to process if they are adjacent to vowels. The differences in phonological functionality of vowels and consonants likewise hold as strongly for Japanese as for other languages: a vowel can stand alone as a syllable and hence as a word, but a consonant in general cannot. Finally, the role of this vowel-consonant asymmetry in segmenting continuous speech is also parallel for Japanese and for other listeners.

Research on spoken-word recognition has amassed abundant evidence that speech input concurrently activates many word candidates which it fully or partially supports; because all languages construct very large vocabularies from relatively few phonemes, words in all languages exhibit a great deal of embedding and overlap, so that many such candidates will be unintended competitors which need to be rejected if the real message is to be recognized. Many mechanisms exist to deal with this competition (see McQueen, 2007, for a review). The PWC, which allows competitors to be rejected if accepting them would strand a vowelless residue of the input, enables English listeners to suppress unwanted activation of *egg* when they hear *leg* or *legacy*, and by the same token it enables Japanese listeners to suppress unwanted activation of *asa* when they hear *kasa* "umbrella." The operation of the PWC runs parallel in English (Norris *et al.*, 1997, 2001) and Japanese (McQueen *et al.*, 2001). Our new experiments have shown that the widespread phenomenon of devoicing in spoken Japanese does not modulate the power of this effect at all. Even consonants which could accompany a devoiced vowel made recognition of an adjacent word hard, to effectively the same extent as consonants which could not have preceded devoicing. Though recognition of *asa* was relatively easy before a vowel, it was hard before a vowelless consonant, even one which might have supported devoicing (e.g., the voiceless fricative /f/).

These segmentation effects operate upon words activated by speech input. The equivalence of the vowelless environments which do support devoicing and those which do not implies that during prelexical processing vowels are not automatically restored or inserted into either environment. In this our results match with those from a recent study by Mash *et al.* (2006), who tested the effect of vowel devoicing on the compensation for coarticulation which shifts identification of a /t-k/ continuum following /s/ versus /ʃ/ (Mann and Repp, 1981). Japanese has no cluster onsets and no coda obstruents, so such sequences are not lexically possible, but /sk/, /st/, /ʃk/, and /ʃt/ sequences can indeed arise from devoicing. The necessary environments for compensation for coarticulation thus exist in practice. But if the underlying vowel had been perceptually restored at the prelexical level,

there would be no such compensation because the consonants would not be sequential, but would be interrupted by a vowel. Mash *et al.* (2006) found, however, that Japanese listeners performing /t/-/k/ categorization in nonwords produced more /k/ responses in *rusko* than in *rushko*—exactly the compensation effect that Mann and Repp (1981) had found. This suggests, in agreement with our own findings, that the vowels are not there in any sense which would affect prelexical processing of the auditory signal.

Our second conclusion, however, is that devoiced vowels will not in practice cause word recognition difficulty. In real speech, vowel devoicing will be encountered in known words. By presenting nonwords in an experiment, we can show that vowelless sequences are not automatically furnished with vowels. But in normal listening, listeners are rarely presented with nonwords. The results of Experiment 3 suggest that vowelless sequences arising by devoicing, though nominally illegal in Japanese phonology, will be effectively licensed as possible words. This may be especially so when stored lexical representations are activated—words with a devoiced initial syllable (e.g., *sukiyaki*, "sukiyaki" beginning /sk/) activate the intended lexical candidates. In Experiment 3, spotting *sake* was as fast in *nyaksake* as in *nyagusake*, and the error difference was much smaller than for the corresponding VCV targets. This may have been because the /ksa/ sequence was consistent with words beginning *kusa-,* i.e., with words which would be pronounced with initial /ksa/. Note that these candidates would, of course, compete for recognition with *sake*, but because *sake* was fully supported by the input while the other candidates were not, *sake* would win this competition. The competition offered by /ksa/ would presumably be no lesser or greater than the competition offered by the voiced-consonant context; /gusa/ would also have activated competitors, such as *gusaku* "rubbish." Note also that, in contrast to Experiments 1 and 2, the potentially devoiced sequences in Experiment 3 preceded the target words. There was thus time for words consistent with those sequences to be retrieved, and hence for the vowelless sequences to be licensed. This result contrasts particularly strongly with that for the VCV words which support no such licensed candidates; here the finding of McQueen *et al.* (2001) was replicated, in that the preceding context effectively blocked access to the embedded words completely. Words such as *asa* embedded in contexts such as *myochasa* were just as hard to detect as items in McQueen *et al.* (2001) such as *uni* in *gyabuni*; in both cases the PWC operated effectively to inhibit a parse of the input which would leave a vowelless residue ([tʃ], [b]). That this did not happen with *sake* in *nyaksake* thus constitutes powerful evidence of how listeners cope with the effects of devoicing in practice. In Experiments 1 and 2, where the potentially devoiced sequences followed the target words, no comparably licensed interpretation was available while the targets were being heard, so the underlying equivalence of a consonantal sequence arising from devoicing and any other (illegal) consonantal sequence emerged; the PWC again played its universal role, and segmentation was interfered with.

Other word recognition evidence from Japanese also suggests that devoiced forms effectively activate lexical representations. Ogasawara and Warner (2009) found that lexical decisions were faster for words such as *hashika* "measles" if the potentially devoiced vowel was reduced than if it was fully articulated. Further, although in nondevoicing environments phoneme detection responses were slower to reduced /i/ than to fully articulated /i/, in a devoicing environment the shorter, less clear, reduced vowel was responded to no more slowly than the longer, clearer full vowel. The lack of disadvantage for the acoustically less clear vowels in Ogasawara and Warner's phoneme detection study strongly suggests that the responses were not based on prelexical processing alone, but drew on lexical evidence (as the phoneme detection task allows; Norris *et al.*, 2000) to support the vowel interpretation.

Thus Japanese listeners appear not to restore devoiced vowels prelexically; the spoken forms with devoicing are perfectly functional in word recognition. Encountering a completely new word with a devoiced vowel could, of course, produce segmentation difficulty. Moreover, another indirect effect of devoicing might cause recognition delay. Consider that an important source of information for word recognition in Japanese is word accent pattern; whether a syllable is accented or not is perceived from only a fraction of the vowel (Cutler and Otake, 1999), and the accent pattern of a spoken fragment allows rapid rejection of alternative words with different accent (Sekiguchi and Nakajima, 1999). If a vowel is devoiced, the pitch information necessary for this efficient use of accent is no longer available, and this has been held to be the reason why devoicing is disfavored in accented syllables (Vance, 1987). However, devoicing of accented syllables does indeed occur, and indeed is becoming more common (Sugito, 1982; Kitahara, 1998). For a single devoiced syllable, accentedness may be accurately apprehended from compensatory pitch modification in an immediately following syllable (Sugito, 1982; Sugito and Hirose, 1988; Maekawa, 1990). If, however, one of a sequence of devoiced syllables is accented, listeners can tell that there was an accented syllable (i.e., they accurately distinguish unaccented sequences from sequences containing an accent), but they cannot reliably tell on which syllable the accent fell (Maekawa, 1990). Thus, as well as possibly delaying segmentation, devoicing could also interfere with word recognition via disruption of accent information.

Our results thus indicate that although devoiced sequences may effectively access lexical representations, without any prelexical restoration of vowels being necessary, the cross-linguistically observed disadvantage for vowelless sequences in speech segmentation holds in Japanese as strongly as it does in other languages.

ACKNOWLEDGMENTS

This research was supported by the NWO-SPINOZA project "Native and Non-Native Listening" (AC) and by Japan Society for the Promotion of Science Grant No. 11610566 (TO). We thank Aoju Chen, Maarten Jansonius, and Bram de Kruijff for technical assistance, and Natasha

Warner, Kikuo Maekawa, and an anonymous reviewer for helpful comments.

APPENDIX A: PHONEMES OF JAPANESE

Vowels: /a,e,i,o,u/; Stops: /p,t,k,b,d,g/; flap: /ɾ/; nasals: /m,n,ɴ/; approximants: /w,j/; affricates: /c/; fricatives: /s,z,h/

[1]Additional analyses compared item subsets. No differences between the sub-groups of the items in which the consonant in the "impossible" environment was voiced (e.g., *gasub*) versus voiceless (e.g., *asap*) appeared in either Experiment 1B or 1C. The sub-group of items in which the impossible-devoicing context contained [t] or [d] were somewhat harder than items containing other consonants in this context, in both Experiments 1B and 1C, but the context effects were the same across sub-groups (no context by consonant interactions were significant).

Cutler, A., Demuth, K., and McQueen, J. M. (**2002a**). "Universality versus language-specificity in listening to running speech," Psychol. Sci. **13**, 258–262.

Cutler, A., McQueen, J. M., Jansonius, M., and Bayerl, S. (**2002b**). "The lexical statistics of competitor activation in spoken-word recognition," *Proceedings of the Ninth Australian International Conference Speech Science and Technology*, edited by C. Bow (Australian Speech Science and Technology Association, Canberra), pp. 40–45.

Cutler, A., and Norris, D. (**1988**). "The role of strong syllables in segmentation for lexical access," J. Exp. Psychol. Hum. Percept. Perform. **14**, 113–121.

Cutler, A., and Otake, T. (**1999**). "Pitch accent in spoken-word recognition in Japanese," J. Acoust. Soc. Am. **105**, 1877–1888.

Dumay, N., Frauenfelder, U. H., and Content, A. (**2002**). "The role of the syllable in lexical segmentation in French: Word-spotting data," Brain Lang **81**, 144–161.

Dupoux, E., Kakehi, K., Hirose, Y., Pallier, C., and Mehler, J. (**1999**). "Epenthetic vowels in Japanese: A perceptual illusion?," J. Exp. Psychol. Hum. Percept. Perform. **25**, 1568–1578.

Kitahara, M. (**1998**). "The interaction of pitch accent and vowel devoicing in Tokyo Japanese," in *Japanese-Korean Linguistics*, Vol. **8** (CSLI & SLA, Stanford, CA), pp. 303–315.

Kondo, M. (**2005**). "Syllable structure and its acoustic effects on vowels in devoicing environments," in *Voicing in Japanese*, edited by J. van de Weijer, K. Nanjo, and T. Nishihara (Mouton de Gruyter, Berlin), pp. 229–246.

Liberman, A. M., Delattre, P. C., Cooper, F. S., and Gerstman, L. J. (**1954**). "The role of consonant-vowel transitions in the perception of the stop and nasal consonants," Psychol. Monogr. **68**, 1–13.

Maekawa, K. (**1990**). "Production and perception of the accent in the consecutively devoiced syllables in Tokyo Japanese," Proceedings of the First International Conference on Spoken Language, ICSLP 90, Kobe, Japan, pp. 517–520.

Maekawa, K. (**2003**). "Corpus of spontaneous Japanese: Its design and evaluation," Proceedings of ISCA/IEEE Workshop Spontaneous Speech Processing and Recognition, SSPR2003, Tokyo.

Maekawa, K., and Kikuchi, H. (**2005**). "Corpus-based analysis of vowel devoicing in spontaneous Japanese: An interim report," in *Voicing in Japanese*, edited by J. van de Weijer, K. Nanjo, and T. Nishihara (Mouton de Gruyter, Berlin), pp. 205–228.

Mann, V. A., and Repp, B. H. (**1981**). "Influence of preceding fricative on stop consonant perception," J. Acoust. Soc. Am. **69**, 548–558.

Mash, D., Kawahara, S., Kingston, J., Brenner-Alsp, K., and Chambless, D. (**2006**). "Sequential contrast versus compensation for coarticulation in Japanese versus English," Paper presented to the Acoustical Society of America, Providence, RI.

McQueen, J. (**1996**). "Word-spotting," Lang. Cognit. Processes **11**, 695–699.

McQueen, J. M. (**1998**). "Segmentation of continuous speech using phonotactics," J. Mem. Lang. **39**, 21–46.

McQueen, J. M. (**2007**). "Eight questions about spoken-word recognition," in *The Oxford Handbook of Psycholinguistics*, edited by G. Gaskell (Oxford University Press, Oxford), pp. 37–53.

McQueen, J. M., and Cutler, A. (**1998**). "Spotting (different types of) words in (different types of) context," Proceedings of the Fifth International Conference on Spoken Language Processing, Sydney, Vol. **6**, pp. 2791–2794.

McQueen, J. M., Otake, T., and Cutler, A. (**2001**). "Rhythmic cues and possible-word constraints in Japanese speech segmentation," J. Mem. Lang. **45**, 103–132.

Norris, D., McQueen, A., Cutler, J. M., Butterfield, S., and Kearns, R. (**2001**). "Language-universal constraints on speech segmentation," Lang. Cognit. Processes **16**, 637–660.

Norris, D., McQueen, J. M., and Cutler, A. (**2000**). "Merging information in speech recognition: Feedback is never necessary," Behav. Brain Sci. **23**, 299–325.

Norris, D., McQueen, J. M., Cutler, A., and Butterfield, S. (**1997**). "The possible-word constraint in the segmentation of continuous speech," Cogn. Psychol. **34**, 191–243.

Ogasawara, N., and Warner, N. (**2009**). "Processing missing vowels: Allophonic processing in Japanese," Lang. Cognit. Processes. In press.

Racine, I., and Grosjean, F. (**2000**). "The influence of schwa deletion on the recognition of words in continuous speech," Année Psychol. **100**, 393–417.

Sekiguchi, T., and Y., Nakajima, (**1999**). "The use of lexical prosody for lexical access of the Japanese language," J. Psycholinguist. Res. **28**, 439–454.

Spinelli, E., McQueen, J. M., and Cutler, A. (**2003**). "Processing resyllabified words in French," J. Mem. Lang. **48**, 233–254.

Sugito, M. (**1982**). *Nihongo akusento no kenkyuu (Studies on Japanese accent)* (Sanseido, Tokyo).

Sugito, M. (**1995**). *Osaka-Tokyo Akusento Onsei Jiten (Osaka-Tokyo Accent Pronunciation Dictionary)* (Maruzen, Tokyo).

Sugito, M., and Hirose, H. (**1988**). "Production and perception of accented devoiced vowels in Japanese," Annual Bulletin Research Institute of Logopedics and Phoniatrics **22**, 19–37.

Suomi, K., McQueen, J. M., and Cutler, A. (**1997**). "Vowel harmony and speech segmentation in Finnish," J. Mem. Lang. **36**, 422–444.

van der Lugt, A. H. (**2001**). "The use of sequential probabilities in the segmentation of speech," Percept. Psychophys. **63**, 811–823.

van Donselaar, W., Kuijpers, C., and Cutler, A. (**1999**). "Facilitatory effects of vowel epenthesis on word processing in Dutch," J. Mem. Lang. **41**, 59–77.

van Son, R. J. J. H., and Pols, L. C. W. (**1995**). "The influence of local context on the identification of vowels and consonants," Proceedings of Eurospeech95, Madrid, pp. 967–970.

Vance, T. J. (**1987**). *An Introduction to Japanese Phonology* (State University of New York Press, Albany, NY).

Warner, N., Kim, J., Davis, C., and Cutler, A. (**2005**). "Use of complex phonological patterns in processing: Evidence from Korean," J. Linguist. **41**, 353–387.

Weber, A., and Cutler, A. (**2006**). "First-language phonotactics in second-language listening," J. Acoust. Soc. Am. **119**, 597–607.

Yip, M. C. (**2004**). "Possible-word constraints in Cantonese speech segmentation," J. Psycholinguist. Res. **33**, 165–173.

Temporal integration in vowel perception

Andrew B. Wallace and Sheila E. Blumstein
Department of Cognitive and Linguistic Sciences, Brown University, Box 1978, Providence, Rhode Island 02912

(Received 7 January 2008; revised 16 December 2008; accepted 12 January 2009)

Psychoacoustic research suggests that multiple auditory channels process incoming sounds over temporal windows of different durations, resulting in multiple auditory representations being available to higher-level processes. The current experiments investigate the size of the temporal window used in vowel quality perception using an acoustic priming paradigm with nonspeech and speech primes of varying duration. In experiment 1, identification of vowel targets was facilitated by acoustically matched nonspeech primes. The magnitude of this effect was greatest for the shortest (25 and 50 ms) primes, remained level at medium (100 and 150 ms) duration primes, and declined significantly at longer prime durations, suggesting that the auditory stages of vowel quality perception integrate sensory input over a relatively short temporal window. In experiment 2, the same vowel targets were primed by speech stimuli, consisting of vowels using the same duration values as those in experiment 1. A different pattern of results emerged with the greatest priming effects found for primes of around 150 ms and less priming at the shorter and longer durations, indicating that longer-scale temporal processes operate at higher levels of analysis. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3077219]

PACS number(s): 43.71.Es, 43.71.Rt, 43.66.Lj [AJ] Pages: 1704–1711

I. INTRODUCTION

Most models of speech perception assume a form of hierarchical processing, in which the acoustic input undergoes a number of transformations as it is mapped onto lexical form (cf. McClelland and Elman, 1986; Blumstein, 1995). The early stages of this process, which appear to involve the extraction of relatively simple auditory features (see Versnel and Shamma, 1998 for one possibility), are likely to play a critical role in speech perception, transforming the sensory input in a way that both shapes and facilitates higher-level phonetic perception. The current study examines one important aspect of this auditory processing: the size of the temporal window of auditory analysis in the early stages of vowel quality perception (i.e., the identification of a vowel's phonetic category, which can be contrasted with the perception of other information encoded in a vowel such as pitch, affect, or prosody).

Temporal integration—the summation of acoustic energy over time—is fundamental to virtually all models of auditory function (cf. Viemeister and Wakefield, 1991). Mathematically, this process is often modeled as a "leaky integrator" whose level of activation is increased by new acoustic stimulation (within a particular frequency band), then decays exponentially. The rate of decay can be characterized by a time constant in the exponential decay function or by the length of the window within which activity decays to a certain percentage of its original value. The size of this temporal window is of considerable interest in psychoacoustics, since it characterizes the temporal resolution of the auditory system.

In attempts to measure the size of this window with behavioral data, different perceptual tasks have yielded vastly different estimates. Studies of gap detection (Plomp, 1964; Penner, 1977; Fitzgibbons, 1983; Shailer and Moore, 1983, 1987), the perception of temporal order (Hirsh, 1959; Pisoni, 1977), the detection of temporal modulation (Viemeister, 1979), and asynchronous masking (Penner and Cudahy, 1973; Moore *et al.*, 1988) have all pointed to a relatively short temporal window of, at most, a few tens of milliseconds for auditory processing. On the other hand, studies of the effect of stimulus duration on hearing thresholds (Plomp and Bouman, 1959; Zwislocki, 1960) or overall loudness (Lifshitz, 1933; Munson, 1947) have pointed to integration over much longer windows of a few hundred milliseconds. Although other explanations have been proposed (cf. Penner, 1978; Viemeister and Wakefield, 1991), this discrepancy has generally been taken to indicate the existence of at least two distinct temporal windows, which comprise separate channels for auditory processing. Poeppel (2003) suggested a short window of 20–40 ms and a long window of 150–250 ms. Incoming sounds are presumably analyzed simultaneously in both channels—different kinds of sounds cannot be directed to different channels, since the sounds can only be identified *after* the initial auditory processing has taken place. The outputs of these channels then map onto higher-level perceptual mechanisms for speech perception, timbre analysis, and so on.

In the case of speech perception, many of the acoustic cues to consonants involve rapid spectral changes, and it is generally assumed that the processing of these spectral changes depends on input from a short-window channel (cf. Zatorre *et al.*, 2002; Poeppel, 2003). Vowels, on the other hand, are encoded by relatively stable spectral patterns. Spectral motion is still important, both in the form of consonantal formant transitions (cf. Strange, 1989) and vowel-intrinsic diphthongization (Nearey and Assmann, 1986; Assmann and Katz, 2000), but these cues are not essential, and synthetic stimuli with steady-state vowels can easily be

 0001-4966/2009/125(3)/1704/8/$25.00

perceived as vowels (Delattre *et al.*, 1952). Because of this, some authors suggested that vowel quality perception depends on input from a long temporal window channel (Samson and Zatorre, 1994; Johnsrude *et al.*, 1997; Poeppel, 2003).

The suggestion that vowel quality perception takes input from a long-window auditory channel has not been empirically studied in great depth, but is tentatively supported by neuroimaging results. Since current neuroimaging methods lack the spatial resolution to isolate specific populations of neurons, this research has focused on potential differences between the two hemispheres in the computational mechanisms involved in auditory processing. Left-hemisphere auditory areas have been found to respond best to rapid acoustic changes, while right-hemisphere areas respond best to steady-state or slowly-changing stimuli (Belin *et al.*, 1998; Giraud *et al.*, 2000; Zatorre and Belin, 2001; Poeppel *et al.*, 2004; Zaehle *et al.*, 2004; Boemio *et al.*, 2005; Jamison *et al.*, 2006). Similarly, the response to the rapidly-changing segmental content of speech is left-lateralized, while the response to the slowly-changing prosodic content is right-lateralized (Hesling *et al.*, 2005; Wartenburger *et al.*, 2007). Further evidence for such hemispheric asymmetries in temporal processing comes from electro- and magnetoencephalographic studies showing stimulus-induced changes in rhythmic activity at frequencies related to the suggested short and long temporal integration windows (Luo *et al.*, 2005; Luo and Poeppel, 2007; Giraud *et al.*, 2007).

Poeppel (2003) suggested that these hemispheric differences can be explained by an asymmetric distribution of neurons with long and short temporal windows across the two hemispheres. Although both types of neurons are bilaterally distributed, long-window (150–250 ms) cells are proposed to be more abundant in the right hemisphere, while short-window (20–40 ms) cells are proposed to be more abundant in the left, creating an overall difference in temporal resolution between the two hemispheres. Differences in window length could also affect spectral resolution, since longer temporal windows allow for a more detailed spectral analysis. Thus, Poeppel (2003) proposed that the right-lateralized, long-window channel has a high spectral resolution, while the left-lateralized, short-window channel has a low spectral resolution.

Using Poeppel's (2003) framework, hemispheric asymmetries in auditory processing might offer evidence of the temporal window involved in the perception of vowel quality. A left-hemisphere advantage would suggest a short temporal window, while a right-hemisphere advantage would suggest a long window. Obleser *et al.* (2007) found bilateral activation during the perception of vowels, with what appeared to be a right-hemisphere preference, although this laterality was not tested statistically. In another functional neuroimaging study, Britton *et al.* (2009) studied the processing of vowels and pure tones at durations of 75, 150, and 300 ms. For both vowels and tones at the longer stimulus durations, more activation was found in the right hemisphere than the left. These findings are consistent with suggestions that vowel quality perception draws its input from a right-lateralized auditory channel with a long temporal integration window.

Behavioral findings, however, seem to contradict the neuroimaging results, suggesting that the perception of vowel quality is based on auditory processing over a short temporal window. Holt *et al.* (2000), applying a method developed by Lotto and Kluender (1998), found a shift in the boundary between 60 ms tokens of the vowels [ɛ] and [ʌ] in a continuum of synthetic stimuli varying only in F2 when these stimuli were flanked by 70 ms pure tones at the onset F2 frequencies of [d] and [b]. The direction of this shift was contrastive, meaning that vowels were less likely to be identified as belonging to the category whose F2 frequencies were closer to those of the flanking tones. The authors used these results to argue that perceptual compensation for coarticulation could be explained by a general auditory contrast mechanism rather than the speech-specific mechanism proposed by Mann (1980). In the context of the current study, however, their findings suggest that the auditory input for vowel quality perception has a high enough temporal resolution to resolve 60–70 ms stimuli, which in turn suggests a relatively short temporal integration window.

Additionally, a series of studies by Chistovich (1985) suggests that the auditory processing of vowels for phonetic perception involves low spectral resolution, a feature that has been linked to the short-window channel (Poeppel, 2003). Previous studies had observed that two formants, in close proximity, combine to form a single perceived spectral peak (Fant, 1956; Carlson *et al.*, 1970), a phenomenon with important implications for vowel quality perception (cf. Stevens, 1972, 1989). Chistovich and Lublinskaja (1979) and Chistovich and Sheikin (1979) showed that the range at which such mergers occurred was around 3–3.5 bark (expressed in the critical-band rate frequency scale of Zwicker (1961) and Zwicker and Terhardt (1980). This 3–3.5 bark critical distance is large relative to the frequency resolution of the peripheral auditory system. Thus, Chistovich (1985) suggested that the auditory analysis of vowel spectra might involve a relatively wide-band spectral integration, which would result in a lower spectral resolution.

Low spectral resolution may not at first appear to be advantageous for vowel quality perception, since the primary determinant of vowel quality is a spectral cue, the pattern of formant frequencies (Delattre *et al.*, 1952). Formants, however, occur on a relatively large spectral scale. Finer levels of spectral detail provide information mainly about harmonics of the fundamental frequency of phonation (F_0). This suggests that the perception of vowel quality can be accomplished at a relatively low spectral resolution. Indeed, increasing spectral resolution to the point where individual harmonics can be resolved might impede the extraction of formants, since the spectral peaks resulting from individual harmonics might be mistaken for formants. Consistent with this is evidence showing that vowel identification becomes less accurate as F_0 increases and individual harmonics become more prominent (Ryalls and Lieberman, 1982; Diehl *et al.*, 1996). Thus, the optimal resolution for vowel quality perception may be found in the low-spectral-, high-temporal-

resolution channel rather than the high-spectral-, low-temporal-resolution channel (but cf. Cheveigné and Kawahara, 1999; Hillenbrand and Houde, 2003).

A number of behavioral methods could be used to experimentally test the temporal resolution of vowel quality perception. Past research has investigated vowel perception by measuring the temporal limits of forward and backward masking. Results have shown backward masking effects of vowel maskers on very short vowel targets, but the maximum temporal intervals for these effects ranged from 80 ms (Pisoni, 1972) to 250 ms (Massaro, 1972). Moreover, these studies found backward masking effects on vowel perception, but not forward masking, and there were large individual differences, all of which suggests that these effects were caused by higher-level interference effects on decision processes rather than low-level auditory integration effects (LaRiviere *et al.*, 1975; Dorman *et al.*, 1977).

Another approach would be to test the accuracy and latency with which subjects can identify gated vowels of varying duration. The results of gating studies, however, are hard to interpret in the context of basic auditory processes. Because duration is itself a secondary cue to vowel quality (cf. Joos, 1948; Peterson and Lehiste, 1960), gating results would be influenced by higher-level processes, which map auditory information onto linguistic representations in order for the subject to ultimately make a phonetic category decision. As a consequence, higher-level linguistic factors could influence the perceptual results. Ultimately, studies that rely solely on the manipulation of speech stimuli will always be influenced by the effects of such manipulations on higher-level processing, making it difficult to isolate low-level auditory processing mechanisms.

An alternative approach uses the effects of nonspeech sounds on speech perception. Lotto and Kluender (1998) pioneered these methods, showing that preceding tones shifted the boundary between [da] and [ba] in a continuum of stimuli. Because the nonspeech tones do not activate higher-level representations strongly enough to create a percept of speech, the influence of such stimuli on perception is more likely to reflect early auditory stages of processing. On the other hand, because the nonspeech effect is measured in relation to subjects' responses to a speech target, it taps those auditory channels which contribute to phonetic perception. Thus, examining the effects of nonspeech sounds on speech perception provides a means of investigating the contribution of low-level auditory processing to the perception of speech.

The current study uses a paradigm similar to that employed by Lotto and Kluender (1998) and Holt *et al.* (2000), exploring the effects of nonspeech tones on vowel identification (Wallace and Blumstein, 2006). Stimuli consist of the vowel targets [i] and [ɑ], preceded by nonspeech primes composed of two sinusoidal components matched to the first two formant frequencies of one or the other of the targets. Initial experiments (Wallace and Blumstein, 2006) showed that subjects respond faster to the vowels in matched prime-target pairs than in mismatched pairs. This priming effect differs from Lotto and Kluender's (1998) method in one important way. In Lotto and Kluender's (1998) study, the effect of the nonspeech tone stimuli on the speech stimuli was found primarily in the boundary region of a continuum of stimuli, where tokens were identified at or near chance for either phonetic category. The current study, in contrast, measures behavioral effects of nonspeech stimuli on good exemplars of phonetic categories. In this way, it is possible to investigate vowel identification independent of the processes involved in resolving phonetic category membership in ambiguous stimuli (though such processes are important in the perception of underspecified, hypoarticulated natural speech).

In the current study, the duration of nonspeech prime tones will be varied parametrically in order to test the size of the temporal window that contributes to vowel quality perception. Prime durations will be 25, 50, 100, 150, 200, 300, and 500 ms, a set that spans the proposed short and long window durations. Because priming effects depend on similarity between prime and target, response times (RTs) are expected to be faster for "match" than for "clash" trials at each prime duration. Prime duration is likely to affect the magnitude of this priming effect, since primes that are shorter than the temporal integration window produce a weaker auditory response than primes with durations equal to or greater than the temporal integration window. This difference in activation is likely to modulate the effect of similarity between primes and targets. Thus, whatever the duration of the window is, increases in prime duration up to this value should increase the magnitude of the priming effect, while increasing the prime duration beyond the length of the window should produce no further increases in priming.

If, as Poeppel (2003) suggested, vowels are analyzed over a long temporal window of 150–250 ms, the shortest primes of the current experiment will be too brief to fully excite that window. These short primes should, therefore, produce less activation and a weaker priming effect than longer primes. Maximum priming should therefore occur at prime durations of 150–250 ms, at which point the prime stimuli will be as long as the temporal integration window. If, on the other hand, vowel perception draws on a short window, then short duration primes should fully excite the auditory representation of the vowel. Maximum priming should then occur at prime durations of 25–50 ms. Long and short temporal integration windows thus make different predictions for the direction of the effect of prime duration on priming magnitude between 20 and 250 ms (i.e., in the 25, 50, 100, 150, and 200 ms conditions of the current experiment). A long window predicts increasing priming over this range, while a short window predicts level or decreasing priming.

II. EXPERIMENT 1

A. Methods

1. Subjects

Subjects were 70 native speakers of North American English who reported having normal hearing and no history of speech or language disorders. They were paid for their participation.

2. Stimuli

Targets were two vowel sounds, [ɑ] and [i], synthesized using the PRAAT software package (Boersma and Weenink, 2007). The linear predictive coding residual of a single token of [i] spoken in isolation by a male native speaker of North American English was passed through a five-formant cascade to create each target vowel. F_0 ranged from 123 to 131 Hz and the stimulus duration was 280 ms. The first three formants were set to constant frequencies unique to each target vowel; F1 and F2 frequencies were determined by measuring the speaker's productions, while F3 was set to a value that produced the best vowel percept as judged by the experimenter (A. Wallace) ([i]: 362, 2191, and 3100 Hz; [ɑ]: 772, 1308, and 2960 Hz). The fourth and fifth formants, as well as all formants' bandwidths, were not altered from the time-varying values measured in the original natural [i] (mean F4=3487 Hz, F5=3997 Hz; BW=82, 39, 261, 204, 238 Hz). The targets' amplitudes were scaled so that they would be equal in perceptual loudness, as predicted by the ISO532B standard [71 dB sound pressure level (SPL) for both targets].

Primes were two nonspeech tone complexes, one matched to each of the targets. Each prime contained two component tones at the frequencies of the first two formants of the matched target. Intensities of the individual components were controlled to ensure: first, that the two sinusoidal components of each prime were equally loud, based on ISO226 equal loudness contours; and second, that the two primes themselves were equally loud, as predicted by ISO532B, resulting in absolute levels of 79 dB SPL for primes matched to [i] and 83 dB SPL for primes matched to [ɑ]. Primes were 25, 50, 100, 150, 200, 300, and 500 ms in duration, including 15 ms cosine-squared onset and offset intensity ramps, except in the 25-ms condition, where 12.5 ms ramps were used. After completing the experiment, subjects were asked what the primes sounded like, and none reported having heard them as speech. Subjects were then asked specifically whether the primes sounded like the target vowels; 3 of 70 subjects reported that they did.

3. Procedure

Subjects were assigned to one of the seven prime duration conditions with ten subjects participating in each condition. Each trial consisted of a prime and a target, separated by a 50 ms inter-stimulus interval. RTs were measured from the onset of the target vowel. Each trial began 750 ms after the onset of the target of the previous trial, unless a response had not yet been recorded, in which case subjects were given up to 3 s to respond. In all, the two targets and two primes combined to form four unique trials (two match conditions and two clash conditions, see Fig. 1). Each of these four was repeated 250 times in pseudorandom order for each subject, while prime duration varied between subjects. Stimuli were presented binaurally over Sony MDR-V6 headphones and subjects responded by pressing one of two buttons labeled with the target vowels in both the international phonetic alphabet and a colloquial phonetic notation ([ɑ]/"ah" and [i]/"ee"). Both performance and reaction-time measures were taken.

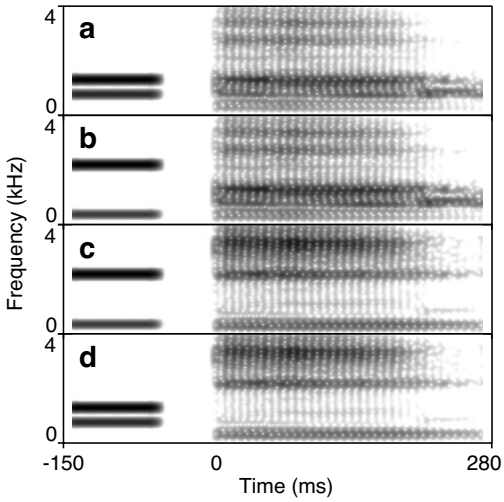


FIG. 1. Spectrograms illustrating four trial types in 100 ms prime condition: (a) [a] target, match prime; (b) [a] target, clash prime; (c) [i] target, match prime; and (d) [i] target, clash prime.

B. Results

Mean RTs were calculated for each subject for all trials of a given type, excluding incorrect responses, responses that were made before the target onset, and trials in which the subject did not respond in the given 3000 ms (1.23% of trials). As shown in Table I, subjects' responses to match trials were both faster and more accurate than their responses to clash trials. Because of the high variability in both error rates (which were low overall) and absolute RTs (which varied greatly between subjects), statistical analyses relied on a

TABLE I. Cross-subject mean error rates and response times for experiment 1.

Prime duration (ms)	Error rate (per 1000 trials)			Response time (ms)		
	Match	Clash	Difference	Match	Clash	Difference
25	8.8	19	10	434	468	34
50	9.8	21	11	468	494	26
100	7.6	16	8.0	451	481	29
150	21	30	9.0	416	441	25
200	5.2	8.4	3.2	443	460	17
300	4.6	9.6	5.0	412	424	11
500	2.8	5.0	2.2	465	471	6

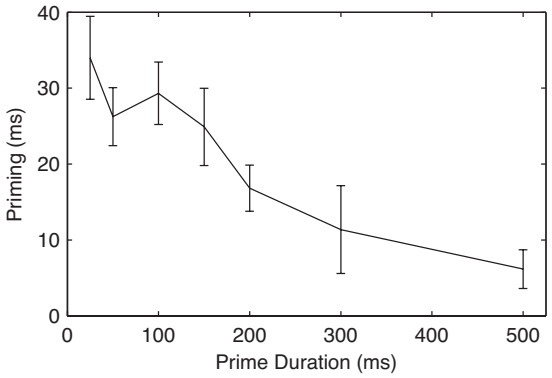


FIG. 2. Results of experiment 1. The magnitude of priming—RT for clash trials minus RT for match trials—is shown as a function of prime duration. Error bars indicate 1 standard error for the cross-subject mean.

single RT priming score, computed independently for each subject as the difference between mean RTs for match and clash trials. A one-sample t-test, with results pooled across all conditions, confirmed that priming scores were significantly greater than zero [$t_{(69)}=11.3$, $p<0.001$]. These results confirm the existence of the priming effect of nonspeech tones in a vowel-identification task.

As Fig. 2 shows, the magnitude of this priming effect varied across prime durations, with a greater magnitude of priming at shorter prime durations. A one-way analysis of variance (ANOVA) found a significant effect of prime duration on the magnitude of the priming effect [$F_{(6,63)}=5.81$, $p<0.001$]. Polynomial trend analysis revealed a significant linear trend [$F_{(1,63)}=31.2$, $p<0.001$] but no significant higher-order trends [quadratic $F_{(6,63)}=1.31$, $p=0.25$; cubic $F_{(6,63)}=0.41$, $p=0.52$; fourth-order $F_{(6,63)}<0.001$, $p=1.0$; fifth-order $F_{(6,63)}=1.74$, $p=0.19$]. To further explore the nature of this trend, contrasts were tested between short (25 and 50 ms), medium (100 and 150 ms), and long (200 and 300 ms) durations, excluding the 500 ms condition. Short and medium prime durations did not differ significantly from each other [$F_{(1,63)}=0.52$, $p=0.47$], but long prime durations produced significantly less priming than either short [$F=15$, $p<0.001$] or medium ($F=9.7$, $p=0.003$) durations. As discussed in the Introduction, a long temporal integration window would predict an increase in priming up to durations equal to that of the window, around 150–250 ms. The fact that such an increase did not occur and that maximal priming occurs at the shorter durations supports a short temporal integration window for vowel quality perception.

C. Discussion

The results of experiment 1 indicate that priming magnitudes are highest at the shortest prime durations, remain level at durations of 100–150 ms, and decline at longer durations. The fact that the short duration primes did not produce a smaller priming effect than the medium duration primes is consistent with a short temporal integration window for the processing of vowels, but not a long one. As discussed in the Introduction, medium to long primes should excite a long window better than short primes. This should result in greater magnitudes of priming for either medium or long primes than for short primes, a pattern which was not found in experiment 1. Thus, the current results indicate that the perception of vowel quality draws on an auditory processing channel with a short temporal integration window of 20–50 ms.

Finally, it is of interest that priming magnitudes actually declined at the longer prime durations. Standard auditory models (cf. Viemeister and Wakefield, 1991) predict that auditory activation (and thus, presumably, the magnitude of the priming effect) would plateau at durations longer than that of the relevant temporal integration window. A few factors that are not incorporated into the basic auditory models might explain the obtained decline in the magnitude of priming. For example, many parts of the auditory system have been found to respond strongly to stimulus onsets (Furukawa and Ishii, 1967; Smith, 1977; Rhode and Smith, 1985; Pfeiffer, 1966; Rhode *et al.*, 1983). Such an onset response might produce a decline in priming at longer prime durations, since increasing duration would increase the temporal separation between prime and target onsets, allowing more time for the auditory response to decay. In addition, it is of interest that no significant difference was found between short (25 and 50 ms) and medium (100 and 150 ms) prime durations, suggesting a delay before priming magnitudes begin to decline. Further research is needed to explore this phenomenon.

III. EXPERIMENT 2

In experiment 1, it was shown that the priming effect produced by nonspeech tone complexes matched to a target vowel's formant frequencies is greatest at the shortest prime durations, suggesting that the spectral properties of vowels are extracted over a relatively short temporal window. To determine whether the effect of prime duration would be the same when primes engage higher levels of phonetic processing along with auditory levels, experiment 2 replicated the duration conditions of experiment 1 using matched or mismatched vowel primes in place of nonspeech tones.

A. Methods

1. Subjects

Subjects were 70 native speakers of North American English, who met the same criteria as the subjects who participated in experiment 1 but who had not taken part in experiment 1.

2. Stimuli

Targets were the same tokens of [i] and [a] used in experiment 1. Primes were vowel stimuli, created by altering the duration of the vowel targets to 25, 50, 100, 150, 200, 300, and 500 ms using the PRAAT software package's (Boersma and Weenink, 2007) implementation of Moulines and Laroche's (1995) pitch-synchronous overlap-and-add algorithm.

3. Procedure

All aspects of the experimental procedure, with the exception of the different primes, were the same as in experi-

TABLE II. Cross-subject mean error rates and response times for experiment 1.

Prime duration (ms)	Error rate (per 1000 trials)			Response time (ms)		
	Match	Clash	Difference	Match	Clash	Difference
25	2.2	11	9.2	446	503	57
50	5.4	21	16	456	536	80
100	9.0	22	13	504	598	94
150	3.4	18	14	480	582	102
200	4.2	11	6.6	476	550	74
300	6.2	13	6.6	417	461	44
500	7.8	11	3.2	507	534	27

ment 1. Subjects were told that vowels would come in pairs, and that they should ignore the first vowel in each pair and only respond to the second by pressing one of two buttons labeled with the target vowels. Both performance and reaction-time measures were taken. Prime duration varied between subjects, with ten subjects participating in each of the seven duration conditions. Subjects completed 250 match and 250 clash trials for each of the two target vowels.

B. Results

Table II lists mean error rates and RT, for match and clash trials at each prime duration of experiment 2. As in experiment 1, mean RTs were calculated for each subject for all trials of a given type, excluding trials for the same reasons as in experiment 1 (1.05% of trials). The subject's mean reaction time for match trials was then subtracted from the mean for clash trials to produce a single difference value, the magnitude of priming for that subject. Figure 3 shows the magnitude of priming as a function of prime duration for the vowel prime stimuli of experiment 2 as well as the nonspeech primes of experiment 1. A single-factor ANOVA revealed a significant effect of prime duration on priming magnitude [$F_{(6,63)}=3.45$, $p=0.005$]. Polynomial trend analysis of the speech-prime conditions found significant or nearly-significant linear [$F_{(1,63)}=10.0$, $p=0.002$], quadratic [$F_{(1,63)}=2.81$, $p=0.098$], and cubic trends [$F_{(1,63)}=7.02$, $p=0.010$], without significant fourth-order [$F_{(1,63)}=0.312$, $p=0.579$] or fifth-order [$F_{(1,63)}=0.13$, $p=0.133$] trends. As in experiment 1, the nature of this trend was further explored by testing contrasts between short (25 and 50 ms), medium (100 and 150 ms), and long (200 and 300 ms) prime durations. In experiment 1, short primes produced as large an effect as medium, and a greater effect than long primes. In contrast, the medium-duration primes of experiment 2 produced a larger effect than short [$F_{(1,63)}=4.2$, $p=0.045$] or long [$F_{(1,63)}=7.2$, $p=0.009$] primes, while short and long primes did not differ significantly [$F_{(1,63)}=0.41$, $p=0.52$].

Finally, the results of experiments 1 and 2 were directly compared in a two-factor (prime duration × prime type) ANOVA, using the results of experiment 1 as a "nonspeech" prime type condition and the results of experiment 2 as a speech prime type condition. Results indicated significant main effects of prime duration [$F_{(6,126)}=5.18$, $p<0.001$] and prime type [speech vs nonspeech; $F_{(1,126)}=67.4$, $p<0.001$], and an interaction that approached significance [$F_{(6,126)}=2.07$, $p=0.061$]. This interaction was consistent with the different patterns observed in experiments 1 and 2.

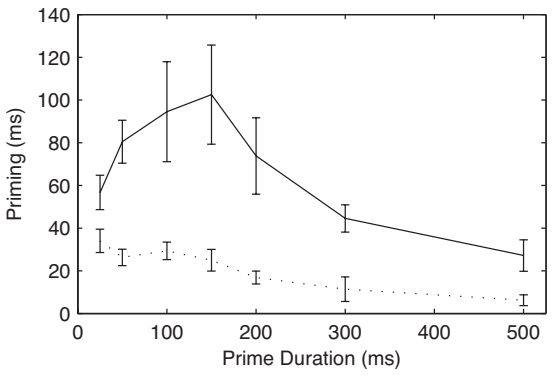


FIG. 3. Results of experiment 2. Solid line shows the magnitude of priming by prime duration for vowel primes and vowel targets. Dotted line shows magnitude of priming by duration for nonspeech primes and vowel targets (data from experiment 1). Error bars indicate 1 standard error for cross-subject mean.

C. Discussion

Experiment 2 found a significantly greater priming effect for speech primes than for nonspeech. This is neither surprising nor particularly interesting given that the primes of experiment 2 were acoustically closer to the targets than were the nonspeech primes of experiment 1, and matched the targets phonetically as well. Of more interest is the effect of prime duration on the speech and nonspeech prime conditions. While experiment 1 found the greatest priming at the shortest prime durations, experiment 2 found priming to be greatest with the medium-duration primes of 100 and 150 ms. These two patterns likely reflect a difference in the processing mechanisms for the two kinds of stimuli. Experiment 1, using nonspeech primes, taps primarily auditory levels of processing, whereas experiment 2, using vowel primes, taps both auditory and phonetic levels. Thus, the data from experiment 2 suggest that phonetic processes operate over a timescale of approximately 150 ms, consistent with the interpretation of backward masking results by Dorman *et al.* (1977) (Massaro, 1972; Pisoni, 1972) in the perception of vowels as reflecting higher-level processing occurring over 100–200 ms.

IV. GENERAL DISCUSSION

The current experiments explored the temporal processing of vowels by manipulating the duration of both nonspeech (experiment 1) and speech (experiment 2) primes, which preceded vowel targets in a vowel-identification task. In both experiments, matched primes facilitated the identification of acoustically similar targets. In experiment 1, the magnitude of priming showed a linear decline across the duration intervals; priming magnitudes were highest at the shortest prime durations, remained level at durations of 100–150 ms, and declined at longer durations. In experiment 2, priming magnitudes showed an initial increase with the greatest priming at medium durations of 100–150 ms followed by a subsequent decline. These results support a model in which vowels undergo an initial auditory analysis within a short temporal window. The output of this analysis is then mapped onto phonetic perceptual mechanisms where longer-scale temporal dynamics play a role.

At the auditory level, the current results are compatible with a range of window durations, including Poeppel's (2003) proposal of 20–40 ms, but they are not compatible with the exclusive involvement of a medium or long window of analysis for vowel quality perception. Instead, the current studies suggest that vowel quality perception draws on the output of a high-temporal-resolution, low-spectral-resolution auditory processing channel. It is possible that the temporal window for the auditory analysis of vowels may be shorter than 25 ms and possibly shorter than 10 ms as proposed by Viemeister (1979); however, given that primes of this duration were not used this hypothesis would need to be tested directly.

Determining the temporal and spectral parameters of this auditory channel will help explain the vowel perception process as a whole. For example, models of vowel perception have tended to avoid direct claims about auditory processing. Syrdal and Gopal (1986), for example, described a model of vowel identification from formant frequencies but did not specify how these frequencies are obtained [they used Peterson and Barney's (1952) measurements]. Molis (2005) contrasted a number of models of vowel perception. Although some of these incorporate knowledge of peripheral auditory processing (i.e., critical-band filtering), none includes the possibility that central auditory processing might introduce different channels with different spectral and temporal resolutions.

The current study has proceeded from the assumption of multiple temporal windows, but there are other potential explanations for the differences that have been found in the temporal resolution of different psychoacoustic tasks. One of these Viemeister and Wakefield's (1991) "multiple looks" model (cf. also Moore, 2003, 2008) proposes that incoming sounds are analyzed within a single, short temporal window, but that specific sounds are identified on the basis of templates, which can span several windows (i.e., multiple looks). Hillenbrand and Houde's (2003) model of vowel perception takes a similar approach in the frequency domain, arguing that spectral smoothing to obscure a vowel's harmonics is not necessary, because a high-spectral-resolution representation of the vowel can instead be compared directly to a smooth template. Experimentally distinguishing between a "multiple windows" model and a "multiple looks" model is rather difficult, and the results of the current study are compatible with both models. Within a multiple looks framework, the current study suggests that steady-state vowels are identified on the basis of short, low-spectral-resolution templates.

Much remains to be learned about the temporal and spectral parameters of the auditory processing of vowels, as well as other speech sounds. Future studies can explore these areas using the basic methodological framework presented in the current paper. Perceptual interactions between speech and nonspeech sounds, because they are likely to occur at relatively early *auditory* stages *of the speech-processing stream*, provide a unique window into this level of processing. By illuminating such mechanisms, this line of research will also shed light on the building blocks of higher-level processes.

ACKNOWLEDGMENTS

This work was supported by NIH Grant Nos. DC000314 and DC006220 to Brown University.

Assmann, P., and Katz, W. (**2000**). "Time-varying spectral change in the vowels of children and adults," J. Acoust. Soc. Am. **108**, 1856–1866.

Belin, P., Zilbovicius, M., Crozier, S., Thivard, L., Fontaine, A., Masure, M.-C., and Samson, Y. (**1998**). "Lateralization of speech and auditory temporal processing," J. Cogn Neurosci. **10**, 536–540.

Blumstein, S. E. (**1995**). "The neurobiology of the sound structure of language," in *The Cognitive Neurosciences*, edited by M. Gazzaniga (MIT, Cambridge, MA), 915–929.

Boemio, A., Fromm, S., Braun, A., and Poeppel, D. (**2005**). "Hierarchical and asymmetric temporal sensitivity in human auditory cortices," Nat. Neurosci. **8**, 389–396.

Boersma, P., and Weenink, D. (**2007**). "Praat: Doing phonetics by computer, version 4.5.16," computer program, http://www.praat.org/ (last viewed February, 2007).

Britton, B., Blumstein, S. E., Myers, E. B., and Grindrod, C. (**2009**). "The role of spectral and durational properties on hemispheric asymmetries in vowel perception," Neuropsychologia (in press).

Carlson, R., Granström, B., and Fant, G. (**1970**). "Some studies concerning perception of isolated vowels," Speech Transmission Laboratory Quarterly Progress Report No. 2-3, KTH, Stockholm.

Cheveigné, A., and Kawahara, H. (**1999**). "Missing-data model of vowel identification," J. Acoust. Soc. Am. **105**, 3497–3508.

Chistovich, L. (**1985**). "Central auditory processing of peripheral vowel spectra," J. Acoust. Soc. Am. **77**, 789–805.

Chistovich, L., and Lublinskaya, V. (**1979**). "The 'center of gravity' effect in vowel spectra and critical distance between the formants: Psychoacoustical study of the perception of vowel-like stimuli," Hear. Res. **1**, 185–195.

Chistovich, L. A., and Sheikin, R. L. (**1979**). "Relevant parameters of steady-state vowel spectra," in *Sensornye Systemy*, edited by G. V. Gershuni (Nauka, Leningrad), pp. 116–129 (in Russian).

Delattre, P. C., Liberman, A. M., Cooper, F. S., and Gerstman, L. J. (**1952**). "An experimental study of the acoustic determinants of vowel colour; observations on one- and two-formant vowels synthesized from spectrographic patterns," Word **8**, 195–210.

Diehl, R., Lindblom, B., Hoemeke, K., and Fahey, R. (**1996**). "On explaining certain male-female differences in the phonetic realization of vowel categories," J. Phonetics **24**, 187–208.

Dorman, M., Kewley-Port, D., Brady, S., and Turvey, M. (**1977**). "Vowel recognition: Inferences from studies of forward and backward masking," Q. J. Exp. Psychol. **29**, 483–497.

Fant, G. (**1956**). "On the predictability of formant levels and spectrum envelopes from formant frequencies," in *For Roman Jakobson*, edited by M. Halle, H. Lunt, and H. Maclean (Mouton, The Hague), pp. 109–120.

Fitzgibbons, P. (**1983**). "Temporal gap detection in noise as a function of frequency, bandwidth, and level," J. Acoust. Soc. Am. **74**, 67–72.

Furukawa, T., and Ishii, Y. (**1967**). "Neurophysiological studies of hearing in goldfish," J. Neurophysiol. **30**, 1377–1403.
Giraud, A.-L., Kleinschmidt, A., Poeppel, D., Lund, T. E., Frackowiak, R. S. J., and Laufs, H. (**2007**). "Endogenous cortical rhythms determine cerebral specialization for speech perception and production," Neuron **56**, 1127–1134.
Giraud, A.-L., Lorenzi, C., Ashburner, J., Wable, J., Johnsrude, I., Frackowiak, R., and Kleinschmidt, A. (**2000**). "Representation of the temporal envelope of sounds in the human brain," J. Neurophysiol. **84**, 1588–1598.
Hesling, I., Dilharreguy, B., Clement, S., Bordessoules, M., and Allard, M. (**2005**). "Cerebral mechanisms of prosodic sensory integration using low-frequency bands of connected speech," Hum. Brain Mapp **26**, 157–169.
Hillenbrand, J., and Houde, R. (**2003**). "A narrow band pattern-matching model of vowel perception," J. Acoust. Soc. Am. **113**, 1044–1065.
Hirsh, I. J. (**1959**). "Auditory perception of temporal order," J. Acoust. Soc. Am. **31**, 759–767.
Holt, L. L., Lotto, A. J., and Kluender, K. R. (**2000**). "Neighboring spectral content influences vowel identification," J. Acoust. Soc. Am. **108**, 710–722.
Jamison, H., Watkins, K., Bishop, D., and Matthews, P. (**2006**). "Hemispheric specialization for processing auditory nonspeech stimuli," Cereb. Cortex **16**, 1266–1275.
Johnsrude, I., Zatorre, R., Milner, B., and Evans, A. (**1997**). "Left-hemisphere specialization for the processing of acoustic transients," NeuroReport **8**, 1761–1765.
Joos, M. (**1948**). "Acoustic phonetics," Language **24**, 5–136.
LaRiviere, C., Winitz, H., and Herriman, E. (**1975**). "Vocalic transitions in the perception of voiceless initial stops," J. Acoust. Soc. Am. **57**, 470–475.
Lifshitz, S. (**1933**). "Two integral laws of sound perception relating loudness and apparent duration of sound impulses," J. Acoust. Soc. Am. **5**, 31–33.
Lotto, A. J., and Kluender, K. R. (**1998**). "General contrast effects in speech perception: Effect of preceding liquid on stop consonant identification," Percept. Psychophys. **60**, 602–619.
Luo, H., and Poeppel, D. (**2007**). "Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex," Neuron **54**, 1001–1010.
Luo, H., Husain, F. T., Horwitz, B., and Poeppel, D. (**2005**). "Discrimination and categorization of speech and non-speech sounds in an MEG delayed-match-to-sample study," Neuroimage **28**, 59–71.
Massaro, D. (**1972**). "Preperceptual images, processing time, and perceptual units in auditory perception," Psychol. Rev. **79**, 124–145.
Mann, V. A. (**1980**). "Influence of preceding liquid on stop-consonant perception," Percept. Psychophys. **28**, 407–412.
McClelland, J. L., and Elman, J. L. (**1986**). "The trace model of speech perception," Cogn. Psychol. **18**, 1–86.
Molis, M. R. (**2005**). "Evaluating models of vowel perception," J. Acoust. Soc. Am. **118**, 1062–1071.
Moore, B. (**2003**). "Temporal integration and context effects in hearing," J. Phonetics **31**, 563–574.
Moore, B. C. J. (**2008**). "Basic auditory processes involved in the analysis of speech sounds," Philos. Trans. R. Soc. London, Ser. B **363**, 947–963.
Moore, B., Glasberg, B., Plack, C., and Biswas, A. (**1988**). "The shape of the ear's temporal window," J. Acoust. Soc. Am. **83**, 1102–1116.
Moulines, E., and Laroche, J. (**1995**). "Non-parametric techniques for pitch-scale and time-scale modification of speech," Speech Commun. **16**, 175–205.
Munson, W. (**1947**). "The growth of auditory sensation," J. Acoust. Soc. Am. **19**, 734–735.
Nearey, T., and Assmann, P. (**1986**). "Modeling the role of inherent spectral change in vowel identification," J. Acoust. Soc. Am. **80**, 1297–1308.
Obleser, J., Zimmermann, J., Van Meter, J., and Rauschecker, J. (**2007**). "Multiple stages of auditory speech perception reflected in event-related fMRI," Cereb. Cortex **17**, 2251–2257.
Penner, M. (**1977**). "Detection of temporal gaps in noise as a measure of the decay of auditory sensation," J. Acoust. Soc. Am. **61**, 552–557.
Penner, M. (**1978**). "A power law transformation resulting in a class of short-term integrators that produce time-intensity trades for noise bursts," J. Acoust. Soc. Am. **63**, 195–202.
Penner, M., and Cudahy, E. (**1973**). "Critical masking interval: A temporal analog of the critical band," J. Acoust. Soc. Am. **54**, 1530–1534.
Peterson, G. E., and Barney, H. L. (**1952**). "Control methods used in a study of vowels," J. Acoust. Soc. Am. **24**, 175–184.
Peterson, G., and Lehiste, I. (**1960**). "Duration of syllable nuclei in English," J. Acoust. Soc. Am. **32**, 693–703.
Pfeiffer, R. R. (**1966**). "Classification of response patterns of spike discharges for units in the cochlear nucleus: Tone-burst stimulation," Exp. Brain Res. **1**, 220–235.
Pisoni, D. B. (**1972**). "Perceptual processing time for consonants and vowels," Haskins Laboratories Status Report on Speech Research No. SR-31/32, Haskins Laboratories, New Haven, CT.
Pisoni, D. (**1977**). "Identification and discrimination of the relative onset time of two component tones: Implications for voicing perception in stops," J. Acoust. Soc. Am. **61**, 1352–1361.
Plomp, R. (**1964**). "Rate of decay of auditory sensation," J. Acoust. Soc. Am. **36**, 277–282.
Plomp, R., and Bouman, M. (**1959**). "Relation between hearing threshold and duration for tone pulses," J. Acoust. Soc. Am. **31**, 749–758.
Poeppel, D. (**2003**). "The analysis of speech in different temporal integration windows: Cerebral lateralization as 'asymmetric sampling in time'," Speech Commun. **41**, 245–255.
Poeppel, D., Guillemin, A., Thompson, J., Fritz, J., Bavelier, D., and Braun, A. R. (**2004**). "Auditory lexical decision, categorical perception, and FM direction discrimination differentially engage left and right auditory cortex," Neuropsychologia **42**, 183–200.
Rhode, W., and Smith, P. (**1985**). "Characteristics of tone-pip response patterns in relationship to spontaneous rate in cat auditory nerve fibers," Hear. Res. **18**, 159–168.
Rhode, W. S., Smith, P. H., and Oertel, D. (**1983**). "Physiological response properties of cells labelled intracellularly with horseradish peroxidase in cat dorsal cochlear nucleus," J. Comp. Neurol. **213**, 426–447.
Ryalls, J. H., and Lieberman, P. (**1982**). "Fundamental frequency and vowel perception," J. Acoust. Soc. Am. **72**, 1631–1634.
Samson, S., and Zatorre, R. (**1994**). "Contribution of the right temporal lobe to musical timbre discrimination," Neuropsychologia **32**, 231–240.
Shailer, M., and Moore, B. (**1983**). "Gap detection as a function of frequency, bandwidth, and level," J. Acoust. Soc. Am. **74**, 467–473.
Shailer, M., and Moore, B. (**1987**). "Gap detection and the auditory filter: Phase effects using sinusoidal stimuli," J. Acoust. Soc. Am. **81**, 1110–1117.
Smith, R. (**1977**). "Short-term adaptation in single auditory nerve fibers: Some poststimulatory effects," J. Neurophysiol. **40**, 1098–1111.
Stevens, K. (**1972**). "The quantal nature of speech: Evidence from articulatory-acoustic data," in *Human Communication, A Unified View*, edited by E. E. David, and P. B. Denes (McGraw-Hill, New York), pp. 51–66.
Stevens, K. (**1989**). "On the quantal nature of speech," J. Phonetics **17**, 3–45.
Strange, W. (**1989**). "Dynamic specification of coarticulated vowels spoken in sentence context," J. Acoust. Soc. Am. **85**, 2135–2153.
Syrdal, A., and Gopal, H. (**1986**). "A perceptual model of vowel recognition based on the auditory representation of American English vowels," J. Acoust. Soc. Am. **79**, 1086–1100.
Versnel, H., and Shamma, S. A. (**1998**). "Spectral-ripple representation of steady-state vowels in primary auditory cortex," J. Acoust. Soc. Am. **103**, 2502–2514.
Viemeister, N. (**1979**). "Temporal modulation transfer functions based upon modulation thresholds," J. Acoust. Soc. Am. **66**, 1364–1380.
Viemeister, N., and Wakefield, G. (**1991**). "Temporal integration and multiple looks," J. Acoust. Soc. Am. **90**, 858–865.
Wallace, A. B., and Blumstein, S. E. (**2006**). "Nonspeech sounds prime acoustically similar words," J. Acoust. Soc. Am. **119**, 3245.
Wartenburger, I., Steinbrink, J., Telkemeyer, S., Friedrich, M., Friederici, A., and Obrig, H. (**2007**). "The processing of prosody: Evidence of interhemispheric specialization at the age of four," Neuroimage **34**, 416–425.
Zaehle, T., Wüstenberg, T., Meyer, M., and Jäncke, L. (**2004**). "Evidence for rapid auditory perception as the foundation of speech processing: A sparse temporal sampling fmri study," Eur. J. Neurosci. **20**, 2447–2456.
Zatorre, R. J., and Belin, P. (**2001**). "Spectral and temporal processing in human auditory cortex," Cereb. Cortex **11**, 946–953.
Zatorre, R., Belin, P., and Penhune, V. (**2002**). "Structure and function of auditory cortex: Music and speech," Trends Cogn. Sci. **6**, 37–46.
Zwicker, E. (**1961**). "Subdivision of the audible frequency range into critical bands (frequenzgruppen)," J. Acoust. Soc. Am. **33**, 248.
Zwicker, E., and Terhardt, E. (**1980**). "Analytical expressions for critical-band rate and critical bandwidth as a function of frequency," J. Acoust. Soc. Am. **68**, 1523–1525.
Zwislocki, J. (**1960**). "Theory of temporal auditory summation," J. Acoust. Soc. Am. **32**, 1046–1060.

Multisyllabic nonwords: More than a string of syllables

Lisa M. D. Archibald[a)]
School of Communication Sciences and Disorders, University of Western Ontario, London, Ontario N6G 1H1, Canada

Susan E. Gathercole[b)]
Department of Psychology, University of York, York YO10 5DD, United Kingdom

Marc F. Joanisse[c)]
Department of Psychology, University of Western Ontario, London, Ontario N6A 5C2, Canada

(Received 18 December 2007; revised 7 January 2009; accepted 8 January 2009)

Nonword repetition is closely associated with the learning of the phonological form of novel words. Several factors influence nonword repetition performance such as short-term memory, phonotactic probability, lexical knowledge, and prosodic factors. The present study examined the influence of list duration, coarticulation, and prosody on nonword repetition by comparing naturally articulated multisyllabic nonwords to multisyllabic nonwords formed by concatenating syllables produced in isolation and serial lists (experiment 1), to multisyllabic forms that incorporated either valid or invalid coarticulatory information (experiment 2), and to multisyllabic forms either with or without common English within-word stress patterns (experiment 3). Results revealed superior recall for naturally articulated nonwords compared to lists of matched duration or sequences with invalid coarticulatory cues. Within-word stress patterns also conveyed a repetition advantage. The findings clearly establish that the coarticulatory and prosodic cues of naturally articulated multisyllabic forms support retention. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3076200]

PACS number(s): 43.71.Es, 43.71.Sy [RSN] Pages: 1712–1722

I. MULTISYLLABIC NONWORDS: MORE THAN A STRING OF SYLLABLES

The ability to repeat multisyllabic nonwords (nonword repetition) has been linked to new word learning in both children (Gathercole *et al.*, 1992) and adults (Atkins and Baddeley, 1998; Gupta, 2003), and in both native (Gathercole and Baddeley, 1989) and foreign language learning (Masoura and Gathercole, 1999, 2005; Service, 1992). Such findings have led to widespread interest in understanding the cognitive processes that underlie nonword repetition. Nonword repetition was first proposed as a relatively pure index of phonological short-term memory (STM) capacity (Gathercole and Baddeley, 1989, 1993). Reports that nonword repetition accuracy decreases with increasing nonword length (Gathercole *et al.*, 1994) and forms a classic serial recall curve (Archibald and Gathercole, 2007a; Gupta, 2005) are consistent with this position. Not all of the available evidence can be explained by STM, however.

Several studies have demonstrated that characteristics of the nonwords themselves influence performance, such as the extent to which the nonwords resemble real words (Gathercole, 1995), the prosodic structure of the nonword (Roy and Chiat, 2004), the phonotactic probability of the inherent phonemes (Edwards *et al.*, 2004), and the articulatory complexity of the nonword (Archibald and Gathercole, 2006). One area that has not been explored as closely concerns the extent to which the coarticulatory information inherent in multisyllabic nonwords influences repetition accuracy. The studies reported in the present paper investigated three factors that may influence nonword repetition, nonword duration, coarticulatory cues, and within-word prosodic patterns.

The role of STM in nonword repetition has been investigated by comparing performance on this task to a classic immediate serial recall paradigm. In a recent study, typically developing children aged 6–12 years repeated nonword syllable lists presented either as whole multisyllabic forms or as lists of individual syllables for serial recall (Archibald and Gathercole, 2007a). Consistent with previous findings (Gupta, 2005), a serial recall curve was found for both conditions. Interestingly, the results clearly established more accurate repetition of the naturally articulated multisyllabic nonwords than lists of (equivalent) syllables presented individually, at least for this age range.

Why should multisyllabic nonwords be easier to recall than lists of individual syllables? One possibility lies in the temporal difference between the conditions. In the study of Archibald and Gathercole (2007a) overall stimulus duration was greater in the serial recall paradigm due to the pause time between list items. Attempting to retain information in STM for longer periods may lead to greater loss for two reasons. One explanation relates to the suggestion that the contents of STM spontaneously decay over time in order to allow for continuous updating and removal of information no longer needed (e.g., Baddeley, 1986; Cowan, 1999). According to this view, lists of longer duration will be subject to greater loss over time resulting in greater forgetting in the longer serial recall lists than the shorter nonword repetition

a)Electronic mail: larchiba@uwo.ca
b)Electronic mail: sg539@york.ac.uk
c)Electronic mail: marcj@uwo.ca

condition. A second view holds that when several elements are held in STM simultaneously, interference arises such that representations may become degraded (e.g., Nairne, 2002). It follows from this that opportunities for interference will be greater when information is retained for longer periods.

An additional factor is the effect of coarticulation, the overlap in the articulation of sound gestures for consecutive segments of an utterance (MacNeilage, 1980). Coarticulatory information is known to influence speech perception. For example, listeners can predict the following vowel when presented with only the aperiodic portion of the consonants that preceded it (e.g., Nittrouer and Whalen, 1989; Tjaden and Sussman, 2006); listeners are faster and more accurate at identifying a vowel when the preceding sounds contain appropriate coarticulatory information (e.g., Martin and Bunnell, 1981, 1982); and listeners have more difficulty identifying vowels when coarticulatory information is distorted such as in speakers with apraxia of speech (Southwood *et al.*, 1997). It therefore appears that coarticulatory cues enhance listeners' ability to perceive the identity of later occurring segments in a way that maximizes the efficiency and speed of speech perception (Ostreicher and Sharf, 1976). It has been suggested that the absence of coarticulatory cues across word boundaries may account for the decrease in immediate sentence recall observed in synthetically as compared to naturally produced speech (Paris *et al.*, 2000).

Coarticulatory cues are available across syllables in multisyllabic nonword repetition but not in serial list recall paradigms. Bidirectional coarticulatory effects across syllables from one vowel (V) to the preceding vowel (i.e., anticipatory effects) and to the vowel of the next syllable (i.e., carry-over effects) have been documented physiologically, acoustically, and perceptually (Sharf and Ohde, 1981). Coarticulatory effects have been found to be influenced by the intervening consonants (Bell-Berti and Harris, 1976; Hansen, 2003), syllable stress (Cho, 2004), and the language spoken (Beddor *et al.*, 2004). For example, languages with a dense vowel space such as American English have been found to allow more V-to-V coarticulation than languages with sparser vowel spaces (Choi and Keating, 1991). The effect of syllable structure on V-to-V coarticulation has also been studied, to some extent. Greater anticipatory than carry-over effects have been observed for open consonant (C)–vowel syllables (CV) such as those employed in the present study (Modarresi *et al.*, 2003). Other studies, however, have reported no effect of syllable structure on V-to-V coarticulation (Mok *et al.*, 2007).

There is growing evidence that the prosodic pattern of a nonword also influences immediate recall. Stress patterns have been found to affect repetition both for English (Gupta *et al.*, 2005; Reeves *et al.*, 2000; Roy and Chiat, 2004) and Japanese speakers (Yuzawa and Saito, 2006), and for typically developing and language impaired children (Chiat and Roy, 2007; Gallon *et al.*, 2007). Typically, errors are more frequent for weak than strong syllables with prestress syllables being the most vulnerable (e.g., Roy and Chiat, 2004). Our previous study of nonword repetition and serial recall (Archibald and Gathercole, 2007a) was not designed to investigate the influence of prosodic factors. Indeed, all of the syllables in the multisyllabic nonwords contained tense vowels and were produced with equal-stress–across-syllables. This design served the dual purpose of improving the acoustic saliency of all syllables, and reducing the "wordlikeness" of the nonwords thereby making them resistant to support from long-term lexical knowledge (Dollaghan and Campbell, 1998). Nevertheless, a suprasegmental contour at the level of the whole nonword would have been present in the multisyllabic forms and may have led to improved recall. The extent to which suprasegmental or within-word stress patterns may facilitate the retention of multisyllabic forms over syllable lists remains unclear.

This paper presents three experiments aimed at investigating the role of phonetic information in nonword repetition performance. In these experiments, we manipulate list duration, coarticulatory information, and prosodic cues within a standard nonword repetition task. In each experiment adults recalled nonword CV syllable strings of matched phonological content under different conditions that manipulated temporal, coarticulatory, or prosodic aspects of the signal. Recall accuracy of naturally articulated multisyllabic nonwords was compared to multisyllabic strings of equal duration without coarticulatory cues in experiment 1, multisyllabic forms with and without valid coarticulatory cues in experiment 2, and multisyllabic forms with and without within-word stress in experiment 3. More accurate recall of natural nonwords and sequences with valid coarticulatory information would point to the role of coarticulation in supporting retention. And better performance on multisyllabic forms with within-word stress patterns would reflect the positive influence of prosodic information.

II. EXPERIMENT 1

In the first experiment, participants repeated syllable strings presented as naturally articulated multisyllabic nonwords, in serial recall lists with one syllable presented at a rate of one per second, and as strings derived by concatenating individually produced CV syllables to obtain a multisyllabic nonword form of equivalent duration to the naturally articulated nonwords. One aim of experiment 1 was to establish the recall advantage of nonword repetition in an adult group. A second goal was to equate the temporal duration of naturally articulated multisyllabic nonwords and syllable strings through the use of simple concatenation. Less accurate performance on the serial recall lists but equivalent performance on the multisyllabic forms would point to the increased trial duration in serial recall as a significant detriment to performance. Superior repetition of naturally articulated than concatenated multisyllabic forms may reflect the importance of coarticulatory cues in enhancing recall.

A. Method

1. Participants

Twenty-four students participated in the present experiment (mean age=17.43 years, range=16.83–18.67). Participants were from a local sixth form in York, England and were completing their final year of high school. English was the native language of all participants.

TABLE I. Syllables used to construct stimuli for experiments 1 and 3.

Practice trials							
/faʊ/	/dʒaʊ/	/gɑ/	/vɑ/	/tʃaɪ/	/uɔɪ/	/vəʊ/	
Experimental trials							
/kaɪ/	/kəʊ/	/daʊ/	/pɔɪ/	/teɪ/	/bɑ/	/gi/	/ku/
/faɪ/	/vəʊ/	/maɪ/	/mɔɪ/	/veɪ/	/tɑ/	/tʃi/	/fu/
/yaɪ/	/tʃəʊ/	/taʊ/	/dɔɪ/	/tʃeɪ/	/dɑ/	/yi/	/vu/

2. Procedure

All participants completed a repetition task under three conditions: repeating naturally spoken nonwords (*natural nonwords*), *concatenated syllables*, and *syllable sequences*. Order of presentation of the conditions was counterbalanced, with eight participants completing each of three presentation orders. In each condition, eight experimental trials preceded by two practice trials consisting of five CV syllables were presented. The five-syllable length was chosen based on pilot testing indicating this sequence length resulted in adequate numbers of errors to yield accuracy rates below ceiling in most participants. The CV syllables comprising the lists were constructed from a pool of phonemes excluding the eight consonants that are late acquired (Shriberg and Kwiatowski, 1994). Only tense vowels were included so that the multisyllabic nonwords could be produced with roughly equal-stress-across-syllables (Dollaghan and Campbell, 1998), thereby minimizing prosodic differences among conditions. The resulting pool of 30 CV syllables, generated by combining 13 consonants and 8 vowels, are shown in Table I. Twenty-four syllables were selected for use in the experimental trials. The remaining six syllables were employed in the practice trials, with the exception that one syllable from the experimental pool was also used to construct the practice items in order to fulfill the criteria described below for sequence construction.

For each condition, the eight sequences were created by combining the syllables from the 24-syllable pool for the experimental tasks with the following constraints: No phonemes were repeated within a sequence; each vowel occurred in each ordinal position for each task; any syllables recurring in a condition occurred in a different ordinal position. Note that each syllable string across all conditions was unique in order to eliminate implicit learning that may arise from the repetition of identical items. A digitized recording was made of a female speaker producing the syllables in isolation and the multisyllabic nonwords.

Presentation of the experimental stimuli was controlled by a specialized computer program written in VISUAL BASIC (Microsoft Corporation, 2003). For the syllable sequence task, the participant was asked to listen to each sequence of sounds and repeat them in the same order. The syllable sequences were presented at the rate of one every 1 s for serial recall. For the natural nonwords' and concatenated syllables' conditions, the participant was told that they would hear a made-up word and asked to repeat it back immediately. All responses were recorded digitally and phonetically transcribed by the first author, a trained speech language pathologist.

The average duration of the syllables in the syllable sequence condition was 456 ms ($SD=38.87$); the total trial duration of each sequence including pauses was approximately 3500 ms. The average duration of the natural nonwords was 2040 ms ($SD=47.10$). The concatenated syllables were created such that each individual syllable was 408 ms, and the total trial duration was 2040 ms, equal to the mean length of the naturally produced nonwords. The concatenated syllables were created by compressing each of the relevant syllables from the syllable sequence condition to 408 ms using the time warp similarity function in GOLDWAVE (Goldwave Inc., 2003), which overlaps similar sections of audio to reduce length, and then concatenating the five syllables in sequence with no intervening silence. Thus, the natural nonword and concatenated syllables' trials each presented continuous speech strings of equal duration, but the concatenated syllables were not naturally coarticulated.

B. Results and discussion

Recall accuracy for each syllable in the natural nonword, concatenated syllables, and syllable sequence conditions was scored using a strict serial order criterion according to which a unit was only scored as correct if it was recalled in its original position within the sequence. Table II summarizes the syllable accuracy data for all of the reported experiments. Figure 1 plots the experiment 1 accuracy results for each recall type and serial position. An initial two-way analysis of variance (ANOVA) (recall type × serial position) completed on the syllable accuracy data revealed significant main effects for task, $F(2,46)=11.829$, $p<0.001$, $\eta^2_p=0.34$, and position, $F(4,92)=80.806$, $p<0.001$, $\eta^2_p=0.78$. The interaction failed to reach significance, $F(8,184)=1.788$, $p=0.082$, $\eta^2_p=0.07$. For the effect of recall type, pairwise comparisons with Bonferroni corrections indicated that syllable accuracy was significantly greater in the natural nonword than either the concatenated syllables ($p<0.001$) or syllable sequence conditions ($p=0.013$) while the latter two conditions did not differ ($p>0.05$). Within-subject contrasts revealed a significant quadratic function for the main effect of position corresponding to standard primacy and recency effects across conditions. The results were also subjected to an item analysis, with the syllables at different serial positions treated as the random factor. Significant effects included recall type, $F2(2,8)=7.125$, $p=0.017$, $\eta^2_p=0.664$, and position, $F2(4,8)=37.742$, $p<0.001$, $\eta^2_p=0.95$, but not the interaction, $F2(8,105)=0.507$, $p>0.05$, $\eta^2_p=0.04$. In pairwise comparisons between recall types, there was one significantly different pair: Accuracy was greater in the natural

TABLE II. Syllable accuracy for each serial position and recall type across three experiments.

			Serial position									
			1		2		3		4		5	
Recall type	Total		*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Experiment 1 ($n=24$)												
Natural nonwords[a]	21.08	7.08	6.46	1.59	4.33	1.71	2.88	1.80	2.92	2.02	4.50	1.69
Concatenated syllables[b]	16.25	5.24	5.08	1.47	3.58	1.79	2.54	1.74	2.00	1.62	2.92	1.35
Syllable sequences	17.17	4.30	5.67	1.17	4.17	1.34	2.50	1.32	1.67	1.05	3.17	1.93
Experiment 2 ($n=14$)												
Naturally coarticulated[c]	16.93	2.20	6.07	1.33	4.14	1.75	2.14	1.35	1.71	1.44	2.86	1.66
Spliced valid	16.07	5.62	5.14	1.66	3.93	1.64	3.07	1.59	1.71	1.33	2.21	1.37
Spliced invalid	12.00	4.57	3.64	1.74	2.50	1.23	2.00	1.47	1.21	1.31	2.21	1.37
Experiment 3 ($n=14$)												
Stressed nonwords	20.79	7.22	5.43	1.70	4.29	2.30	3.36	1.55	3.14	1.79	4.57	1.83
Neutral prosody[a]	17.36	7.42	5.57	2.10	3.86	1.99	2.64	1.91	1.86	1.41	3.43	1.95
Concatenated syllables[b]	15.21	7.19	5.07	1.64	2.57	2.03	2.50	2.07	2.29	2.13	2.79	1.72

[a]Constitutes the same recall type of the same items.
[b]The concatenated syllables of experiment 1 were compressed prior to splicing whereas those of experiment 2 were not compressed.
[c]Constitutes the same recall type as in (a) but different phoneme sequences.

nonwords' than the concatenated syllables' condition ($p=0.034$; all remaining comparisons, $p>0.05$).

In this experiment, recall accuracy was significantly greater when repeating naturally articulated nonwords than nonwords of equal duration created by concatenating single syllables or lists of single syllables. These results clearly establish a recall advantage for multisyllabic nonwords over syllable strings of matched content in adults, replicating previous findings for children (Archibald and Gathercole, 2007a). The more accurate repetition of naturally articulated nonwords than concatenated syllables of matched duration indicates that the multisyllabic benefit is not simply due to the more rapid presentation of multisyllabic forms over syllable lists, ruling out explanations related to temporal decay in the serial list paradigm as the source of the difference. One possible explanation of the superior repetition in the natural nonword than concatenated syllable conditions in the present experiment relates to the synthetic manipulation (compression) of the concatenated syllables. Experiment 3 addresses this issue by providing a replication of this finding without the synthetic manipulation. A more interesting possibility is that the naturally articulated nonwords provide additional cues to support retention, and experiment 2 was designed to address one potential source of such cues, coarticulation.

III. EXPERIMENT 2

Results of experiment 1 established that the repetition advantage for multisyllabic nonwords does not lie in the shorter overall duration of such stimuli. One suggestion has been that multisyllabic forms contain additional phonetic signals of segmental structure in the form of coarticulatory and prosodic cues. These cues may in turn facilitate a richer encoding of multisyllabic sequences. Experiment 2 was designed to investigate the extent to which coarticulatory information influences nonword repetition. In this experiment, adults recalled nonword strings as naturally articulated multisyllabic nonwords, multisyllabic nonwords in which each syllable was spliced from a different repetition in a way that maintained coarticulatory information, and as multisyllabic nonwords in which each syllable was spliced in a way that created invalid coarticulatory information. As this experiment represents one of the first investigations of coarticulation in nonword repetition, invalid coarticulatory cues were maximized by producing the required target syllable in a different sequence where phonemes in adjacent syllables were maximally different from the target. For example, selected consonants differed from the target sequence in place and manner of articulation, and voicing. That is, if the target syllable was preceded by a syllable containing a back, voiceless plosive, /k/, it was produced in a context preceded by a labiodental, voiced fricative, /v/. Similarly, if the target syllable was followed by a high front vowel, /i/, it was produced in a context followed by a low back vowel, /ɑ/.

As the aim of experiment 2 was to focus on coarticulatory cues, it was necessary to control prosodic information in the nonwords. In order to achieve this, items in both spliced conditions were created such that each syllable was cut and pasted into an appropriate ordinal location. For example, the initial syllable of a given nonword was always derived from the initial syllable of the original nonword. This process assured that each syllable was prosodically valid, eliminating the possibility that differences in performance were due to intonation and duration cues that were inappropriate for a syllable in a given position. It should be noted also that the use of phonologically equivalent nonwords across all three conditions ensured that phonological content did not influence any set of items to differing extents.

The inclusion of two spliced conditions, containing either valid or invalid articulatory information, addressed the concern that nonword repetition performance could be weak-

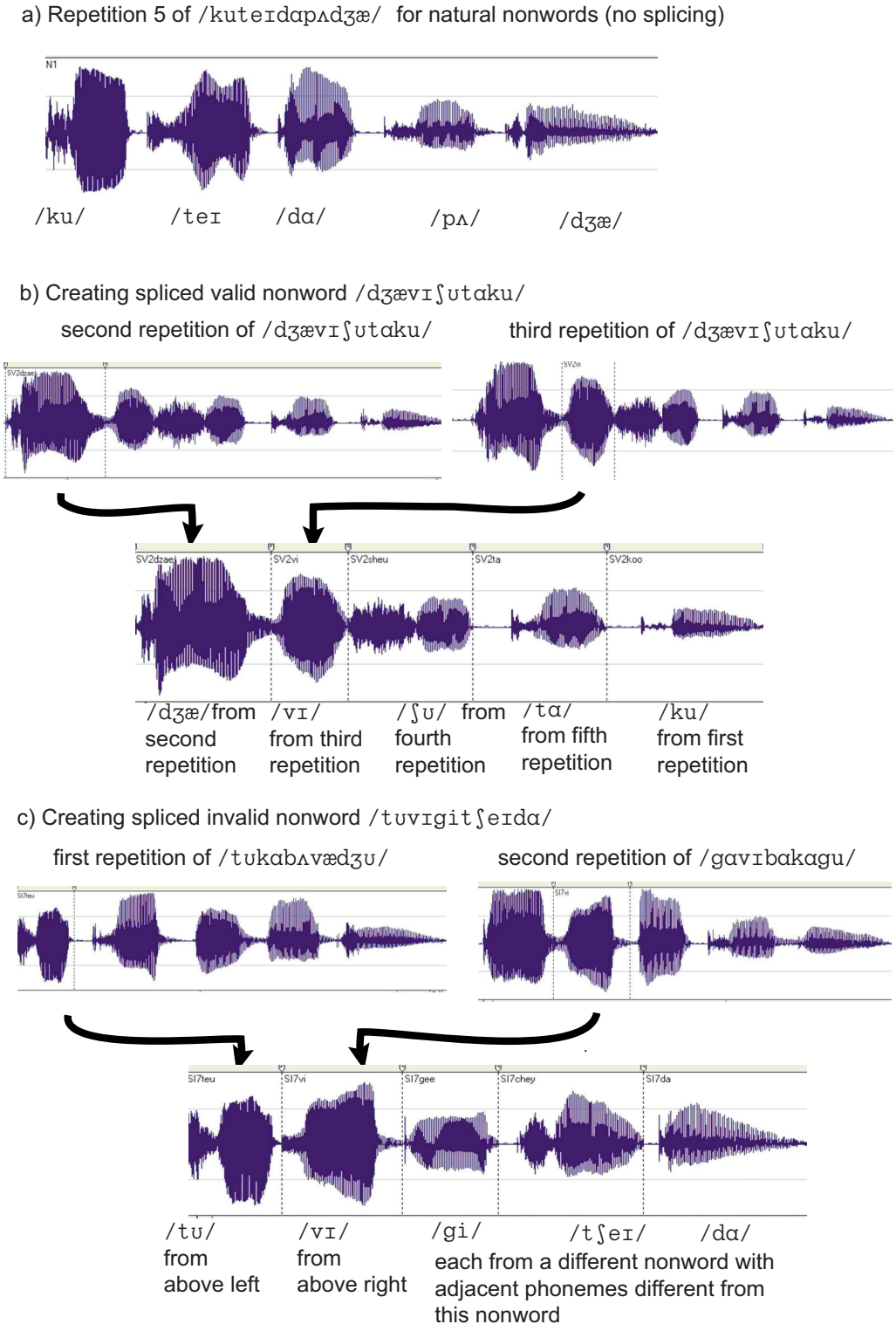


FIG. 1. (Color online) Examples of an item created for each condition.

ened simply due to the presence of spliced syllables. That is, although care was taken to splice syllables in a way that retained the naturalness of the nonwords, the act of splicing the waveform could nevertheless have disrupted the signal in a way that would lead to less accurate repetition. This would be indicated by equally poor repetition accuracy in the spliced conditions containing valid and invalid coarticulatory information. In contrast, if performance was poorer in the invalid condition, it would indicate that coarticulatory cues do, in fact, influence recall in nonword repetition.

A. Method

1. Participants

Fourteen adults ranging in age from 18 to 35 years from the London, ON community were recruited to participate in the present experiment. All spoke English as their primary language.

2. Procedure

All participants completed a repetition task under three conditions, the repetition of naturally produced nonwords

TABLE III. Syllables used to construct stimuli for experiment 2.

Practice trials							
/fʊ/	/wʌ/	/gɑ/	/jʊ/				
Experimental trials							
/kae/	/ʃʊ/	/vɪ/	/pʌ/	/teɪ/	/bɑ/	/gi/	/ku/
/fae/	/dʊ/	/kɪ/	/ʃʌ/	/veɪ/	/tɑ/	/tʃi/	/fu/
/dʒae/	/tʊ/	/tʃɪ/	/dʌ/	/tʃeɪ/	/dɑ/		

(*naturally coarticulated*) and multisyllabic forms spliced from either valid (*spliced valid*) or invalid (*spliced invalid*) contexts. Condition order was counterbalanced across participants such that there were three orders with eight participants each. Items were created following the same principles as employed in experiment 1. CV syllables were constructed from 11 consonants and 8 vowels chosen because they could be paired with phonemes of highly contrastive articulation, reliably segmented in the acoustic waveform, and create nonwords in CV structures. The vowels corresponded to the extreme points of English vowel space: high-mid front, /i,ɪ,eɪ/, high back, /u,ʊ/, mid back, /ʌ/, low front, /æ/, or low back, /ɑ/, and the consonants included /p,b,t,d,k,g,f,v,ʃ,tʃ,dʒ/. There were a total of 26 syllables, 4 of which appeared only in the practice trials (see Table III).

A digitized recording was made of an adult female speaker producing seven repetitions of each nonword. The first and seventh repetitions were rejected in order to eliminate prosodic differences that typically occur for initial and final items in a list. The remaining repetitions were used to create the auditory stimuli in each condition as follows: For the naturally coarticulated condition [Fig. 2(a)], the ten items (two practice and eight trials) were used as recorded, two each from the first through fifth repetitions of the respective nonword.

The target items for the two spliced conditions were phonologically equivalent to those in the natural nonword condition. However, stimulus tokens were constructed by segmenting the recordings of nonwords into individual syllables and re-splicing them. Syllables were segmented by finding the offset of the vowel at the end of each syllable on the acoustic waveform using the software program SOUND FORGE 6.0 (Sonic Foundry Inc., 2002). The spliced valid condition consisted of syllables re-spliced from different tokens of the same nonword item [see Fig. 2(b)]. For example, the first syllable was taken from the first token, the second from the second token, and so on. Order of syllable selection was counterbalanced across stimulus items. Note that this splicing scheme maintained the validity of coarticulatory information in the nonword since each syllable had been produced in the same phonetic context to which it was spliced. In contrast, the spliced invalid stimuli contained spliced syllables taken from five-syllable nonwords in which adjacent syllables had phonemes with highly contrastive articulation demands from those of the target stimulus [Fig. 2(c)]. That is, none of the syllables in the stimulus word had originally been produced with the same adjacent phonemes as the target. For example, to create the spliced invalid nonword /kuʃʊtʃeɪtɑvɪ/, the syllable /ku/ was taken from the nonword /kudibɑgʊteɪ/, where /di/ is distinguished from the target second syllable—/ʃʊ/—in terms of place, manner, and voice of the consonant, and place of the vowel. Similarly, /ʃʊ/ was taken from /væʃʊbɑgifu/, where /væ/ is similarly distinguished from the target first syllable /ku/, and /ba/ from the target third syllable /tʃeɪ/, and so on. Note that additional five-syllable sequences were recorded as necessary to provide appropriate syllables to construct target stimuli. The same process was employed for selecting syllables to be spliced as in the valid condition above.

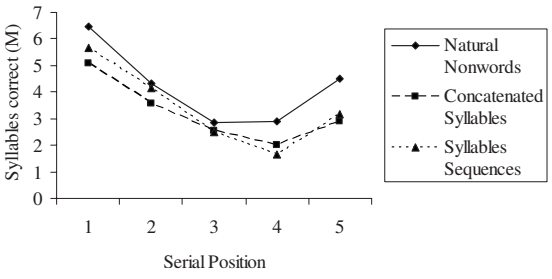


FIG. 2. Mean syllable accuracy (out of eight trials) at each serial position for the natural nonwords, concatenated syllables, and syllable sequences conditions.

It should be noted that the splicing procedures were identical for the valid and invalid conditions; thus any detriment caused by the act of splicing would be equivalent across both conditions. Nevertheless, it was the case that the syllable transitions were less smooth for the invalid than valid conditions. In order to ensure that the resulting invalid sequences were perceived as the target sequence, a trained research assistant unaware of the correct forms or nature of the experiment phonetically transcribed the spliced invalid nonwords. For 11 of the 50 syllables, changes in the oral aperture related to anticipatory coarticulatory processes occurring before vowel offset resulted in the perception of the presence of an additional consonant. That is, the syllable, /ku/, spliced from the sequence /kudi../ was perceived as /kud/ despite the fact that the alveolar closure had not yet been achieved. The change in the oral aperture leading to this misperception was clearly visible as a change in the vowel's acoustic waveform. This section of the waveform was attenuated to eliminate the misperception. On average, the duration of the attenuation was 0.06 s (SD=0.03). While this process maintained the invalid coarticulatory cues and duration, it reduced opposing coarticulatory information.

Stimulus presentation was controlled by the same computer software program as employed in experiment 1. Participant responses were digitally recorded and phonetically transcribed. Recall accuracy in the naturally coarticulated,

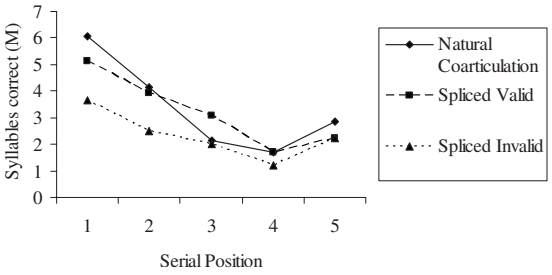


FIG. 3. Mean syllable accuracy (out of eight trials) at each serial position for the naturally coarticulated, spliced valid, and spliced invalid nonword conditions.

spliced valid, and spliced invalid conditions was scored by the first author using the same serial order criterion as employed in experiment 1.

B. Results and discussion

Figure 3 plots the experiment 2 accuracy data across serial position for each recall type (naturally coarticulated, spliced valid, and spliced invalid). An initial two-way ANOVA (recall type × serial position) completed on the syllable accuracy data revealed a significant main effect of recall type, $F(2,26)=13.860$, $p<0.001$, $\eta^2_p=0.52$, and position, $F(4,52)=46.346$, $p<0.001$, $\eta^2_p=0.78$, and a significant interaction, $F(8,104)=3.059$, $p=0.004$, $\eta^2_p=0.19$. For the main effect of recall type, pairwise comparisons with Bonferroni correction revealed significantly less accurate performance in the spliced invalid condition than both the naturally coarticulated ($p=0.001$) and spliced valid conditions ($p=0.002$) while the latter two conditions did not differ ($p>0.05$). The effect of recall type on serial position was investigated using simple pairwise *t*-tests. There was no difference in the recency effects across conditions as evidenced by no task differences at the fourth and fifth serial positions ($p>0.05$, all cases). Primacy effects were significantly reduced in the spliced invalid condition as indicated by significantly less accurate performance in this than either of the other two conditions at positions 1 and 2 and at position 3 compared to the spliced valid condition, $t(13)>2.4$, $p<0.05$, all cases. In the corresponding items analysis, there was a significant main effect of position, $F2(4,8)=15.023$, $p=0.001$, $\eta^2_p=0.88$. The task term did not reach significance, $F2(2,8)=2.657$, $p=0.13$, $\eta^2_p=0.40$, likely a reflection of the low power in the item analysis as a result of the small number of items. Importantly, the effect size for the task term in the item analysis was comparable to the F1 ANOVA and the means reflected a similar trend with higher proportions of participants repeating syllables accurately in the naturally coarticulated ($M=0.42$; $SD=0.27$) and spliced valid conditions ($M=0.40$; $SD=0.24$) compared to the spliced invalid task ($M=0.32$; $SD=0.24$). The interaction was not significant, $F2(8,105)=1.077$, $p>0.05$, $\eta^2_p=0.08$.

Thus, syllable accuracy was greater in the two conditions presenting valid coarticulatory information (naturally coarticulated and spliced valid nonwords) compared to the spliced invalid condition. These results demonstrate the positive influence of valid over invalid coarticulatory cues in immediate repetition. It is possible that coarticulatory processes enable a richer encoding of temporary phonological representations because they transmit information about more than one phoneme at a time (Liberman *et al.*, 1967). One interesting aspect of the present results is the finding that there was very little cost to the splicing of the acoustic waveform in that recall accuracy did not differ between the naturally coarticulated and spliced valid conditions. Recent evidence has demonstrated superior repetition for natural vs synthetic speech (e.g., Humes *et al.*, 1993). Nevertheless, the present findings suggest that immediate repetition processes can withstand small disruptions to the acoustic waveform. Such a skill may be adaptive, as it would assist with recognition of highly similar (but not identical) acoustic forms.

The coarticulatory effect in the present experiment was more pronounced in the primacy portion of the recall curve. The presence of primacy and recency effects in nonword repetition has been established previously (Gupta, 2005), as have differential primacy and recency effects in immediate recall generally (Cowan *et al.*, 2002; Oberauer, 2003). Typically, primacy effects are attributed to an automatic attentional gradient operating during encoding such that items presented first receive more attention and therefore are encoded (and recalled) more accurately (e.g., Brown *et al.*, 2000; Farrell and Lewandowsky, 2002; Oberauer, 2003; Page and Norris, 1998). The present findings suggest that the presence of valid coarticulatory information may further enhance encoding especially when the attentional gradient is highest.

These findings indicate that the more accurate recall of multisyllabic forms than syllable lists relates, at least in part, to the coarticulatory information inherent in the multisyllabic strings. Experiment 3 investigates the influence on retention of another factor that differs between these forms, prosody.

IV. EXPERIMENT 3

Prosodic structure has been found to play a role in nonword repetition (Gallon *et al.*, 2007; Roy and Chiat, 2004; Yuzawa and Saito, 2006). Nevertheless, experiments 1 and 2 were designed to minimize prosodic influences on recall by employing equal-stress-across-syllables. It could be argued that such stimuli remain list-like leading to the underestimation of the importance of processes related to the learning of more plausible novel words. Experiment 3 was designed to examine the influence on nonword repetition of information inherent to more plausible nonwords, prosodic information. In this final experiment, stress patterns common to five-syllable words in English were imposed on nonwords creating within-word stress contours. This condition was compared to the naturally articulated nonwords employed in the previous experiments that had neutral prosody (i.e., equal-stress-across-syllables). An aprosodic condition was also created by concatenating syllables recorded in isolation. In order to avoid the temporal and manipulated confound present in experiment 1, the isolated syllables were produced with a naturally short duration such that no compression was required and the sequence duration was not significantly longer.

Graded and decreasing performance across all three of these conditions, within-word stress, neutral prosody, and

aprosodic nonwords, would provide strong evidence of the positive influence of prosodic information on immediate recall of multisyllabic forms. More accurate recall in the within-word stress relative to the neutral prosody conditions would point to the importance of syllable level prosodic information rather than strictly a suprasegmental benefit.

A. Method

1. Participants

The 14 adults who participated in experiment 2 also completed experiment 3. Order was fixed such that all participants performed experiment 2 first, with a 5-min break between sessions.

2. Procedure

All participants completed a repetition task under three conditions, the repetition of naturally produced nonwords with within-word stress (*stressed nonwords*), naturally produced nonwords with equal-stress-across-syllables (*neutral prosody*), and an aprosodic condition (*concatenated syllables*). Condition order was counterbalanced across participants such that there were three orders with eight participants each. The phonological form of items was the same as those employed in experiment 1 (which differed from those in experiment 2).

A digitized recording was made of an adult female speaker producing four repetitions of each syllable sequence used in the stressed nonword and neutral prosody conditions, and the item was selected from either the second or third repetition. Note that fewer repetitions were needed because the items were not spliced from different repetitions but taken as a single production. Items in the stressed nonword condition were produced with one of three stress patterns common to English multisyllabic words [i.e., weak (W)-strong[1] (S)-WS[2]W as in "accumulation;" S[2]WS[1]WW as in "hippopotamus;" S[1]WWS[2]W as in "qualification"]. Items in the neutral prosody condition were produced with equal–stress-across-syllables. For the concatenated syllables' condition, syllables were naturally produced in isolation with short duration so that the average duration of the syllable string once concatenated ($M=1.91$ s, $SD=0.08$) would be equivalent to that of the stressed nonword condition ($M=1.85$ s, $SD=0.17$; $t(18)=-1.105$, $p>0.05$). This procedure eliminated the need to artificially compress syllable length as occurred in experiment 1. Note that the items in the neutral prosody group had the longest average duration ($M=2.08$ s, $SD=0.08$) compared to either of the other two conditions, $t(18)>3.9$, $p<0.05$, both cases.

Stimulus presentation was controlled by the same computer software program as employed in experiment 1. Participant responses were digitally recorded and phonetically transcribed. Recall accuracy was scored by Archibald using the same serial order criterion as employed in experiment 1.

B. Results and discussion

Figure 4 plots the experiment 3 syllable accuracy data across serial positions for the stressed, neutral, and concatenated syllables' conditions. An initial two-way ANOVA (recall type × serial position) completed on the accuracy data revealed significant effects of recall type, $F(2,26)=14.266$, $p<0.001$, $\eta^2_p=0.52$, position, $F(4,52)=18.107$, $p<0.001$, $\eta^2_p=0.58$, and recall type and position, $F(8,104)=2.071$, $p=0.045$, $\eta^2_p=0.14$. Pairwise comparisons with Bonferroni correction for recall type revealed significantly more accurate repetition in the stressed than both the neutral ($p=0.009$) and concatenated syllable conditions ($p=0.003$). Also, there was a marginally significant difference between the latter two conditions ($p=0.066$). The interaction was explored through pairwise *t*-tests. The most consistent effect was the recency advantage for stressed nonwords compared to both the remaining conditions at serial position 5, and to the neutral nonwords at positions 3 and 4. The reduced primacy effect for the concatenated syllables observed in experiment 1 was present here at position 2. Remaining comparisons were nonsignificant ($p>0.05$). In the corresponding items' analysis, significant effects were found for recall type, $F2(2,8)=8.537$, $p=0.01$, $\eta^2_p=0.68$, and position, $F2(4,8)=20.426$, $p<0.001$, $\eta^2_p=0.91$, but not for the interaction term, $F2(8,105)=0.662$, $p>0.05$, $\eta^2_p=0.05$. In pairwise comparisons for task type, repetition accuracy was significantly greater in the stressed nonword than concatenated syllables ($p=0.004$). All remaining comparisons were nonsignificant ($p>0.05$, all cases).

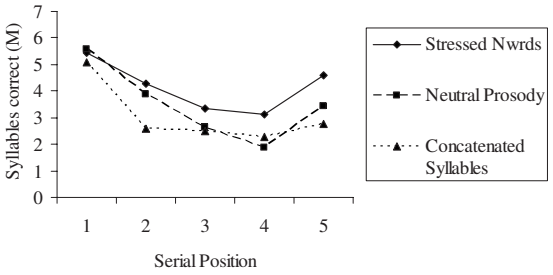


FIG. 4. Mean syllable accuracy (out of eight trials) at each serial position for the stressed nonwords', neutral prosody, and concatenated syllables' conditions.

It should be noted that in one further analysis comparing syllable accuracy for the strong vs weakly stressed syllables in the stressed nonword condition, no significant difference was found between stress types, $t(26)=-1.206$, $p>0.05$.

In this experiment, syllable recall was more accurate in the F2 analysis for the within-word stress condition than concatenated syllables and in the F1 analysis than either of the other two conditions, an effect that was greater in the recency portion of the curve. Consistent with previous findings (Roy and Chiat, 2004), within-word prosodic pattern facilitated recall. It could be argued that the presence of a within-word stress pattern increased the wordlikeness of the nonwords, thereby facilitating recall through the lexical effect (e.g., Gathercole, 1995). Lexicality, however, has traditionally been associated with improved primacy rather than recency (e.g., May and Sande, 1982). One explanation of recency effects is the temporal distinctiveness hypothesis, which holds that items presented more recently will be more distinctive on average and thereby enjoy a recall advantage (Brown *et al.*, 2000). It may be that the differential stress patterns across syllables in the stressed nonwords of the

present experiment enhanced their distinctiveness leading to more accurate recall overall, and in the final serial positions, in particular.

It should be noted that experiment 3 essentially replicated the finding from experiment 1 that naturally articulated nonwords were repeated more accurately than concatenated syllables of matched duration. In addition to the more accurate performance on stressed nonwords, the higher scores on the neutral prosody nonwords (equivalent to the natural nonwords; condition in experiment 1) than the concatenated syllables of shorter duration were marginally significant. While this latter result may reflect a smaller effect size due to the elimination of the synthetic manipulation in the concatenated syllables' condition, the marginally significant outcome with this substantially smaller sample size (experiment 3, $n=14$; experiment 1, $n=24$) in favor of the (longer) neutral prosody sequences provides evidence that temporal duration does not account for the nonword repetition advantage. It is possible that a larger sample size may have reduced variability in experiment 3 and lead to significant results matching those of experiment 1.

V. GENERAL DISCUSSION

The present study examined the influence of three acoustic-phonetic factors—list duration, subphonetic detail, and prosody—on nonword repetition performance. In the first experiment, we compared repetition of naturally articulated five-syllable nonwords to recall of concatenated syllable strings of equal duration but which were created from syllables produced in isolation. In a second experiment, we more closely examined the influence of coarticulatory cues by comparing nonword repetition of multisyllabic nonwords containing either valid or invalid coarticulatory information, and naturally produced nonwords. Finally, we compared multisyllabic forms varying in prosodic detail including common English stress patterns, suprasegmental contours only, and aprosodic forms created by concatenating syllables recorded in isolation. The results established superior immediate recall for multisyllabic nonword forms over single syllable lists in adults. The repetition advantage was not a result of the shorter duration of multisyllabic strings. Valid coarticulatory cues enhanced recall, especially over the first syllables of the nonword. Within-word stress facilitated repetition, particularly over the final positions of the sequence.

These results clearly establish a recall advantage for coarticulated multisyllabic nonwords that cannot be attributed to their overall shorter duration. One reason that multisyllabic nonwords may enjoy a repetition advantage over syllable lists or concatenated syllables is that the structure of the nonwords may promote chunking within STM. It has been suggested that chunks form when stimuli already form a unit or have external punctuation serving to group stimuli (McLean and Gregg, 1967), both of which apply in the case of multisyllabic nonwords. The use of nonwords in the present studies limits long-term memory processes forcing greater reliance on STM estimated to have a capacity of 4 ± 1 chunks (Cowan, 2001). Thus, the five-syllable syllable lists that resist chunking may have been more difficult to recall than multisyllabic forms.

Not only was nonword repetition more accurate than syllable list recall in the present studies, scores were higher when valid coarticulatory information was present in the multisyllabic form. The presence of invalid coarticulatory cues disrupted recall to the greatest extent, as indicated by the lowest performance across the three studies. The coarticulation benefit was a robust finding with valid coarticulatory information preserved despite splicing of the waveform. The greater primacy effect may indicate that coarticulation information enhances encoding, especially when the attentional gradient is highest. Traditionally, primacy effects have been considered to reflect long-term memory support of traces retained in memory for the longest time period (Craik, 1970). It may be that valid coarticulatory information enhances encoding, which in turn facilitates activation of long-term phonological representations.

It is interesting to speculate on how sensitivity to coarticulatory information may influence nonword repetition. One suggestion has been that coarticulation speeds processing because it transmits information about more than one phoneme at a time (Liberman *et al.*, 1967). Speeded processing may have positive effects on speech perception, memory, and language in that inputs may be recognized earlier and additional time afforded for rehearsal and other processing requirements. Relatedly, children with specific language impairment (SLI), an unexplained difficulty learning language, have more difficulty repeating multisyllabic nonwords than single syllable nonword lists (Archibald and Gathercole, 2007b). It may be that children with SLI are less sensitive to, or able to capitalize on, coarticulatory cues in connected speech.

The present findings also indicate that prosodic pattern positively influences immediate repetition of novel word forms. Familiar within-word stress patterns conveyed a marked advantage that was not tied simply to the more accurate recall of strongly stressed syllables. Natural prosodic patterns may further enhance the chunking processes described above by highlighting distinctions between groups of syllables. Increased distinctiveness between syllables may account for the greater recency effect observed in the present study as well.

One other possible explanation of the prosodic advantage in the present study relates to list position. It is well established in serial recall research (e.g., Treiman and Danis, 1988) that migrated or transposed units in responses are much more likely to move to similar prosodic locations (i.e., a word-initial phoneme exchanges with a word-initial rather than word-final phoneme). It may be that the presence of fewer syllables with equivalent stress markings in the within-word stress condition in the present study decreased opportunities for same-position moves compared to the other conditions with equal-stress-across-syllables.

VI. CONCLUSION

Nonword repetition was examined in three studies exploring the influence of nonword duration, coarticulatory in-

formation, and prosodic cues. The results indicate that phoneme strings presented as a multisyllabic form rather than an isolated syllable list are more accurately recalled. This advantage cannot be readily explained by STM accounts related to temporal decay. Both the presence of valid coarticulatory information and within-word stress patterns enhanced recall. Coarticulatory effects were observed over the primacy portion of the recall curve whereas prosodic influences were linked to greater recency. These findings suggest that several processes operating in conjunction in new word learning enable rich encoding and retention of novel phonological forms facilitating transfer to long-term memory. It may be that individuals with word learning deficits such as those with SLI are impaired in one or more of these processes.

Archibald, L. M. D., and Gathercole, S. E. (**2006**). "Nonword repetition: A comparison of tests," J. Speech Lang. Hear. Res. **49**, 970–983.

Archibald, L. M. D., and Gathercole, S. E. (**2007a**). "Nonword repetition and serial recall: Equivalent measures of verbal short-term memory?," Appl. Psycholinguist. **28**, 587–606.

Archibald, L. M. D., and Gathercole, S. E. (**2007b**). "Nonword repetition in specific language impairment: More than a phonological short-term memory deficit," Psychon. Bull. Rev. **14**, 919–924.

Atkins, P. W. B., and Baddeley, A. D. (**1998**). "Working memory and distributed vocabulary learning," Appl. Psycholinguist. **19**, 537–552.

Baddeley, A. D. (**1986**). *Working Memory* (Oxford University Press, Oxford).

Beddor, P. S., Harnsberger, J. D., and Lindemann, S. (**2002**). "Language-specific patterns of vowel-to-vowel coarticulation: Acoustic structures and their perceptual correlates," J. Phonetics **30**, 591–627.

Bell-Berti, F., and Harris, K. S., and Tuller, B. (**1989**). "Some experiments on anticipatory coarticulation," J. Acoust. Soc. Am. **64**, S92–S92.

Brown, G. D. A., Preece, T., and Hulme, C. (**2000**). "Oscillator-based memory for serial order," Psychol. Rev. **107**, 127–181.

Chiat, S., and Roy, P. (**2007**). "The Preschool Repetition Test: An evaluation of performance in typically developing and clinically referred children," J. Speech Lang. Hear. Res. **50**, 429–443.

Cho, T. H. (**2004**). "Prosodically conditioned strengthening and vowel-to-vowel coarticulation in English," J. Phonetics **32**, 141–176.

Choi, J. D., and Keating, P. (**1991**). "Working Papers in Phonetics," WPP, **78**, Department of Linguistics, UCLA. Paper No. 78, http://repositories.cdlib.org/uclaling/wpp/No78

Cowan, N., (**1999**). in *Models of Working memory*, edited by A. Miyake and P. Shah (Cambridge University Press, New York), pp. 62–101.

Cowan, N. (**2001**). "The magical number 4 in short-term memory: A reconsideration of mental storage capacity," Behav. Brain Sci. **24**, 87–114.

Cowan, N., Saults, J. S., Elliott, E. M., and Moreno, M. (**2002**). "Deconfounding serial recall," J. Mem. Lang. **46**, 153–177.

Craik, F. I. M. (**1970**). "The fate of primary memory items in free recall," J. Verbal Learn. Verbal Behav. **9**, 143–148.

Dollaghan, C., and Campbell, T. F. (**1998**). "Nonword repetition and child language impairment," J. Speech Lang. Hear. Res. **41**, 1136–1146.

Edwards, J., Beckman, M. E., and Munson, B. (**2004**). "The interaction between vocabulary size and phonotactic probability effects on children's production accuracy and fluency in nonword repetition," J. Speech Lang. Hear. Res. **47**, 421–436.

Farrell, S., and Lewandowsky, S. (**2002**). "An endogenous model of ordering in serial recall," Psychon. Bull. Rev. **9**, 59–79.

Gallon, N., Harris, J., and van der Lely, H. (**2007**). "Nonword repetition: An investigation of phonological complexity in children with grammatical SLI," Clin. Linguist. Phon. **21**, 435–455.

Gathercole, S. E. (**1995**). "Is nonword repetition a test of phonological memory or long-term knowledge? It all depends on the nonwords," Mem. Cognit. **23**, 83–94.

Gathercole, S. E., and Baddeley, A. D. (**1989**). "Evaluation of the role of phonological STM in the development of vocabulary in children: A longitudinal study," J. Mem. Lang. **28**, 200–213.

Gathercole, S. E., and Baddeley, A. D. (**1993**). *Working Memory and Language* (Erlbaum, Hillsdale, NJ).

Gathercole, S. E., Willis, C., Emslie, H., and Baddeley, A. D. (**1992**). "Phonological memory and vocabulary development during the early school years: A longitudinal study," Dev. Psychol. **28**, 887–898.

Gathercole, S. E., Willis, C., Emslie, H., and Baddeley, A. D. (**1994**). "The children's test of nonword repetition: A test of phonological working memory," Memory **2**, 103–127.

Goldwave Inc., (**2003**). GOLDWAVE DIGITAL AUDIO EDITOR, St. John's, NL.

Gupta, P. (**2003**). "Examining the relationship between word learning, nonword repetition, and immediate serial recall in adults," Q. J. Exp. Psychol. **65A**, 1213–1236.

Gupta, P. (**2005**). "Primacy and recency in nonword repetition," Memory **13**, 318–324.

Gupta, P., Lipinski, J., Abbs, B., and Lin, P.-H. (**2005**). "Serial position effects in nonword repetition," J. Mem. Lang. **53**, 141–162.

Hansen, G. F. (**2003**). "Lip gestures in Danish vowels: Articulation and coarticulation," Phonum **9**, 65–68.

Humes, L. E., Nelson, K. J., Pisoni, D. B., and Lively, S. E. (**1993**). "Effects of age on serial recall of natural and synthetic speech," J. Speech Hear. Res. **36**, 634–639.

Liberman, A. M., Cooper, E. S., Shankweiler, D. P., and Studdert-Kennedy, M. (**1967**). "Perception of the speech code," Psychol. Rev. **74**, 431–61.

MacNeilage, P. (**1980**). "The control of speech production," in *Child Phonology, Vol. 1: Production*, edited by G. Yeni-Komshian, J. Kavanagh, and C. A. Ferguson (Academic, New York), pp. 9–21.

Martin, J. G., and Bunnell, H. T. (**1981**). "Perception of anticipatory coarticulation effects," J. Acoust. Soc. Am. **69**, 559–567.

Martin, J. G., and Bunnell, H. T. (**1982**). "Perception of anticipatory coarticulation effects in vowel-stop consonant-vowel sequences," J. Exp. Psychol. Hum. Percept. Perform. **8**, 473–488.

Masoura, E. V., and Gathercole, S. E. (**1999**). "Phonological short-term memory and foreign vocabulary learning," Int. J. Psychol. **34**, 383–388.

Masoura, E. V., and Gathercole, S. E. (**2005**). "Phonological short-term memory skills and new word learning in young Greek children," Memory **13**, 422–429.

May, R. B., and Sande, G. N. (**1982**). "Encoding expectancies and word frequency in recall and recognition," Am. J. Psychol. **95**, 485–495.

McLean, R. S., and Gregg, L. W. (**1967**). "Effects of induced chunking on temporal aspects of serial recitation," J. Exp. Psychol. **74**, 455–59.

Microsoft Corporation (**2003**). VISUAL BASIC.NET, Redmond, WA.

Modarresi, G., Sussman, H., Lindblom, B., and Burlingame, E. (**2004**). "An acoustic analysis of the bidirectionality of coarticulation in VCV utterances," J. Phonetics **32**, 291–312.

Mok, P. K. P. (**2007**) "Effects of syllable structure on V-to-V coarticulation" (Thai vs. English). *In Proceeding of the 16th International Congress of Phonetic Sciences*, Saarbrücken, Germany, 421–424.

Nairne, J. S. (**2002**). "Remembering over the short-term: The case against the standard model," Annu. Rev. Psychol. **53**, 53–81.

Nittrouer, S., and Whalen, D. H. (**1989**). "The perceptual effects of child-adult differences in fricative-vowel coarticulation," J. Acoust. Soc. Am. **86**, 1266–1276.

Oberauer, K. (**2003**). "Understanding serial position curves in short-term recognition and recall," J. Mem. Lang. **49**, 469–483.

Ostreicher, H. J., and Sharf, D. J. (**1976**). "Effects of coarticulation on the identification of deleted consonant and vowel sounds," J. Phonetics **4**, 285–301.

Page, M. P. A., and Norris, D. (**1998**). "The primacy model: A new model of immediate serial recall," Psychol. Rev. **105**, 761–781.

Paris, P. A., Thomson, M. H., Gilson, R. D., and Kincaid, J. P. (**2000**). "Linguistic cues and memory for synthetic and natural speech," Hum. Factors **25**, 17–32.

Reeves, C., Schmauder, A. R., and Morris, R. K. (**2000**). "Stress grouping improves performance on an immediate serial list recall task," J. Exp. Psychol. Learn. Mem. Cogn. **26**, 1638–1654.

Roy, P., and Chiat, S. (**2004**). "A prosodically controlled word and nonword repetition task for 2-to 4-year olds: Evidence from typically-developing children," J. Speech Lang. Hear. Res. **47**, 223–234.

Service, E. (**1992**). "Phonology, working memory and foreigh-language learning," Q. J. Exp. Psychol. **45A**, 21–50.

Sharf, D. J., and Ohde, R. N. (**1981**). in *Speech and Language: Advances in Basic Research and Practice*, edited by N. Lass (Academic, New York),

Vol. **5**, pp. 153–247.
Shriberg, L., and Kwiatowski, J. (**1994**). "Developmental phonological disorders I: A clinical profile," J. Speech Hear. Res. **10**, 828–835.
Sonic Foundry Inc., (**2002**). SOUND FORGE 6.0, Madison, WI.
Southwood, H., Dagenais, P. A., Sutphin, S. M., and Garcia, J. M. (**1997**). "Coarticulation in apraxia of speech: A perceptual, acoustic, and electropalatographic study," Clin. Linguist. Phonetics **11**, 179–203.
Tjaden, K., and Sussman, J. (**2006**). "Perception of coarticulatory information in normal speech and dysarthria," J. Speech Lang. Hear. Res. **49**, 888–902.
Treiman, R., and Danis, C. (**1988**). "Short-term memory errors for spoken syllables are affected by the linguistic structure of the syllables," J. Exp. Psychol. Learn. Mem. Cogn. **14**, 145–152.
Yuzawa, M., and Saito, S. (**2006**). "The role of prosody and long-term phonological knowledge in Japanese children's nonword repetition performance," Cognit. Dev. **21**, 146–157.

Modeling the effect of channel number and interaction on consonant recognition in a cochlear implant peak-picking strategy

Carl Verschuur
Hearing and Balance Centre, Institute of Sound and Vibration Research, University of Southampton, Highfield, Southampton SO16 7ES, United Kingdom

(Received 17 March 2008; revised 21 November 2008; accepted 14 December 2008)

Difficulties in speech recognition experienced by cochlear implant users may be attributed both to information loss caused by signal processing and to information loss associated with the interface between the electrode array and auditory nervous system, including cross-channel interaction. The objective of the work reported here was to attempt to partial out the relative contribution of these different factors to consonant recognition. This was achieved by comparing patterns of consonant feature recognition as a function of channel number and presence/absence of background noise in users of the Nucleus 24 device with normal hearing subjects listening to acoustic models that mimicked processing of that device. Additionally, in the acoustic model experiment, a simulation of cross-channel spread of excitation, or "channel interaction," was varied. Results showed that acoustic model experiments were highly correlated with patterns of performance in better-performing cochlear implant users. Deficits to consonant recognition in this subgroup could be attributed to cochlear implant processing, whereas channel interaction played a much smaller role in determining performance errors. The study also showed that large changes to channel number in the Advanced Combination Encoder signal processing strategy led to no substantial changes in performance. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3075554]

PACS number(s): 43.71.Ky, 43.71.Es [PEI] Pages: 1723–1736

I. INTRODUCTION

The majority of current cochlear implant users obtain high levels of open set speech recognition in quiet (Meyer *et al.*, 2003). Nevertheless, even the most successful cochlear implant users are poorer than normal hearing listeners at speech discrimination, particularly in adverse listening conditions. One of the sources of perceptual difficulties for cochlear implant user is thought to be impaired spectral resolution (Donaldson and Nelson, 2000). This occurs because cochlear implant processing converts the incoming acoustic signal into a small number of envelope values. However, impaired spectral resolution and associated perceptual deficits may also be due to cross-electrode spread of excitation, or channel interaction. This limited spectral resolution can be seen, indirectly, in the types of consonant errors that are typically shown by implant users. A number of studies have shown poor perception of place of articulation perception as compared to manner and voicing in cochlear implant users (e.g., Donaldson and Nelson, 2000; Dorman *et al.*, 1997; Dorman *et al.*, 1998; Van Tasell *et al.*, 1992; Verschuur, 2005). Consonant place contrasts are signaled by spectral cues such as the spectral peak of a plosive burst or fricative noise, or the onset frequency of a formant transition (Stevens and Blumstein, 1978). These cues rely on relatively fine frequency resolution, whereas cues to voicing and manner in consonants rely more on temporal and amplitude resolution, such as the comparison of the onset between voiced and voiceless components of a signal.

It is not wholly clear to what extent cochlear implant users' perceptual deficits are due to information loss associated with signal processing and to what extent they are due to loss of information caused by channel interaction or other signal distortions occurring at the interface between the electrode array and auditory nervous system. Indirect evidence for the role of channel interaction above and beyond the role of signal processing is provided by a number of studies that have shown that cochlear implant users' performance does not generally improve beyond the level of performance obtained with between six and ten active channels, even where it is possible to provide a larger number of active channels in the device (Dorman and Loizou, 1997; Dorman and Loizou, 1998; Friesen *et al.*, 2001; Loizou *et al.*, 1999). It is possible that the upper limit in the number of "perceptual channels" available to the implant user could be due to cross-channel current spread, known as channel interaction (Throckmorton and Collins, 1999, 2002). However, it is also possible that implant processing means that additional useful information is not available above eight channels. In order to clarify this issue, acoustic model studies can be useful in providing indirect evidence as to whether channel interaction plays a role in determining the performance ceiling in implant users (Friesen *et al.*, 2001). Some studies have attempted to equate speech recognition performance in implant users with acoustic models that vary by channel number. Fu *et al.* (1998) found that a group of three users of the Nucleus 22 device showed similar levels of speech recognition in noise to normal hearing subjects listening to an acoustic model with a relatively small number of channels and concluded that loss of spectral resolution was implicated in poor performance in noise by implant users. However, channel number may not

be a complete or accurate model of channel interaction. Consequently, some authors have used an acoustic model in which the number of analysis channels is kept fixed but channel overlap is varied, or where channel number and channel overlap are co-varied. Shannon *et al.* (1998) used a simulation with overlap of filter skirts of the noise bands, thus creating an effective spectral smearing effect. They found that channel overlap made little difference to speech recognition, although it should be noted that the acoustic model used in the study had only four spectral channels; hence spectral information was somewhat limited even without overlap. Throckmorton and Collins (2002) described acoustic models of channel interaction and also other spectral anomalies that are associated with cochlear implant use, such as pitch reversals, indiscriminable electrodes, and forward masking. They compared sentence and consonant recognition abilities between the different acoustic models to determine which might have the greatest impact on speech perception abilities. The authors found that models of spectral channel interaction had the greatest detrimental effect on consonant recognition.

A few other studies have attempted to compare different acoustic models incorporating spectral channel interaction against cochlear implant user performance directly. Fu and Nogaki (2005) compared number of channels with changes in spectral resolution using spectral smearing. The outcome measure used was release from masking, as shown by sentence recognition in noise. The acoustic models were based on a fixed-channel strategy using infinite impulse response (IIR) filter banks and varied by channel number (16, 8, and 4) and spectral overlap between channels (24 dB/octave or 6 dB/octave slope). The authors found that release from masking in sentence recognition was modeled best by acoustic models with broadly overlapping filters (6 dB/octave slope) and that better-performing implant users' performance was approximated with either an 8-channel or 16-channel model, while worse-performing users' performance was best approximated by a 4-channel model. However, the ten implant subjects used a variety of implant devices, meaning that the study did not control for signal processing characteristics across implant users or between implant and acoustic model conditions.

Laneau *et al.* (2006) undertook a series of experiments in which spectral overlap between adjacent channels was systematically varied and the affect on fundamental frequency perception was determined and compared with a previous experiment (Laneau *et al.*, 2004) with implant users. The authors used an acoustic model with noise-band carrier stimuli and assessed pitch discrimination abilities as a function of degree of carrier overlap varying from no overlap to overlap equivalent to 10 mm spread of excitation. The precise pattern of filter overlap was based on the model of channel interaction of Black and Clark (1980) and assumed asymmetric spread of excitation along the cochlea. Two separate experiments showed a close match between Nucleus 24 users and acoustic models with 1 mm spread of excitation. The acoustic model used the same filter bank as was used in the group of Nucleus 24 users against which performance was compared (Laneau *et al.*, 2004); hence the comparison between acoustic model and cochlear implant user results is particularly compelling.

Channel interaction may also be a contributor to cochlear implant variation below the performance ceiling; e.g., it could help to explain why some implant users perform less well or reach a lower asymptote in performance as a function of channel number. One way to understand variance between cochlear implant users is to compare overall speech perception with more specific perceptual abilities. Munson *et al.* (2003) applied this approach by assessing variation in perception of particular phoneme features between better- and worse-performing cochlear implant users, where overall performance was defined by overall phoneme identification ability. The underlying assumption in their study was that if a larger difference is obtained between better- and worse-performing cochlear implant users for a particular consonant feature, then this would suggest possible reasons for inter-user variation, given that consonant features are related to particular perceptual abilities (in the case of consonant place, this would be spectral resolution). The authors assessed consonant confusion errors in 30 cochlear implant users, of whom 12 were users of the Nucleus 22 device and 18 were users of the Clarion device implementing a range of strategies. Consonant recognition was analyzed in terms of transmission of the features voicing, place, manner, and duration. Relative performance for different consonant features did not differ between better and worse performers (overall performance being defined by total percent correct score on the consonant recognition task); i.e., the same pattern applied to both better and worse performers, with percentage information transmitted being better for voicing and manner than place. The authors suggested, on the basis of this, that it is more likely that cochlear implant processing information loss may explain the *relative* transmission of consonant features, while individual differences relate to absolute performance levels. However, the authors' findings do not exclude the possibility that better performers' perceptual limitations were due to both processing loss and electrical/neural interface information loss.

A more powerful approach might be to combine this type of analysis of inter-user variation with a matching acoustic model that varied some parameters of interest. The present research aimed to use this combined approach and, moreover, to focus on two methodological factors which could impact on the validity of results, namely, the homogeneity of signal processing characteristics used by implant subjects and the careful matching between acoustic model and implant user characteristics. A possible confounding factor affecting the predictive power of acoustic model studies is the degree of correspondence between cochlear implant users' processor characteristics and acoustic model characteristics, while another factor is the degree of heterogeneity in signal processing characteristics used by the implant subjects. As noted, most studies that have made a direct comparison between implant users and acoustic models have not applied the same signal processing parameters across all implant users and acoustic models. It is therefore useful to con-

sider the extent to which correspondence between cochlear implant user and acoustic model performance can be attributed to the *specific* set of cochlear implant processing parameters (e.g., pre-emphasis, strategy type, type of filter bank, channel number, and channel stimulation rate) as opposed to the *general* properties of cochlear implant processing shared across device types (e.g., envelope extraction and relatively small number of channels). An assumption behind much previous work using acoustic models of cochlear implant user performance is that a range of different models may account for cochlear implant user performance so long as those models have these general properties of cochlear implant processing that are perceptually important, e.g., the relatively small number of channels and the absence of temporal fine structure within channels. Most acoustic model studies to date have used fixed-channel models, whereas most recipients of this device use peak-picking strategies rather than fixed-channel strategies, and few studies have applied acoustic model methodology to peak-picking strategies (but see Loizou *et al.*, 2000a). Most cochlear implant systems are able to implement both fixed-channel and peak-picking (or "n-of-m") strategies, in which only the highest amplitude analysis channels are coded to corresponding electrode channels during any analysis cycle. Moreover, most recipients of the Nucleus 24 device, which is the focus of the present study, use a peak-picking strategy rather than a fixed-channel strategy, and few studies have applied acoustic model methodology to peak-picking strategies (but see Dorman *et al.*, 2002; Fu *et al.*, 1998; Throckmorton and Collins, 2002). Fu *et al.* (1998) took care to use similar signal processing in cochlear implant users and corresponding acoustic models and also used both a fixed-channel and peak-picking strategy; however, the number of implant users was small ($N=3$), and the acoustic model did not explicitly attempt to model channel interaction/overlap separately from channel number.

The objective of the present study was to assess consonant recognition in both quiet and noise as a function of peak/channel number in a group of cochlear implant listeners using a peak-picking strategy. The inclusion of a noise-contaminated condition was motivated by the need to better understand how particular features might be affected by noise in cochlear implants. The present study extends the general approach of Fu *et al.* (1998) by ensuring homogeneity of processing characteristics among a larger number of cochlear implant subjects and by comparing implant user performance with acoustic models that differed by degree of simulated channel interaction or overlap, in a similar vein to Fu and Nogaki (2005) but here based on the physiologically motivated model of Laneau *et al.* (2006). In contrast to other published studies comparing consonant recognition in implant users and acoustic models, in the present study the cochlear implant subjects all used the same device (Nucleus 24), the same set of processing parameters, e.g., the same strategy type, filter bank method, frequency allocation to electrodes, and pulsatile stimulation rate. In order to provide variation in channel number, both number of peaks selected and total channel number were co-varied to provide large changes in the spectral information provided by the processors. In order to provide variation in stimulus noise conditions, testing was undertaken in quiet and in background speech-shaped noise at +10 dB signal-to-noise ratio (SNR).

The aims of the cochlear implant user experiment (described in Sec. II) were therefore to determine if noise or channel number would have an effect on (the pattern of) consonant feature recognition and to determine if there was any variation in the pattern of consonant feature recognition across users, either in absolute terms or as a function of noise or channel number. The aim of experiment 2 (described in Sec. III) was to attempt to model implant user performance by assessing responses in normal hearing subjects listening to a "carefully matched" noise-band acoustic model that applied the same spectral processing characteristics as used by the cochlear implant recipients. A further advantage of using the model was that the bandwidth of the noise carriers was altered to approximate varying degrees of spectral channel interaction, as per Laneau *et al.* (2006). The overall aims of comparing results from the two experiments were therefore to determine if an acoustic model could predict the patterns of consonant feature transmission in the implant users, to determine whether the incorporation of simulated spectral channel interaction could improve the predictive power of the model, and to assess whether any inter-user differences could be accounted for by differences in channel interaction in the model.

II. EXPERIMENT 1. CONSONANT RECOGNITION IN QUIET AND NOISE IN COCHLEAR IMPLANT USERS AS A FUNCTION OF MAXIMA/CHANNEL NUMBER

A. Methods

1. Test material

For both experiments, stimuli consisted of nonsense syllables recorded from an adult female speaker speaking with a Standard Southern British accent, in the form /iCi/ where one of 20 English consonants is preceded and followed by the vowel /i/. Choice of this vowel environment was motivated by the desire to avoid ceiling effects and because this vowel environment may be more sensitive to cochlear implant parameter effects. Jiang *et al.* (2006) found that the /iCi/ vowel environment yielded a larger effect of background stationary noise on stop consonant recognition than for /aCa/ in normal hearing listeners. Loizou *et al.* (2000b) showed that consonant recognition in the /iCi/ vowel environment was more sensitive to stimulation rate in cochlear implant users than with the /aCa/ vowel environment, and, more generally, performance was poorer than that for /aCa/ where ceiling effects were obtained in some conditions. This was also confirmed by preliminary experimental work using acoustic model data, which indicated worse overall consonant recognition of around 10% with the /iCi/ format compared to the /aCa/ format, with the latter yielding some ceiling effects in quiet. The 20 consonants, along with the phonological feature matrix used in the feature transmission analysis, are shown in Table I. Where appropriate, speech-shaped noise [white noise filtered to have the same long-term average spectrum as the Bamford–Kowal–Bench (BKB) sentences (Bench *et al.*,

TABLE I. Consonants used in the recognition task, with feature matrix used for confusion matrix analysis.

Consonant/feature	Voicing	Fricative	Nasal	Place
b	Yes	No	No	bil
d	Yes	No	No	alv
ʤ	Yes	Yes	No	ret
ʃ	No	Yes	No	ret
f	No	Yes	No	lad
g	Yes	No	No	vel
j	Yes	No	No	alv
k	No	No	No	vel
l	Yes	No	No	alv
m	Yes	No	Yes	bil
n	Yes	No	Yes	alv
e	No	Yes	No	den
p	No	No	No	bil
r	Yes	No	No	ret
s	No	Yes	No	alv
t	No	No	No	alv
ʧ	No	Yes	No	ret
v	Yes	No	No	lad
w	Yes	No	No	bil
z	Yes	Yes	No	alv

1979) spoken by an adult female speaker] was mixed with the speech stimuli at the appropriate SNR for noise-contaminated listening conditions.

2. Subjects

Subjects for experiment 1 were nine post-lingually deafened adult users of the Nucleus CI24M® or Nucleus CI24R Contour® cochlear implant devices, with age range of 25–85 (mean age of 61), of whom six were male. All had experience of using the Advanced Combination Encoder or ACE® (Holden *et al.*, 2002) strategy for at least 3 months and had been cochlear implant users for at least 9 months. All were users of the Esprit 3G® speech processor. Subjects had all obtained more than 50% in the BKB sentence recognition test in quiet at their most recent review session. Six were monaural implantees, while three subjects normally used bilateral implants; for the bilateral implantees, testing was undertaken using the first-implanted side (which in all cases was the better-performing ear when tested in isolation using the BKB sentence recognition test). Payment was made for travel expenses but not for participation in the experiment *per se*. All subjects gave signed informed consent and the study protocol was approved by the Institute of Sound and Vibration Human Experimentation Safety and Ethics Committee (University of Southampton). Further subject details are given in Table II.

3. Procedure

All nine subjects were tested in a sound field within a sound-treated audiometric booth while using their normal speech processor. Each subject was tested while using two different "MAPs," MAP is the term coined by Cochlear Corporation® for the implant program containing parameters including electrical comfort and threshold values, stimulation rate, channel and/or maxima number, etc., and is adopted here. All MAPs were implemented with the ACE® (see Skinner *et al.*, 2002) processing strategy, one with 12 maxima from 20 total channels (12-of-20) or "maximum channel" condition and the other with 4 maxima out of 7 total channels (4-of-7) or "reduced channel" condition. For the reduced channel condition, a MAP was created using electrode channels 1, 4, 7, 10, 13, 16, and 19, leaving all other MAP parameters such as stimulation rate and electrical comfort/threshold levels unaltered compared to the maximum channel condition. Pulsatile stimulation rate was maintained at 900 pps/ch (pulses per second per channel) in all

TABLE II. Subject details for cochlear implant user experiment. BKB score=most recent number of words correct obtained in quiet, listening only.

Subject number	Sex	Age	BKB score	Duration of implant use	Implant type	Normal strategy	Normal MAP parameters	Normal listening mode
1	M	25	81	1 yr 5 m	CI24R	ACE	900 pps 12 of 20	Unilateral
2	M	70	92	$2\frac{1}{2}$ yr	CI24R	ACE	900 pps 12 of 20	Unilateral
3	M	65	90	1 yr	CI24R	ACE	900 pps 12 of 20	Unilateral
4	F	73	94	L–6yr R–4 yr	CI24M	SPEAK	250 pps 8 of 20	Bilateral
5	M	85	57	R–7 yr L–$3\frac{1}{2}$ yr	CI24M	ACE	720 pps 8 of 20	Bilateral
6	F	62	80	2 yr	CI24R	ACE	900 pps 12 of 20	Unilateral
7	M	49	98	2 yr	CI24R	ACE	900 pps 12 of 19	Unilateral
8	M	72	94	L–6 yr R–3 yr	CI24M	SPEAK	250 pps 8 of 20	Bilateral
9	F	48	100	1 yr	CI24R	ACE	900 pps 12 of 20	Unilateral

cases to ensure that this was not a confounding variable across subjects. Frequency allocation to electrode channels for the two channel conditions was determined by standard Cochlear look-up tables (these values are shown in Figs. 5 and 6 in the test material section for experiment 2—it should be noted that these values were re-created for the acoustic models used in that experiment).

Order of MAP condition was alternated between subjects, and testing was conducted first in quiet then in noise for each MAP condition. For three subjects it was necessary to create new MAPs to standardize channel/maxima number, strategy type, and stimulation rate across all the subjects. Re-mapping was undertaken using the Cochlear Custom-Sound® programming software by the author. This was necessary in order to re-set electrical comfort and threshold levels to take account of loudness differences associated with changes in stimulation rate in order to ensure that loudness was as much as possible not a confounding factor between subjects.

No formal attempt was made to choose "better" and "worse" performers *a priori*. Two subjects normally used the spectral peak (SPEAK) processing strategy (Skinner *et al.*, 2002) in the bilateral condition, but both of these subjects had experience using the ACE® strategy since receiving their implants. The three subjects who had bilateral implants performed the tests using only their first implant. Where unfamiliar MAPs were used by subjects (the reduced channel/maxima MAP, changed stimulation rate, or maxima/channel number in three cases, unilateral use in three cases, as noted), acclimatization time of at least half an hour was provided to allow time to get used to novel MAPs. In this context, it should be noted that the *post hoc* separation into worse and better cochlear implant users, described in Sec. II B, did not co-vary with distinctions between those who normally used the 900 pps/ch strategy vs those who did not, nor did it co-vary with those who normally used bilateral cochlear implants vs those who did not. Moreover, subjects all commented that they did not find altered MAPs difficult to get used to compared to normal MAPs, and in all cases subjects had experience with altered parameters (all three bilateral implantees had at least 2 years as monaural implantees prior to bilateral implantations; the SPEAK users had had at least 3 months experience with ACE® use).

Subjects were tested using a 20-alternative forced choice consonant recognition task. Each test session was preceded by a practice run in which each of the 20 consonants was played once. Following preliminary experimental work, it was established that three repetitions per stimulus was the minimum number that should provide adequate data reliability; hence each test run consisted of a series of 60 presentations per condition (three repetitions per stimulus, randomized presentation order). There were two stimulus conditions of quiet and noise (stationary noise at +10 dB SNR, as noted) multiplied by two channel/maxima number conditions ("4-of-7" and "12-of-20"). Subjects were seated in a double-walled sound-treated room, approximately 1.5 m away from a loudspeaker positioned at 0° azimuth. Testing was undertaken with one speech processor only and no contralateral hearing aid. Autosensitivity and other forms of automatic gain control were de-activated in subjects' MAPs. A touch screen was used for visual presentation of response options and to code subjects' responses. Stimulus presentation and data collection were controlled using scripts developed from the PRAAT (Version 4.1) speech analysis toolkit (Boersma and Weenink, 2005). Calibration was undertaken using a speech-shaped noise stimulus that had the same mean rms value as the speech stimuli in order to ensure that speech was presented at a mean presentation level of 70 dB (A-weighted) at the cochlear implant microphone of each subject.

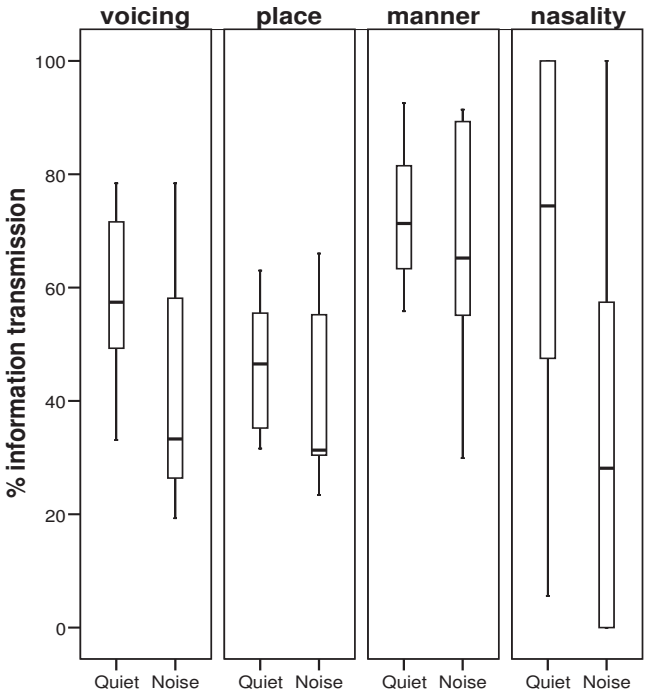


FIG. 1. Percentage information transmitted for consonant features for Nucleus 24 users using the ACE® strategy with 12-of-20 MAP.

Subject responses were tabulated and converted into consonant confusion matrices for each of the four listening conditions (two channel number settings in quiet and noise). Derivation of confusion matrices and subsequent information transmission analysis were undertaken using the software packages FIX and SCORE (www.phon.ucl.ac.uk/resource/software.html). Each confusion matrix was analyzed in terms of information transmission for the features voicing, place, manner, and nasality according to the feature matrix in Table I, as per Miller and Niceley (1955). Feature transmission values were then entered as dependent variables for the purpose of statistical analysis.

B. Results and discussion

Information transmission scores are shown separately for the two channel number conditions in Figs. 1 and 2. As with subsequent figures, feature transmission data are presented in the form of boxplots. These figures suggest differences in transmission between features and a large effect of noise on nasality and possibly an effect for voicing, but probably not other features. Variance was much greater for nasality than for the other features. However, the two figures show

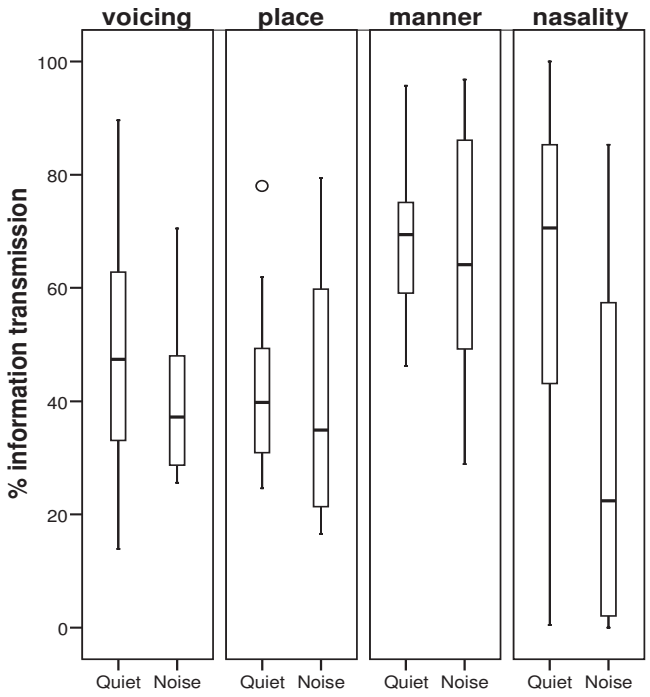


FIG. 2. Percentage information transmitted for consonant features for Nucleus 24 users using the ACE® strategy with 4-of-7 MAP.

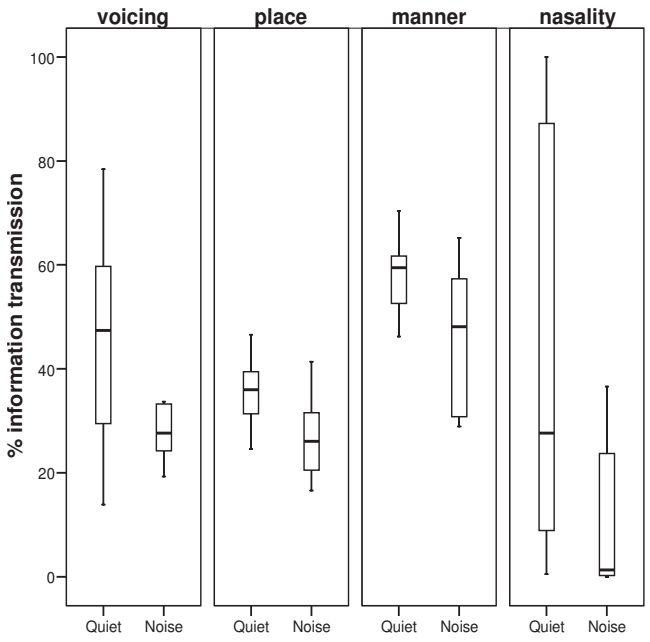


FIG. 3. Percentage information transmitted for consonant features for Nucleus 24 users obtaining less than 50% overall consonant recognition score using the ACE® strategy, averaged across both channel number conditions.

very similar patterns, suggesting little difference between the full and reduced channel number conditions. A multivariate analysis of variance (MANOVA) was performed with four dependent variables (percentage information transmission for the features voicing, place, manner, and nasality) and two within-subject factors with two levels each (channel number and noise). For this and subsequent analyses, only significant factors or interactions are highlighted. The noise factor had a significant effect on nasality $[F(1,32)=2.665;\ p<0.05]$ but no other feature. Interestingly, channel number did not have a significant effect on any feature. There were also no significant factor interactions for any feature.

It was also of interest to consider differences between individual cochlear implant users. In order to do this, it was necessary to define better vs worse performer. This was based on consonant recognition total correct in quiet in the "full-channel number" MAP condition. The mean consonant recognition score was 51% with a range from 37% to 67%. Of the nine subjects, five had baseline consonant recognition scores in quiet with the full-channel number MAP of 50% or more while the other four had scores of less than 50%. Therefore, separate analyses of these two subgroups were undertaken. Figure 3 shows phonological feature transmission for the worse-performing implant users, while Fig. 4 shows feature transmission values for the better-performing subgroup. In both figures, feature transmission values are averaged across channel number conditions. These figures suggest a marked difference in absolute transmission levels, and they also suggest a greater effect of noise interference for the worse-performing subjects. A 2×2 MANOVA, again with channel number and noise as the two factors, was undertaken for each subgroup separately. For the MANOVA using data from the worse-performing subgroup only, noise had a significant effect on the features voicing $[F(1,12)=6.311;\ p<0.05]$ and place $[F(1,12)=5.219;\ p<0.05]$ and just failed to reach significance for the features nasality $[F(1,12)=3.682;\ p=0.08]$ and manner $[F(1,12)=4.550;\ p=0.05]$. There was no significant effect of channel number, nor was there an interaction between the factors. By contrast, for the better-performing subgroup, noise approached signifi-

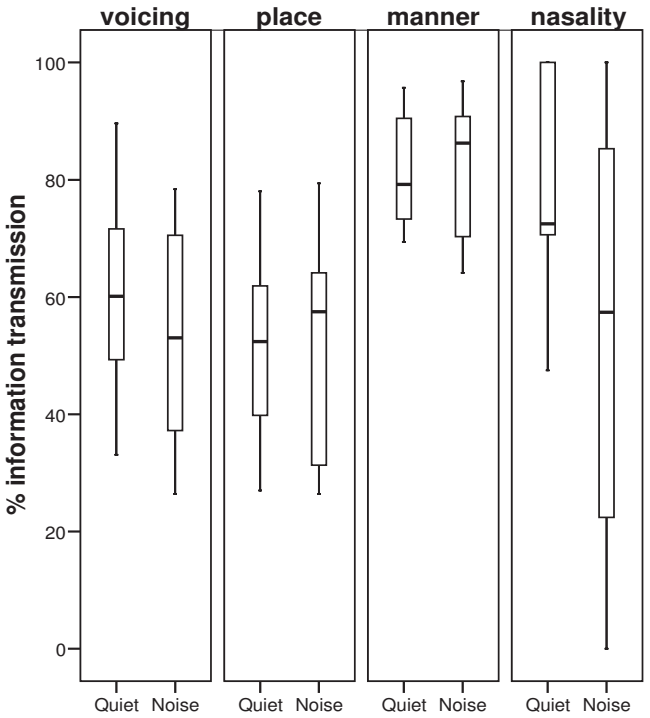


FIG. 4. Percentage information transmitted for consonant features for Nucleus 24 users obtaining greater than 50% overall consonant recognition score using the ACE® strategy, averaged across both channel number conditions.

cance for the nasality feature [$F(1,16)=4.104$; $p=0.06$] but was not significant or even approaching significance for the other three features. As with the worse-performing subgroup, there were no significant effects of channel number and no interaction. Some caution needs to be applied to the statistical interpretation of such small subgroups; nevertheless, the effect appears to suggest a difference in noise susceptibility between the two groups of listeners but no clear difference in pattern of feature transmission in quiet at least.

In order to determine whether the pattern of feature transmission was the same across better and worse users, a repeated measures ANOVA was also undertaken, although it should be noted that only "feature" and its interactions would be expected to show novel results over the MANOVA analyses. Three within-subject factors were entered—noise, feature, and channel number—and one between-subject factor was entered (cochlear implant subgroup, as defined previously). As expected, channel number did not have a significant effect, nor was there any two- or three-way significant interaction between channel number and any of the other factors. Noise had a significant effect overall [$F(1,7)=8.993$; $p<0.05$] and also showed a significant interaction with feature [$F(3,21)=3.945$; $p<0.05$] due to the effect on nasality but not other features. Feature was also significant overall [$F(3,21)=8.026$; $p<0.005$] but did not show any significant interaction apart from the above-mentioned interaction with noise. *Post hoc t*-tests were undertaken to explore the feature factor and its interaction with noise. These showed that manner was significantly better than voicing, place, and nasality [$p<0.05$] but no other differences between features overall, although all differences between place and the other three features approached statistical significance. The between-subject factor "group" was significant overall [$F(1,7)=8.432$; $p<0.05$] but did not show any significant interaction. However, it should be noted that the interaction between groups and presence/absence of noise approached significance [$F(1,7)=5.224$; $p=0.056$], suggesting that noise interference may have been greater for the worse-performing subgroup.

The data appear to support findings from Munson *et al.* (2003) that there are no significant differences in the pattern of feature transmission between better and worse implant users in quiet. However, there was a difference in noise susceptibility in that worse-performing users (defined from their quiet-condition overall consonant recognition scores) showed greater noise effects across a larger range of features. A further interesting point is that an approximately threefold change in channel and maxima number had no impact on any aspect of consonant recognition.

The findings from this experiment led to a number of predictions for testing the validity of an associated acoustic model (described in Sec. III). First, there should be no effect of changing from a 4-of-7 to a 12-of-20 ACE® MAP for any performance measure. Second, data provide a baseline for relative and absolute feature transmission levels; performance in quiet with the stimuli should show poorer place of articulation transmission than manner, voicing, or nasality transmission, but manner transmission should also be better than nasality or voicing. Third, noise interference (at SNR and noise type used here) should affect nasality transmission, not transmission of other features, if the model fits better-user performance, but should affect a range of features if the model is a better predictor of worse-user performance. If the model can predict ceiling performance without assumed spectral channel interaction, this would indicate that performance limitations for better users can be explained by the effects of signal processing information loss. It would also be of interest to determine whether a model with some or greater channel interaction can better estimate performance patterns in worse-performing users. Alternatively, it might be that channel interaction is needed in the model to predict performance in better users.

III. EXPERIMENT 2. EFFECT OF CHANNEL NUMBER AND INTERACTION ON CONSONANT RECOGNITION IN NORMAL HEARING SUBJECTS LISTENING TO ACOUSTIC MODELS OF THE NUCLEUS 24 COCHLEAR IMPLANT

A. Methods

1. Test material

Stimuli used for experiment 2 were pre-processed offline using the MATLAB toolbox, designed to mimic the processing of the Nucleus 24 system in order to replicate as much as possible the signal processing stages used by subjects from the first experiment. Stimuli used as input to the processing were the same recorded nonsense syllables as described in Sec. II. Stimuli were first processed via a pre-emphasis filter with a Hamming window and a FFT size of 8192 (in order to mimic the normal pre-emphasis of the processors used by the implant subjects). The signal was then analyzed using a 128-point FFT, as occurs with the Nucleus Sprint and Freedom processors. This yielded bin center frequencies that were linearly spaced at multiples of 125 Hz. These bins were combined by summing powers to provide 20 or 7 frequency bands (for the two channel number conditions) using the same frequency allocation tables as are implemented in the Nucleus 24 device when one of these two total channel numbers are chosen. The envelope of each filter was calculated as a weighted sum of the corresponding FFT bin powers where the weights determined the frequency boundaries of the bands. Upper and lower frequency bands for both channel number conditions are shown in Figs. 5 and 6. A final stage in the algorithm involved the selection of N channels with greatest amplitude, where N was the number of spectral maxima, or peaks, defined in each of the two channel conditions (12 or 4).

Noise bands were modulated according to the fluctuations in the envelopes of the corresponding filters. In order to model spectral channel interaction, the frequency response of the filters used to generate the noise-band carriers was determined according to the Laneau *et al.* (2006) model and is defined by

$$F(x(f)) = \exp\left(\frac{-\text{abs}(x\text{electrode} - x(f))}{\lambda}\right),$$

where λ = distance along cochlear in millimeters (the conversion of distance on a cochlear into the frequency domain

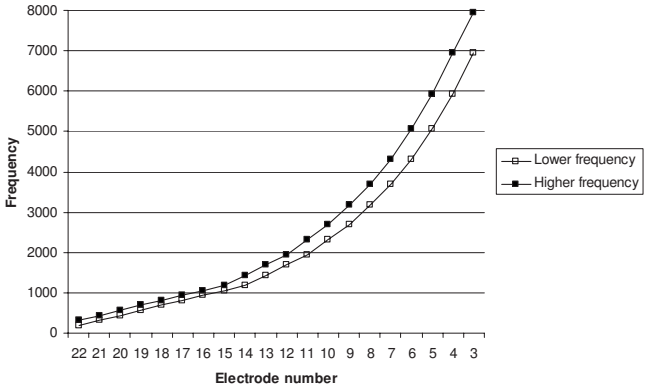


FIG. 5. Upper and lower frequency boundaries for 12-of-20 channel MAPs.

assumed the Greenwood formula), *x*electrode=the position of the simulated electrode, $x(f)$ implements the conversion to distance along the cochlea from frequency according to Greenwood (1990): Equation 1. Filter transfer function used to model spectral channel interaction from Laneau *et al.* (2006).

The model assumed a 35 mm cochlear length and 25 mm electrode array insertion. The desired frequency response was obtained by implementing a linear phase finite impulse response filter in MATLAB. Three channel overlap (interaction) conditions were chosen: first, no overlap between noise-band carriers; second, overlap equivalent to 1 mm spectral spread; and, finally, overlap equivalent to 3.3 mm spectral spread. Therefore, the three models were identical except for the setting of λ and therefore the overlap between adjacent noise-band filters. In order to generate the final stimuli, the noise bands were added together and the rms level of the resulting signal was adjusted to be equal to that of the original signal.

2. Subjects

Eleven normal hearing listeners, students and staff at the University of Southampton, all aged between 18 and 25, were recruited to the study. All were checked via screening audiometry and found to have hearing within normal limits across octave frequencies from 250 to 8000 Hz. Otoscopy was undertaken on all subjects to check for external or middle ear abnormalities and to ensure that insert earphone presentation was possible.

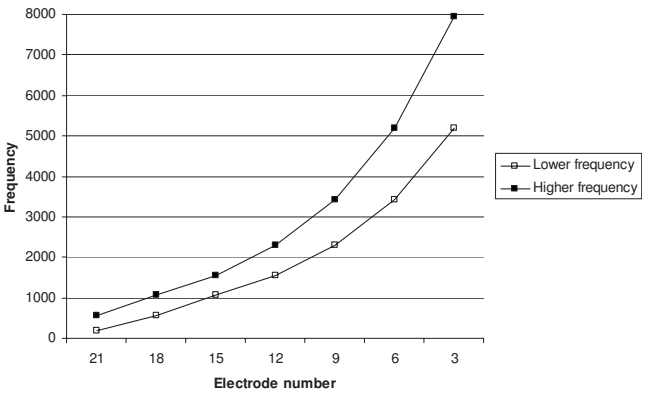


FIG. 6. Upper and lower frequency boundaries for 4-of-7 channel MAP.

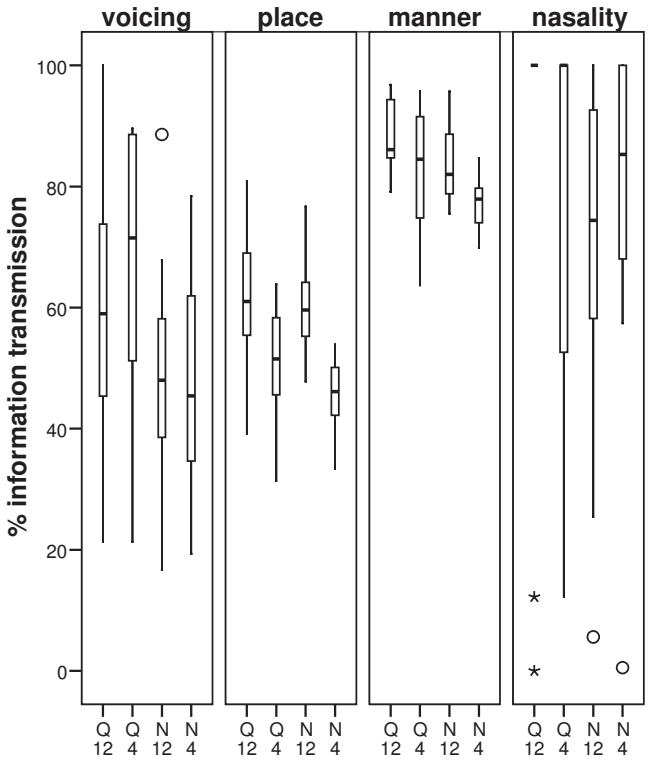


FIG. 7. Percentage information transmitted for consonant features with the no channel interaction acoustic model. Key to conditions: Q12=quiet, 12-of-20 MAP; Q4=quiet, 4-of-7 MAP; N12=noise, 12-of-20 MAP; N4=noise, 4-of-7 MAP.

3. Procedure

The same testing regime was used as described in Sec. II, e.g., a 20-alternative forced choice task. Stimuli were routed through an INKEL MX-880E stereo mixer that delivered a mono signal to an ER-3A insert earphone worn by the subjects. Insert earphone presentation was used for acoustic model experiments in order to minimize the effect of the outer ear transfer function on the stimuli. Monaural presentation was used as this mimics the normal listening condition for the majority of cochlear implant users, and the ear to which stimuli were presented was alternated between subjects. Stimuli were presented to subjects via routing of the signal into an adjacent soundproof booth. As with experiment 1, a touch screen was used for visual presentation of response options and to code subjects' responses. Mean presentation level was adjusted to a mean rms value of 65 dB across conditions (A-weighted).

B. Results and discussion

As with experiment 1, consonant confusion matrices were derived for each listening condition, and information transmission values were then derived for the four features voicing, place, manner, and nasality. Feature transmission values as a function of channel number and presence/absence of noise are shown for the three channel interaction models in Figs. 7–9. These figures suggest that differences between channel interaction models are small, as are differences between channel number conditions. In order to determine the effects of noise and channel number across the different

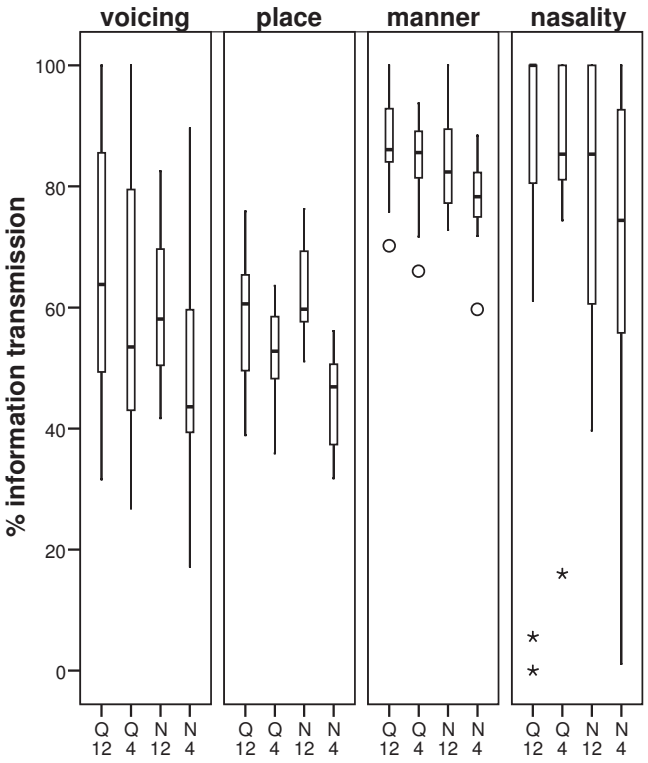


FIG. 8. Percentage information transmitted for consonant features with the 1 mm channel interaction acoustic model. Key to conditions: Q12=quiet, 12-of-20 MAP; Q4=quiet, 4-of-7 MAP; N12=noise, 12-of-20 MAP; N4=noise, 4-of-7 MAP.

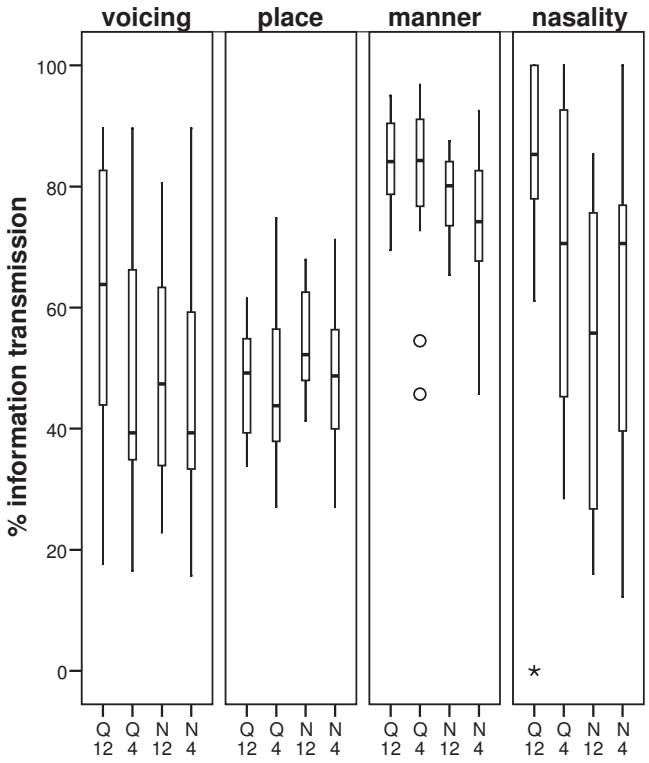


FIG. 9. Percentage information transmitted for consonant features with the 3 mm channel interaction acoustic model. Key to conditions: Q12=quiet, 12-of-20 MAP; Q4=quiet, 4-of-7 MAP; N12=noise, 12-of-20 MAP; N4=noise, 4-of-7 MAP.

acoustic models, three MANOVAs were undertaken (one for each of the three versions of the model) with two within-subject factors (noise and channel number) and the four features as dependent variables. Again, only statistically significant factors and interactions are described here.

For the "3.3 mm channel interaction" model, the noise factor was found to have a significant effect on nasality only [$F(1,40)=4.540$; $p<0.05$], and there were no effects of channel number on any feature and no interaction between the two factors for any feature. However, for the "1 mm channel interaction" model, the noise factor was not significant for any feature, nor was there any interaction, but there was an effect of channel number on place transmission [$F(1,40)=17.447$; $p<0.05$]. For the "no channel interaction" model, the noise factor was found to have a significant effect on the features voicing [$F(1,40)=4.589$; $p<0.05$] and manner [$F(1,40)=4.550$; $p<0.05$], and channel number also had an effect on place transmission [$F(1,40)=19.613$; $p<0.05$].

This analysis suggests a good fit between implant and model results for the 3.3 mm channel interaction model in that this model correctly predicted the noise effect on nasality and the absence of channel number effects on any feature. However, this does not make clear how well the models fit implant user data with respect to relative and absolute feature transmission. Figures 10–13 therefore show feature-specific comparisons across the three models and two implant user subgroups. These figures suggest that the acoustic models were good predictors of performance for better cochlear implant users overall in terms of absolute magnitude of consonant feature transmission of noise or the presence of channel number effects. Figure 11 also shows the one advantage of the 3.3 mm channel interaction model in that it correctly predicts the magnitude of place transmission for both channel number conditions, whereas the other two models predict better performance in the full-channel condition. In order to provide a clearer quantitative comparison between the results of the two experiments, a series of MANOVAs were also undertaken in which data were combined from each of the two implant user subgroups with data from one of the three acoustic models, yielding six analyses in total. In each case, a group factor was added (with two levels, cochlear implant user vs AM subject), in addition to the "noise" and "channel number" factors. As with previous analyses, the four dependent variables of voicing, place, manner, and nasality transmission were included. For the sake of clarity, only the significance or otherwise of the group factor or its interactions are described here. For all three MANOVAs combining "worse user" and acoustic model data, the group factor was significant for all four features (in each case $p<0.001$) although there were no interactions between the group factors and the other factors. This indicates a significant difference between model and worse-performing implant user data across conditions and suggests that the acoustic models were poor predictors of worse-user performance. By contrast, for all three MANOVAs combining acoustic model data with data from better cochlear implant users, the group factor was not significant for any of the features (the lowest p-value being greater than 0.20). For the no channel interaction and 1 mm channel interaction models, there was a significant inter-

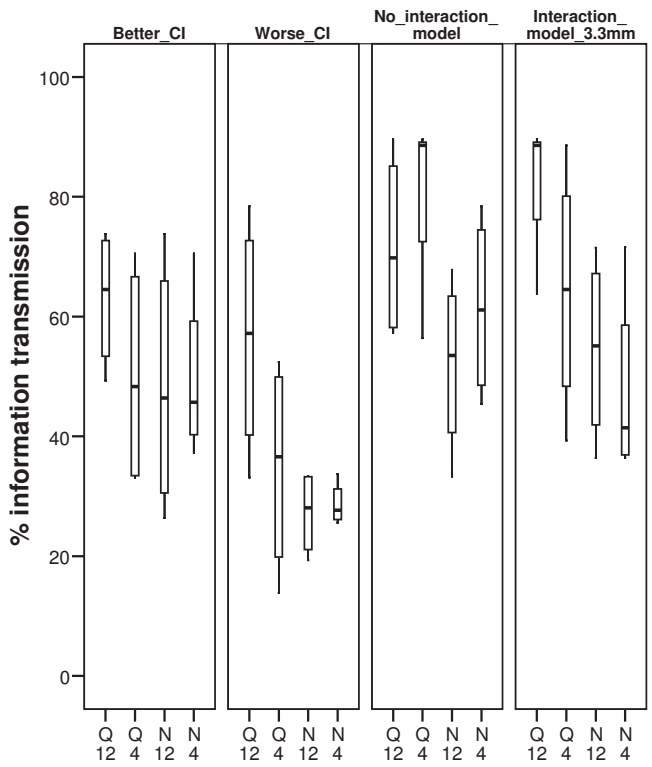


FIG. 10. Percentage information transmitted for the consonant feature "voicing" as a function of noise and channel number, across acoustic models and implant user subgroups. Key to conditions: Q12=quiet, 12-of-20 MAP; Q4=quiet, 4-of-7 MAP; N12=noise, 12-of-20 MAP; N4=noise, 4-of-7 MAP.

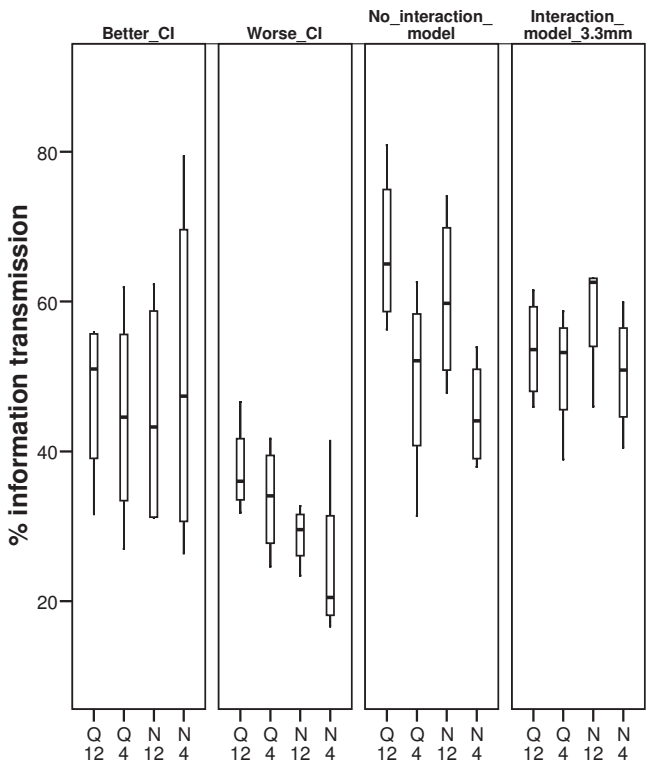


FIG. 11. Percentage information transmitted for the consonant feature "place of articulation" as a function of noise and channel number, across acoustic models and implant user subgroups. Key to conditions: Q12 =quiet, 12-of-20 MAP; Q4=quiet, 4-of-7 MAP; N12=noise, 12-of-20 MAP; N4=noise, 4-of-7 MAP.

action ($p<0.05$) between channel number and group for place, reflecting the fact that these models showed a channel number effect for place, whereas this was not shown for implant users, while there was no interaction between the group factor and any other factor for the "3 mm channel interaction" group.

Overall, it is clear that the acoustic models work well for better implant users. The model with the greatest degree of channel interaction works best in the sense that the predicted pattern of noise and channel number effects among the better implant users were obtained with that model but not the other models. At the same time, it is interesting that the differences between the channel interaction models are small and that the fit between model and better implant user performance does not vary according to the degree of channel interaction (presumably because inter-user variation, even within each subgroup, is greater than variation between the three models). Equally, it is clear that none of the models provides a good fit with data obtained from the worse-performing implant users, either in terms of the (low) absolute feature transmission values or the susceptibility to noise effects shown across features by the worse-performing implant users. Consequently, it cannot be claimed that channel interaction (up to the degree used in the models here) provides a good model of inter-user variation.

IV. DISCUSSION

Deficits in consonant recognition experienced by cochlear implant users may be explained by loss of phonemically relevant acoustic information through signal processing, or they may be explained by other factors, in particular the further loss of information thought to be due to cross-channel spread of excitation. This research sought to differentiate more clearly between these two sources of information loss within the domain of consonant recognition. In order to do this, the pattern of consonant feature recognition as a function of both presence/absence of noise and number of active channels in a peak-picking strategy (ACE®) was first assessed in a group of cochlear implant users. The implant users were as homogenous as possible with respect to device characteristics; e.g., they all used the same device, speech processing strategy, and other signal processing parameters, such as filter bank type, frequency allocation, and stimulation rate; this meant that loss of information due to signal processing was, as much as possible, uniform across the subjects. This homogeneity in terms of processing meant that stronger inferences about the possible reasons for inter-user variation could be made, as these could be assumed to be due to individual differences rather than differences in device signal processing characteristics. A further experiment applied the same perceptual tasks and stimuli to a group of normal hearing subjects listening to a noise-band vocoder model that replicated the signal processing stages employed in the device used by subjects in the first experiment. Spectral overlap between adjacent noise bands was varied in order to mimic the effects of spectral channel interaction, as per Laneau *et al.* (2006). This was done in order to determine whether the inclusion of spectral channel interaction would

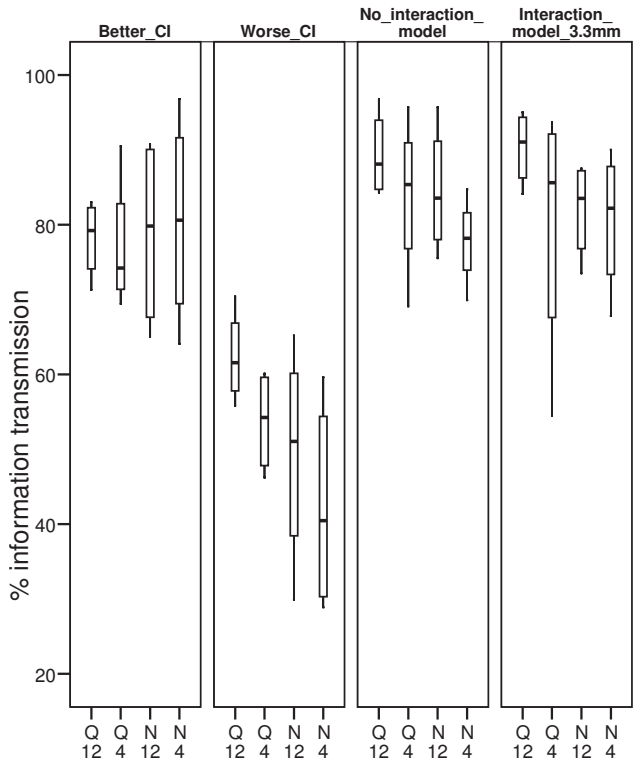


FIG. 12. Percentage information transmitted for the consonant feature "manner" as a function of noise and channel number, across acoustic models and implant user subgroups. Key to conditions: Q12=quiet, 12-of-20 MAP; Q4=quiet, 4-of-7 MAP; N12=noise, 12-of-20 MAP; N4=noise, 4-of-7 MAP.

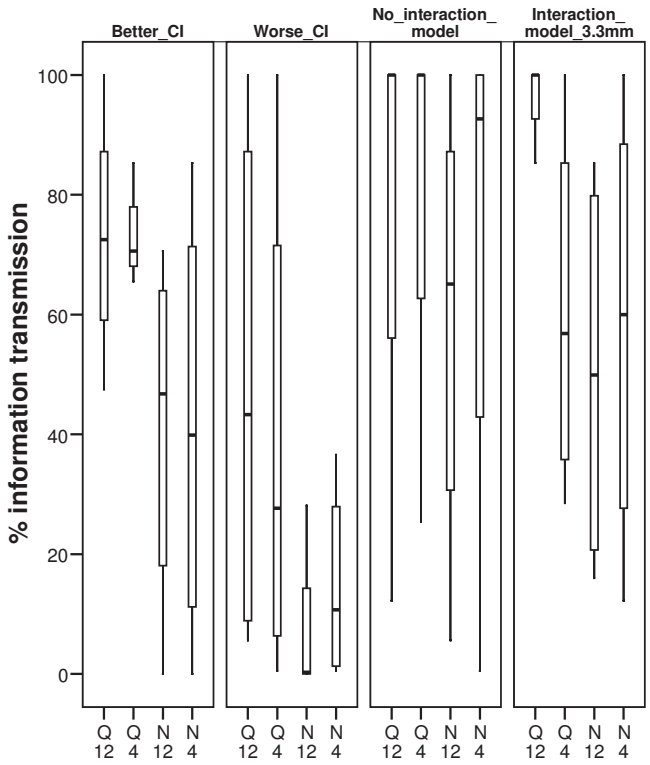


FIG. 13. Percentage information transmitted for the consonant feature "nasality" as a function of noise and channel number, across acoustic models and implant user subgroups. Key to conditions: Q12=quiet, 12-of-20 MAP; Q4=quiet, 4-of-7 MAP; N12=noise, 12-of-20 MAP; N4=noise, 4-of-7 MAP.

improve the match between acoustic model and implant user performance and to see whether this could account for any inter-user differences.

The most important inference from the findings of the two experiments is that channel interaction appears to play a considerably smaller role in explaining consonant recognition deficits in better-performing implant users than loss of information from signal processing. Despite the differences in auditory nerve responses to noise-band vocoder and electrical stimulation, the correspondence between the better users and acoustic model subjects was high with respect to relative and absolute feature transmission values, pattern of noise effects, and the absence of channel number effects. Moreover, variations in simulated spectral channel interaction in the acoustic model had only a modest effect on performance, although there was a slight advantage for the 3.3 mm channel interaction model in terms of matching patterns of implant user performance. The finding that only small differences were obtained across channel overlap conditions in the present study differs from some other noise-band acoustic model studies that have incorporated spectral overlap between adjacent noise bands, although these differences may relate to the type of outcome measure. Laneau *et al.* (2006) provided the most direct comparison with the present study as the authors used the same acoustic model within the context of the same device and processing scheme. The authors found equivalent performance in pitch perception between Nucleus 24 user performance and performance with an acoustic model using simulated channel interaction equivalent to 1 mm spectral spread. In the present study 1 mm spectral spread had almost no effect on performance, though it is likely that pitch discrimination is likely to be more highly correlated with frequency resolution than is the case for consonant recognition. Fu and Nogaki (2005) found that sentence recognition by the best cochlear implant users is approximated by an acoustic model with 8 to 16 channels but only where there was some channel overlap.

Results of the present experiment indicated that consonant place perception can be best modeled in better implant users if some channel interaction is assumed, and it also suggests that better users would improve place transmission by about 10% if no channel interaction were present. In experiment 2, channel number did have an effect on transmission of place (the consonant feature which is most reliant on spectral information), but only with acoustic models with no channel interaction or 1 mm channel interaction. By contrast, the 3.3 mm channel interaction model showed no effect for channel number for any feature, and the same finding was obtained in the cochlear implant users. Moreover, the magnitude of place transmission was accurately estimated by the no channel interaction model. This set of results suggests that channel interaction is implicated in the lack of improvement in place transmission (of around 10%) when changing from 4-of-7 to 12-of-20 channel number condition for the cochlear implant users.

An important inference from the present study is that implant signal processing appears to play a much larger role in limiting place of articulation transmission than does chan-

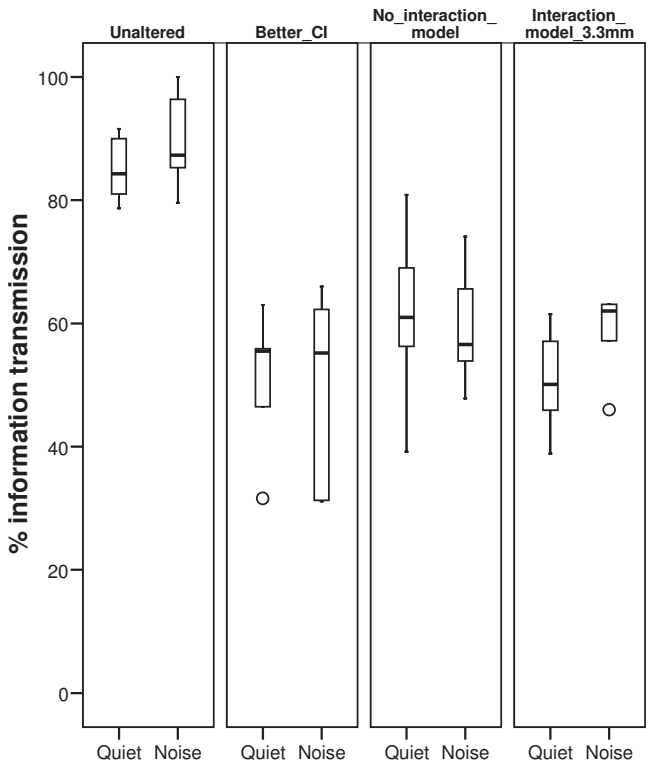


FIG. 14. Percentage information transmitted for the consonant feature place of articulation for unaltered stimuli, for the 12-of-20 acoustic model with and without channel interaction and for better cochlear implant users.

nel interaction. In order to compare these effects, it is useful to compare the deterioration in performance which might be attributed to processing (e.g., by comparing normal performance levels with the best implant or best acoustic model conditions) with the deterioration that might be attributed to channel interaction (by comparing models with and without channel interaction). This comparison is illustrated in Fig. 14, which includes baseline performance with unaltered stimuli. Normative data with unaltered stimuli were obtained from the subjects listening to the acoustic models (although these data were not presented in Sec. III B); transmission of voicing, manner, and nasality all exceeded 97% in quiet or noise, while transmission of place was around 86% in both quiet and noise. Figure 14 shows place transmission in quiet and noise for (normal hearing subjects listening to) unaltered stimuli, better cochlear implant users, the 12-of-20 acoustic model with no channel interaction, and the 12-of-20 acoustic model with 3.3 mm channel interaction. It can be seen that the best acoustic model or implant user performance was still only around or just below 60%, e.g., 25%–30% below normal hearing listeners' abilities, whereas the difference between the acoustic models with or without channel interaction was much smaller in quiet (and non-existent for the noisy condition). This suggests that the effect of removing spectral detail through cochlear implant processing with even the 12-of-20 processing condition is far greater than the effect of adding channel interaction to the acoustic model. Moreover, place transmission differed little between the 4-of-7 and 12-of-20 MAP listening conditions, suggesting that the signal processing used in the present study leads to (consonant place) information loss, which is barely different for the two conditions, despite the presumably substantial differences in spectral details between them.

It has been hypothesized that degree of spectral channel interaction could predict differences between individual cochlear implant users. One corollary of this in terms of consonant feature transmission among implant users would be that the place feature should show the greatest variance between users, as this feature relies on spectral resolution to a greater extent than other features. This was not the case here or in the study by Munson *et al.* (2003). The other corollary would be that an acoustic model with more simulated spectral channel interaction would mimic performance in worse users; this was clearly not the case, as the acoustic model in the present study failed as a predictor of inter-user variance; e.g., performance differences between better and worse users were certainly not mirrored by performance differences in acoustic models with differing degrees of channel interaction. This does not preclude the possibility that degree of channel interaction could be a predictor of performance variation, only that the maximum degree of simulated channel interaction used in the present study was insufficient to produce substantive deterioration in performance. Moreover, channel interaction is almost certainly more complex than accounted for by the similar model of spectral band overlap here (see Throckmorton and Collins, 2002). It could therefore be worthwhile to repeat the experiment with greater degrees of, and more sophisticated models of, simulated channel interaction than modeled here.

The transmission of nasality information showed a number of interesting patterns across implant users and acoustic models that merit particular mention and analysis. Given the paucity of evidence concerning the effect of noise interference on transmission of particular phonemic errors in cochlear implant users, it is interesting to consider why this particular feature was so susceptible to a degree and type of noise that had no effect on performance for better users or acoustic models. Examination of consonant confusion patterns showed that the increased nasality errors were due to increased confusions between the nasal consonants /m/ and /n/ with semivowels /l/ and /w/. For normal hearing listeners, nasality is associated with a low frequency resonance or "murmur," with low intensity slow-changing formant transitions and with spectral zeroes (Fujimura, 1962). Semivowels have the slow-changing formant transitions in common but not the other features of nasal consonants. It is tempting to assume that at least some of these cues should be lost through standard implant processing, although the high levels of nasality transmission in quiet by better implant users and acoustic models suggest that enough information is conveyed by the implant to allow successful transmission in quiet. Figures 15 and 16 show visual representations of electrode output for the nonsense syllables /ini/ and /ili/ (both having approximately the same place of articulation). Electrode number is shown on the *y*-axis, with the most apical (low frequency) electrode having the lowest number, while amplitude (corresponding to current level in implant users or noise-band level in the acoustic models) is represented as height within each channel. The figures were generated using

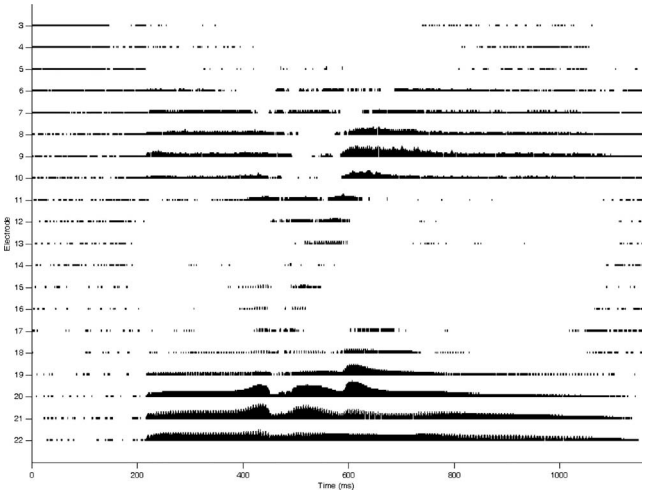


FIG. 15. Electrode activation pattern for the nonsense syllable /ili/.

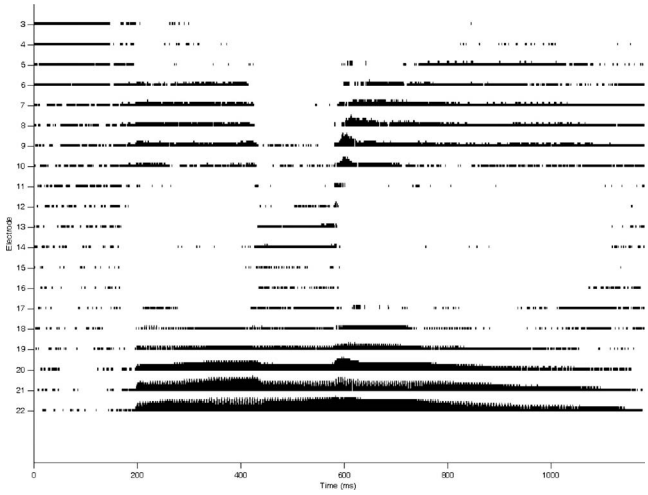


FIG. 16. Electrode activation pattern for the nonsense syllable /ini/.

the same MATLAB software as used to generate the acoustic simulations in experiment 2 using the ACE® parameters for the 12-of-20 condition. It can be seen that the pattern of activation in apical electrodes 20–22 differentiates the two stimuli, with the semivowel having significant modulations in energy between about 400 and 600 ms and the nasal having consistently lower energy than adjacent vowels without modulation. It would appear that the energy in speech-shaped noise in low frequencies is sufficient to obscure these differences in modulation patterns in apical electrodes. It is also worth noting that although no statistical differences were found between better- and worse-performing implant users in terms of patterns of feature transmission, there were some differences that approached significance (perhaps because the study was under-powered to show inter-user variation adequately). The smallest differences between better and worse users were found for voicing (around 15%), while much greater differences were found for nasality (35%). Interestingly, there was a high degree of variance associated with nasality transmission across implant users and normal hearing listeners. Given the evidence from Figs. 15 and 16, it may be that variation in temporal envelope modulation perception plays a greater role in determining these performance variations than spectral resolution. This suggests a different source of inter-user variation from spectral channel interaction, although precisely which aspect of nasality is most important (e.g., temporal resolution, amplitude resolution, or perception within apical channels more generally) requires further investigation. It is also worth noting that the worse-performing cochlear implant users were more susceptible to noise interference across all feature types than better-performing users. A parsimonious interpretation of the data could simply be that worse-performing users were at the less favorable end of their individual SNR functions and that a more favorable SNR (e.g., +15 dB SNR) would be needed for those individuals to tease out differences between features. Nevertheless, it is striking that those implant users with poorer baseline performance in quiet should also show greater susceptibility to noise interference.

V. SUMMARY AND CONCLUSION

The research presented in this paper provides some evidence for the idea that cochlear implant user performance, at least in the limited domain of consonant recognition, can be modeled with reasonable accuracy by using a carefully matched acoustic model in which great care is taken to match processing parameters to those used by the cochlear implant subjects and where cochlear implant users accessed homogenous signal processing characteristics. However, the level of accuracy of the model was high for better-performing cochlear implant users and poor for worse-performing cochlear implant users. Deficits in consonant recognition in better-performing cochlear implant users could be attributed primarily to information loss caused by cochlear implant processing, with channel interaction playing a markedly smaller role. Consonant recognition in worse-performing cochlear implant users was markedly worse than that predicted by a model with the highest degree of channel interaction used in the study. This suggests that either a greater degree of channel interaction is needed to model performance in these cases or that some other factor is playing a role in determining why some users are particularly poor at consonant recognition, especially in noise. The finding of worse nasality transmission and worse performance in noise more generally for the cochlear implant users who had poorer overall consonant recognition in quiet suggests that variations in temporal resolution, particularly in apical channels, could be at least as important as spectral channel interaction in explaining performance variations. However, further work is needed to compare performance against higher degrees of channel interaction and also to explore the precise reasons for the greater degree of variation in nasality perception and the marked susceptibility of nasality perception to noise interference associated with cochlear implant processing. Finally, the extent to which device-specific characteristics, including the use of a peak-picking strategy, rather than a fixed-channel strategy might have impacted on results could be assessed by applying a similar methodology to use of other devices. The present study does suggest that consonant recognition within the context of the particular speech processing strategy and device used (ACE strategy with the

Nucleus 24 device) is actually rather robust to interference with stationary noise, although this finding cannot be generalized to more negative SNRs or noise types other than that used here. By comparing data obtained from a set of cochlear implant users who all accessed identical signal processing characteristics and by comparing those data with equivalent acoustic model data in which the same set of processing parameters is implemented, it is possible to make strong inferences about the relative contribution of different factors in determining deficits in speech perception abilities experienced by cochlear implant users. The use of a detailed phonological analysis of consonant confusions can also reveal perceptual abnormalities that cannot be analyzed using more generic or global speech perception measures.

ACKNOWLEDGMENTS

The author wishes to thank Cochlear Corporation, in particular Brett Swanson, for providing the MATLAB toolbox. Thanks also go to Johan Laneau and Michael Buechler, who developed the additional "AMO" files for generation of acoustic simulations. The author would like to thank the staff and patients at the South of England Cochlear Implant Centre for subject recruitment. The author also thanks Bethan Lewis for assistance with data collection.

Bench, J., Kowal, A., and Bamford, J. (**1979**). "The BKB (Bamford-Kowal-Bench) sentence lists for partially-hearing children," Br. J. Audiol. **13**, 108–112.

Black, R. C., and Clark, G. M. (**1980**). "Differential electrical excitation of the auditory nerve," J. Acoust. Soc. Am. **67**, 868–874.

Boersma, P., and Weenink, D. (**2005**). "Praat: Doing phonetics by computer," Version 4.4 (computer program).http://www.praat.org/ (last viewed 1/29/09).

Donaldson, G. S., and Nelson, D. A. (**2000**). "Place-pitch sensitivity and its relation to consonant recognition by cochlear implant listeners using the MPEAK and SPEAK speech processing strategies," J. Acoust. Soc. Am. **107**, 1645–1658.

Dorman, M. F., and Loizou, P. C. (**1997**). "Speech intelligibility as a function of the number of channels of stimulation for normal-hearing listeners and patients with cochlear implants," Am. J. Otol. **18**, S113–S114.

Dorman, M. F., and Loizou, P. C. (**1998**). "The identification of consonants and vowels by cochlear implant patients using a 6-channel continuous interleaved sampling processor and by normal-hearing subjects using simulations of processors with two to nine channels," Ear Hear. **19**, 162–166.

Dorman, M. F., Loizou, P. C., and Fitzke, J. (**1998**). "The identification of speech in noise by cochlear implant patients and normal-hearing listeners using 6-channel signal processors," Ear Hear. **19**, 481–484.

Dorman, M. F., Loizou, P. C., and Rainey, D. (**1997**). "Speech intelligibility as a function of the number of channels of stimulation for signal processors using sine-wave and noise-band outputs," J. Acoust. Soc. Am. **102**, 2403–2411.

Dorman, M. F., Loizou, P. C., Spahr, A. J., and Maloff, E. (**2002**). "A comparison of the speech understanding provided by acoustic models of fixed-channel and channel-picking signal processors for cochlear implants," J. Speech Lang. Hear. Res. **45**, 783–788.

Friesen, L. M., Shannon, R. V., Baskent, D., and Wang, X. (**2001**). "Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants," J. Acoust. Soc. Am. **110**, 1150–1163.

Fu, Q.-J., and Nogaki, G. (**2005**). "Noise susceptibility of cochlear implant users: The role of spectral resolution and smearing," J. Assoc. Res. Otolaryngol. **6**, 19–27.

Fu, Q. J., Shannon, R. V., and Wang, X. (**1998**). "Effects of noise and spectral resolution on vowel and consonant recognition: Acoustic and electric hearing," J. Acoust. Soc. Am. **104**, 3586–3596.

Fujimura, O. (**1962**). "Analysis of nasal consonants," J. Acoust. Soc. Am. **34**, 1865–1875.

Greenwood, D. D. (**1990**). "A cochlear frequency-position function for several species—29 years later," J. Acoust. Soc. Am. **87**, 2592–2605.

Holden, L. K., Skinner, M. W., Holden, T. A., and Demorest, M. E. (**2002**). "Effects of stimulation rate with the Nucleus 24 ACE speech coding strategy," Ear Hear. **23**, 463–476.

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.

Laneau, J., Moonen, M., and Wouters, J. (**2006**). "Factors affecting the use of noise-band vocoders as acoustic models for pitch perception in cochlear implants," J. Acoust. Soc. Am. **119**, 491–506.

Laneau, J., Wouters, J., and Moonen, M. (**2004**). "Relative contributions of temporal and place pitch cues to fundamental frequency discrimination in cochlear implantees," J. Acoust. Soc. Am. **116**, 3606–3619.

Loizou, P. C., Dorman, M, and Tu, Z. (**1999**). "On the number of channels needed to understand speech," J. Acoust. Soc. Am. **106**, 2097–2103.

Loizou, P. C., Dorman, M. F., Tu, Z., and Fitzke, J. (**2000a**). "Recognition of sentences in noise by normal-hearing listeners using simulations of speak-type cochlear implant signal processors," Ann. Otol. Rhinol. Laryngol. Suppl. **185**, 67–68.

Loizou, P. C., Poroy, O., and Dorman, M. (**2000b**). "The effect of parametric variations of cochlear implant processors on speech understanding," J. Acoust. Soc. Am. **108**, 790–802.

Meyer, T. A., Frisch, S. A., Pisoni, D. B., Miyamoto, R. T., and Svirsky, M. A. (**2003**). "Modeling open-set spoken word recognition in postlingually deafened adults after cochlear implantation: Some preliminary results with the neighborhood activation model," Otol. Neurotol. **24**, 612–620.

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

Munson, B., Donaldson, G. S., Allen, S. L., Collison, E. A., and Nelson, D. A. (**2003**). "Patterns of phoneme perception errors by listeners with cochlear implants as a function of overall speech perception ability," J. Acoust. Soc. Am. **113**, 925–935.

Shannon, R. V., Zeng, F. G., and Wygonski, J. (**1998**). "Speech recognition with altered spectral distribution of envelope cues," J. Acoust. Soc. Am. **104**, 2467–2476.

Skinner, M. W., Arndt, P. L., and Staller, S. J. (**2002**). "Nucleus 24 advanced encoder conversion study: Performance versus preference," Ear Hear. **23**, 2S–17S.

Stevens, K. N., and Blumstein, S. E. (**1978**). "Invariant cues for place of articulation in stop consonants," J. Acoust. Soc. Am. **64**, 1358–1368.

Throckmorton, C. S., and Collins, L. M. (**1999**). "Investigation of the effects of temporal and spatial interactions on speech-recognition skills in cochlear-implant subjects," J. Acoust. Soc. Am. **105**, 861–873.

Throckmorton, C. S., and Collins, L. M. (**2002**). "The effect of channel interactions on speech recognition in cochlear implant subjects: Predictions from an acoustic model," J. Acoust. Soc. Am. **112**, 285–296.

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.

Verschuur, C. A. (**2005**). "Effect of stimulation rate on speech perception in adult users of the Med-El CIS speech processing strategy," Int. J. Audiol. **44**, 58–63.

The neural processing of masked speech: Evidence for different mechanisms in the left and right temporal lobes

Sophie K. Scott[a)]
Institute of Cognitive Neuroscience, University College London, 17 Queen Square, London WC1N 3AR, United Kingdom

Stuart Rosen
Division of Psychology and Language Sciences, University College London, London WC1E 6BT, United Kingdom

C. Philip Beaman and Josh P. Davis
Department of Psychology, University of Reading, Whiteknights, Reading RG6 6AL, United Kingdom

Richard J. S. Wise
MRC Clinical Sciences Centre, London W12 0NN, United Kingdom

(Received 5 November 2007; revised 3 November 2008; accepted 13 November 2008)

It has been previously demonstrated that extensive activation in the dorsolateral temporal lobes associated with masking a speech target with a speech masker, consistent with the hypothesis that competition for central auditory processes is an important factor in informational masking. Here, masking from speech and two additional maskers derived from the original speech were investigated. One of these is spectrally rotated speech, which is unintelligible and has a similar (inverted) spectrotemporal profile to speech. The authors also controlled for the possibility of "glimpsing" of the target signal during modulated masking sounds by using speech-modulated noise as a masker in a baseline condition. Functional imaging results reveal that masking speech with speech leads to bilateral superior temporal gyrus (STG) activation relative to a speech-in-noise baseline, while masking speech with spectrally rotated speech leads solely to right STG activation relative to the baseline. This result is discussed in terms of hemispheric asymmetries for speech perception, and interpreted as showing that masking effects can arise through two parallel neural systems, in the left and right temporal lobes. This has implications for the competition for resources caused by speech and rotated speech maskers, and may illuminate some of the mechanisms involved in informational masking. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3050255]

PACS number(s): 43.71.Rt, 43.71.Qr [RYL] Pages: 1737–1743

I. INTRODUCTION

The properties of masking sounds affect the extent to which they compete for the same resources—central or peripheral—as the target. The aspects of these properties can be very broadly captured by the terms informational and energetic masking, where in the latter the effects are largely due to competition at the auditory periphery, whereas in the former competition for resources seems to be associated with more central auditory processes. For any particular masking signal, the masking effects typically arise from a combination of energetic and informational factors. For example, while masking speech with steady-state noise will presumably be dominated by energetic effects, masking speech with speech will involve both energetic and informational masking. In this paper we used functional neuroimaging to contrast two different speech-related masking signals. Our aim was to identify any difference in their effects in cortical processing, which could be linked to competition for central auditory processing resources, and thus to aspects of informational masking.

We have previously presented data from a positron emission tomography (PET) study indicating that neural correlates of the functional differences between informational and energetic masking can be distinguished (Scott *et al.*, 2004). Subjects were instructed to listen to a single talker in the presence of either a concurrent, continuous masking noise (energetic masking) or speech from another talker (energetic plus informational masking). Each masker type was presented at four different signal-to-noise ratios (SNRs). For the noise masker, there were level dependent effects in the left ventral prefrontal cortex and supplementary motor area, and level independent effects in the left prefrontal and right posterior parietal cortex. For the speech masker, in contrast, there was a level independent activation extensively through auditory association cortex in regions, lateral, anterior, and posterior to primary auditory cortex. In the left hemisphere, these regions have been previously demonstrated to be important for the early perceptual processing of speech (Jacquemot *et al.*, 2003; Scott and Johnsrude, 2003; Wise *et al.*, 2001), and in the right hemisphere these regions have been

[a)]Author to whom correspondence should be addressed. Electronic mail: sophie.scott@ucl.ac.uk

associated with nonverbal aspects of speech perception (Scott *et al.*, 2000; Patterson *et al.*, 2002). We interpreted such activation as evidence implicating neural systems important in speech processing when speech is masked by speech—perhaps due to competition for perceptual resources, or because some of these regions are important in the representation of multiple sources of acoustic information (Zatorre *et al.*, 2004).

A limitation of our previous study was that unmodulated noise with the same spectrum as speech was used as the energetic masking condition. Continuous noise was selected as the energetic masker as it leads to the greatest levels of masking, but this did mean that neural activation in the speech masking conditions associated with "glimpses" of the target and masking speech could not be distinguished from processes more strongly linked to informational masking generally (Festen and Plomp, 1990). Thus, some of the results seen in auditory cortical fields could have been associated with essentially energetic processes by allowing glimpses of the target.

A second limitation of this study is that the precise nature of the speech masking effects is hard to determine—we are unable to distinguish between the effects of the acoustical or lexical properties of the masking speech. Although it can be hard to specifically draw a line between informational and energetic masking effects, it has been established that the maximal informational masking is achieved when masking a talker with the same talker, which indicates an important role for acoustic properties. There is also some role for lexical information in speech masking effects (Brungart, 2001), since intrusions from masking speech occur at rates higher than those expected by chance.

Either of these mechanisms (acoustic or linguistic processing), as well as glimpses (which are naturally acoustic in nature), could be responsible for the activation seen in our previous study. Functional imaging studies are well positioned to be able to determine the contributions of these different factors to masking by speech. The bilateral temporal lobe systems recruited by speech perception can be fractionated, both in terms of hemispheric asymmetries and along anatomical lines. Functional imaging studies have shown a clear dominance for left superior temporal areas in the processing of linguistic information in speech (Scott *et al.*, 2000; Narain *et al.*, 2003; Jacquemot *et al.*, 2003). In contrast, right superior temporal areas consistently respond to signals with pitch variation, be these in speech or music (Patterson *et al.*, 2002; Scott *et al.*, 2000; Zatorre and Belin 2001). Functional imaging thus has the power to differentiate linguistic from nonlinguistic processing of masking speech. The aim of the current study is to identify the way in which masking speech competes for central auditory processes, and the extent to which this relates to linguistic processes, and to attempt to control for the possibility of glimpses contributing to the effects previously reported.

Several behavioral papers have interrogated aspects of informational masking by using speech and time-reversed speech as maskers (Hawley *et al.*, 2004; Rhebergen *et al.*, 2005; Johnstone and Litovsky, 2006). In this study we use spectrally rotated speech as a comparison masker (Blesser, 1972), since it has several advantages over reversed speech in terms of its acoustic profile (Scott and Wise, 2004). Hence, three different stimuli were used as maskers: speech, rotated speech (Blesser, 1972), and noise with the same long-term spectrum as speech, modulated by the envelope of speech [speech-modulated noise (SMN); Festen and Plomp, 1990]. These stimuli have different acoustic and lexical characteristics, and imaging the processing of these maskers will give some indication about which characteristics are processed in masking signals. Furthermore, we will be able to establish whether the neural systems responsible for processing characteristics of maskers are similar to those already implicated in cortical speech processing (e.g., Scott *et al.*, 2000; Mummery *et al.*, 1999).

The first two maskers—speech and rotated speech—have very similar auditory profiles, although only the speech also has lexical information. Rotated speech has the spectrotemporal complexity of speech, and maintains much of the sense of voice pitch variation, but is unintelligible. Masking from rotated speech would therefore be associated with prelexical, acoustic aspects of the signal. SMN has the same amplitude modulations as the original speech signal but none of the spectral complexity, structure, or sense of pitch. As a masker it thus allows glimpses of the target speech. The use of this as a baseline masking condition allows us not only to contrast speech masking conditions with noise masking conditions but to also control for the possibility of glimpses, periods during which the masker energy is relatively low, so that the target speech is more readily heard (Festen and Plomp, 1990). This ensures that activation detected when contrasting speech-in-speech over speech-in-noise does not arise simply from "glimpsing" during amplitude "dips."

We have two main hypotheses. By contrasting speech-in-speech and speech-in-rotated-speech with speech-in-SMN, we are controlling for glimpses of the target stimuli. If we see cortical activation associated with these speech based maskers, therefore, we can conclude that this is associated with central auditory processing of the masking signal. Our second hypothesis is that there will be differences in the central processing of the speech and rotated speech maskers, with speech being processed bilaterally (as it contains both acoustic and linguistic elements of speech) while rotated speech will be associated with right temporal lobe activation (as it does not contain the linguistic elements of speech).

II. METHODS: STIMULUS PREPARATION

Three different sets of stimuli were constructed: speech-in-speech, speech-in-rotated-speech, and speech-in-SMN. Oscillograms and spectrograms for each masker type are shown in Fig. 1. All stimulus materials were drawn from digital representations (sampled originally at 44.1 kHz) of simple sentences recorded in an anechoic chamber by a male and a female talker of standard Southern British English. The target sentences were always Bamford–Kowal–Bench (BKB) sentences (Foster *et al.*, 1993) spoken by a female whereas maskers were based on the Institute of Hearing Research audio-visual sentences lists spoken by a male (MacLeod and Summerfield, 1987). All sentences were low-pass filtered at

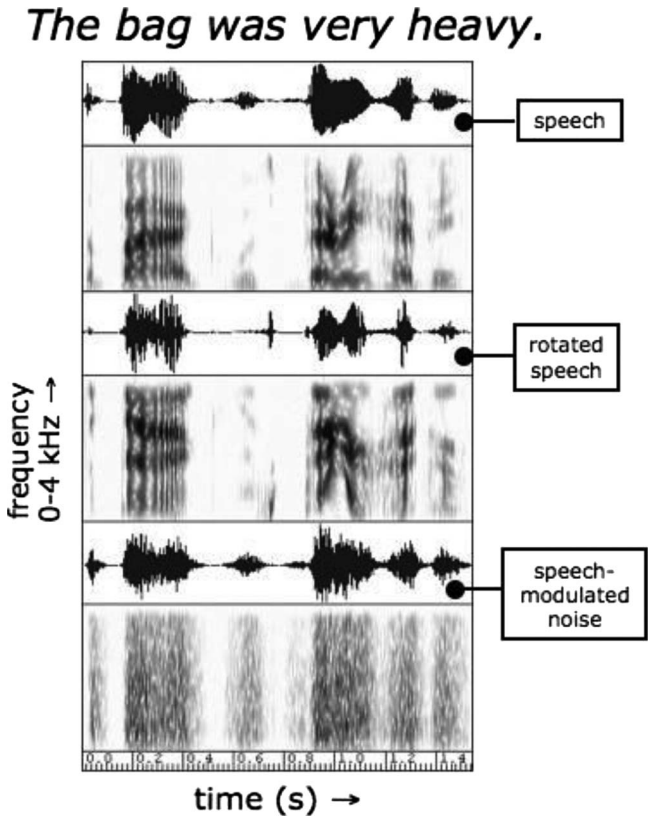


FIG. 1. Oscillograms and spectrograms for the three masking stimuli: speech, rotated speech, and SMN.

3.8 kHz (6th-order elliptical filter, both forward and backward, so as to ensure zero-phase filtering equivalent to a 12th-order filter), and then downsampled to 11.025 kHz to save space. For the speech masker conditions, the masker sentences underwent no further processing. Rotated speech maskers were spectrally inverted using a digital version of the simple modulation technique described by Blesser (1972). In order to preserve somewhat more of the high frequency energy in the original speech signal, here the signals were inverted around 2 kHz, instead of the 1.6 kHz used by Blesser (1972). Because normal and spectrally inverted signals would lead to sounds with very different long-term spectra, the speech signal was first equalized with a filter (essentially high pass) that would make the inverted signal have approximately the same long-term spectrum as the original. This equalizing filter was constructed on the basis of recent extensive measurements of the long-term average spectrum (LTAS) of speech (Byrne *et al.*, 1994), and implemented in finite impulse response (FIR) form. The equalized signal was then amplitude-modulated by a sinusoid at 4 kHz, followed by forward-backward low-pass filtering at 3.8 kHz as described above. The total rms level of the inverted signal was set equal to that of the original low-pass filtered signal.

SMN was created by modulating a speech-shaped noise with envelopes extracted from the original wide-band masker speech signal by full-wave rectification and second-order Butterworth low-pass filtering at 20 Hz. The speech-shaped noise was based on a smoothed version of the LTAS of the male masker sentences. All 270 masker sentences (sampled at 22.05 kHz) were subjected to a spectral analysis using a fast Fourier transform (FFT) of length 512 sample points (23.22 ms), with windows overlapping by 256 points, giving a value for the LTAS at multiples of 43.1 Hz. This spectrum was then smoothed (in the frequency domain) with a 27-point Hamming window that was two-octaves wide, over the frequency range 50 Hz–7 kHz. The smoothed spectrum was then used to construct an amplitude spectrum for an inverse FFT (assuming a sampling rate of 11.025 kHz) with component phases randomized with a uniform distribution over the range $0-2\pi$.

Sentences at different SNRs were created by digital addition, with SNRs determined by a simple rms calculation across the entire waveform. All combined waves were normalized to the same rms value. Because sentences were typically of different durations, summation of the original sentences would have meant that either the target or the masker would have sound energy at its end occurring in a period of silence of its pair (assuming onsets were synchronized). Sentence pairs were thus modified in duration to their mean using the synchronized overlap-and-add (SOLA) technique (Roucus and Wilgus, 1985) as implemented by Huckvale (2007). This alters the duration of speech without changing its fundamental frequency or spectral properties. SOLA cannot, in fact, guarantee any particular final duration, but analysis of the sentences after processing showed them all to fall within a 15 ms range (around a mean of 1.545 s). The shorter sentence in each pair was padded with an appropriate number of zeros before the final addition.

III. BEHAVIORAL TESTING

The intelligibility of the three different masker conditions was assessed in 12 normally hearing adults (ages 26–50, with six men), none of whom subsequently participated in the PET study. Conditions were presented in a randomized order. Sixteen sentences, with a total of 48 key words, were presented per condition. Sentences were presented diotically over headphones and listeners were asked to repeat back the words that they could hear from the female talker. These sentences have a very simple semantic and syntactic structure with three or four key words (e.g., the CLOWN had a FUNNY FACE). Responses were scored in terms of the number of key words correctly repeated. This was done for a range of SNRs for each masker type: 0, −3, and −6 dB SNRs for the SMN masker; −3, −6, and −9 dB SNRs for the speech masker; and −6, −9, and −12 dB SNRs for the rotated speech masker (see Fig. 2). There is a clear effect of masking condition and level on the intelligibility of the sentences. These data were used to select the SNR conditions for the PET scanning in which intelligibility was ~80%: −3 dB SNR for the SMN masker, and −6 dB SNR for the speech and rotated speech masker. Performance across the conditions at these levels was not significantly different when compared in a repeated measures analysis of variance (ANOVA) or with t-tests ($p > 0.05$). These levels were used for every presentation of the specific masking condition in the PET study.

The eight subjects for the PET study were tested prior to scanning. They were played individual BKB sentences and

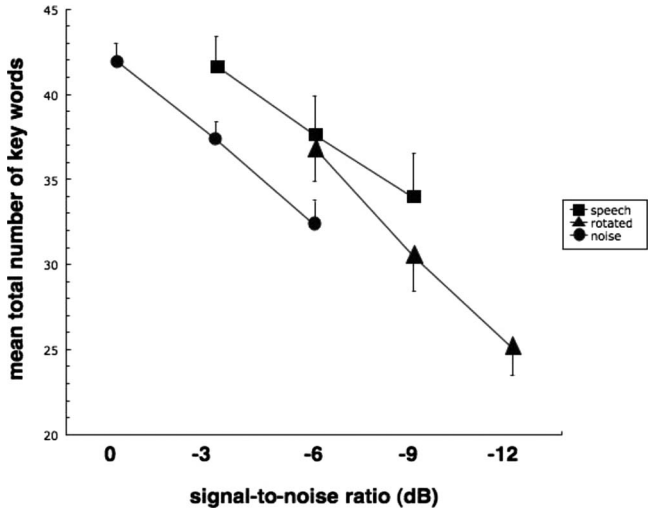


FIG. 2. Intelligibility in three different masking conditions (speech, rotated speech, and SMN), as a function of SNR, from the pilot testing conditions. The error bars show standard errors.

the masking stimuli diotically over headphones and repeated back what they could hear. Sixteen sentences, with a total of 48 key words, were presented per condition (none of which were repeated in the subsequent PET study). Intelligibility was scored by an experimenter who recorded the number of correct key words per condition as a score out of 48. This gave a score for each subject and masking condition. The order of conditions was randomized.

All of the PET subjects were able to perceive speech in the different conditions during prescan training. The average number of key words per condition was 40.4 (SD 2.61) for the speech in masking speech (=84%, with a maximum of 92% and a minimum of 75%), 39.0 (SD 2.82) for speech in masking rotated speech (=81%, with a maximum of 88% and a minimum of 73%), and 37.9 (SD 2.53) for speech-in-SMN (=79%, with a maximum of 88% and a minimum of 73%). A repeated measures ANOVA revealed that performance differed statistically across the three conditions ($F=4.62$, $df=2,7$, and $p=0.027$). *Post hoc* t-tests revealed that this arose from a significant difference between the speech and SMN masking conditions ($p=0.023$), where performance in SMN was poorer. There was no significant difference between performance on the speech-in-speech and speech-in-rotated-speech conditions, or between the speech-in-rotated-speech and speech-in-SMN conditions. The difference in intelligibility between speech-in-speech and speech-in-SMN was just over 5%.

IV. PET SCANNING

Eight right-handed native English-speaking volunteers, none of whom reported any hearing problems, were recruited and scanned. The mean age was 42, with a range 35–57. Each participant gave informed consent prior to participation in the study, which was approved by the Research Ethics Committee of Imperial College School of Medicine/Hammersmith, Queen Charlotte's & Chelsea & Acton Hospitals. Permission to administer radioisotopes was given by the Department of Health (London, UK).

PET scanning was performed with a Siemens HR++ (966) PET scanner operated in high-sensitivity three-dimensional mode. Sixteen scans were performed on each subject, using the oxygen-15-labeled water bolus technique. All subjects were scanned while lying supine in a darkened room with their eyes closed.

The stimuli were presented diotically at a comfortable level determined for each subject, and this level was kept constant over the scanning sessions. The sentence presentations began 15 s before the scanning commenced, and each sentence presented was novel (i.e., there were no repeats). As in our previous study, we used a target female talker and a male masking talker as this enabled us to give the subjects the simple instruction of "listen to the female talker." The subjects were instructed to listen passively to the female talker "for meaning" in the scanning sessions. Passive listening (i.e., with no overt responses) reduces the likelihood that activation seen is due to controlled processing aspects of the task, which would be involved if the subjects were required to make explicit responses or try and remember the sentences they heard (Scott and Wise, 2003). Such requirements have been shown to influence responses in auditory cortex (Brechmann and Scheich, 2005).

V. ANALYSIS

The images were analyzed using statistical parametric mapping (SPM99, Wellcome Department of Cognitive Neurology, http://www.fil.ion.ucl.ac.uk/spm), which allowed manipulation and statistical analysis of the grouped data. All scans from each subject were realigned to eliminate head movements between scans and normalized into a standard stereotactic space [the Montreal Neurological Institute template was used, which is constructed from anatomical magnetic resonance imaging (MRI) scans obtained on 305 normal subjects]. Images were then smoothed using an isotropic 10 mm, full width at half maximum, Gaussian kernel, to allow for variation in gyral anatomy and to improve the SNR.

VI. RESULTS

Three main contrasts were performed, both based on subtractions. In the first, regions more activated by speech-in-speech than speech-in-SMN were identified. This revealed activation confined to the left and right superior temporal gyri (STGs) (Table I), anterior to primary auditory cortex, and extending to the dorsal bank of the STS (Fig. 3). In the second, regions more activated by speech-in-rotated-speech than speech-in-SMN were identified. This revealed activation in the right STG (Table I), anterior to primary auditory cortex, and again extending to the dorsal bank of the STS (Fig. 4). Of the two peaks in this region, one lies within 2 mm in each dimension of the peak response to speech-in-speech, and thus likely represents the same peak of activation, with the spatial resolution available using PET. In the third contrast, regions more activated by the speech-in-speech than speech-in-rotated-speech were identified. This contrast did not reveal any significant activity. Finally, a conjunction analysis of both informational masking conditions

TABLE I. Peak activations for various planned contrasts.

Contrast	Region	Z score	X	Y	Z
Speech-in-speech > speech-in-SMN	Left STG	4.43	−68	−12	0
	Right STG	5.69	66	−2	−6
Speech-in-rotated-speech > speech-in-SMN	Right STG	6.53	58	−10	4
		5.34	64	−2	−4
Conjunction of speech in speech and rotated speech > speech in SMN	Right STG	6.21	64	−2	−4
		6.11	60	−10	2

revealed a peak in the right STG (Table I), which was just 2 mm more medial than the peak response in the speech-in-speech contrast, and essentially therefore reflects the same peak of activation. An additional analysis investigated any overall response to intelligibility, without regard to masker type, by using the subjects' pretesting scores as covariates across all conditions. No regions were significantly activated by this, possibly because the range of intelligibility was reduced in this study, relative to studies that expressly vary intelligibility (across all the subjects, scores ranged from 73% to 92%). In our previous study of masking, intelligibility ranged over a wider range (from 50% to 100%) and significant intelligibility related regions were seen (Scott *et al.*, 2004).

VII. DISCUSSION

Our previous study (Scott *et al.*, 2004) showed extensive bilateral superior temporal activation associated with informational masking of speech: We interpreted this as central perceptual processing of the masking speech signal, consistent with a central competition of resources in informational masking. However, we could not rule out a contribution of glimpses of the target signal in the speech masking condition as a basis of at least some of the activation, nor could we determine the nature of the central resources—acoustic or linguistic—for which there was perceptual competition. The results of the current study allow us to address these issues.

First, the activation in the speech masker condition in this study is less extensive than that seen in the previous study, suggesting that some of the changes in activation in the previous study were indeed a result of glimpses of the target signal allowed by the modulated masker. This seems to primarily affect the activations seen in more posterior audi-

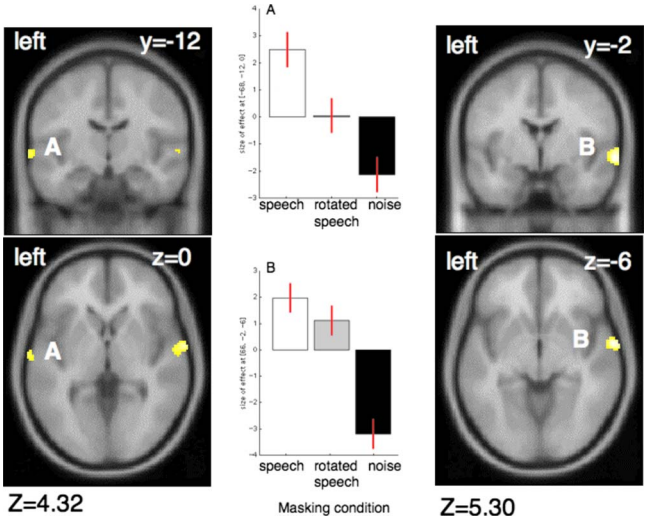


FIG. 3. (Color online) Activation for the contrast of the conditions "speech-in-speech" over the conditions "speech-in-modulated-noise" (analyzed in SPM99, $p<0.0001$, cluster size >40 voxels). This subtraction reveals activations that are significantly greater to the masking speech than to the noise masker. The peak activations in the left and right temporal lobes are projected on the MNI TI template from SPM99: The panels on the left of the figure show the activation peak in the left hemisphere, and the panels on the right show the peak activation in the right hemisphere. The upper panels show the activation on a coronal image of the brain, and the lower panels show the activation on a transverse image. The graphs show the effect sizes as percentage signal change across conditions: While the comparison is of the activity for speech-in-speech > speech-in-noise, the activity in this peak for the speech-in-rotated-speech condition is also shown. Note that in the left temporal lobe the response to speech-in-rotated-speech is reduced relative to the response to speech-in-speech, whereas in the right temporal lobe the responses for both speech-in-speech and speech-in-rotated-speech are more similar.

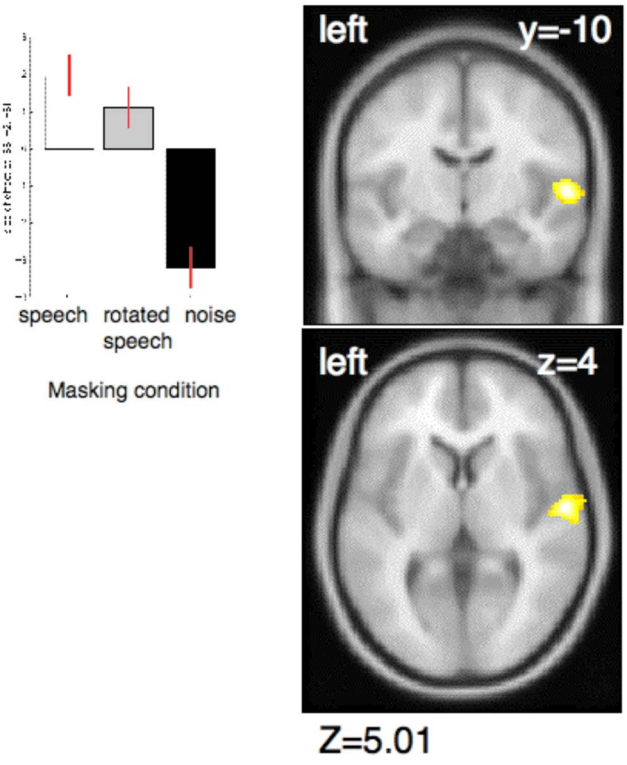


FIG. 4. (Color online) Activation for the contrast of the conditions speech-in-rotated-speech over speech-in-modulated-noise (analyzed in SPM99, $p<0.0001$, cluster size >40 voxels). This subtraction reveals activations that are significantly greater to the masking rotated speech than to the noise masker. The peak activation in the right temporal lobe is projected on the MNI TI template from SPM99. The graph shows the effect size as percentage signal change across conditions. Note that the activation lies posterior and dorsal to the right STG peak for the speech-in-speech > speech-in-modulated-noise contrast: However, there is a subpeak (64, −2, −4, Z=5.34) which lies within 2 mm of this.

tory regions: In the current study there are no peaks farther back than $y=-15$, i.e., lateral to the anterior extent of primary auditory cortex. In our previous study there were bilateral peaks extending as far back as $y=-30$, lateral to the posterior extent of primary auditory cortex. This suggests that activation associated with masking speech in anterior STG regions is not driven solely by glimpses of the target stimulus. Zatorre *et al.* (2004) recently demonstrated, by varying the number of acoustic sources perceived, that the right anterior superior temporal sulcus (STS) is associated with representation of multiple auditory objects. Potentially, it has a role in representing and selecting between competing auditory sources. Our data are consistent with this finding, suggesting a role for these anterior auditory fields in the processing of multiple auditory "streams" of information (Scott, 2005). Since these same regions are strongly implicated in the auditory processing of spoken language *without* maskers (Mummery *et al.*, 1999), these would be prime candidates for a locus of central processing of masking speech.

Second, we are able to distinguish differences between different speech masking conditions, which relate to the nature of competition for central auditory resources in masking. The presence of speech as a masker is associated with bilateral STG responses, whereas rotated speech as a masker results in right STG responses only, relative to SMN. This suggests that masking speech recruits more extensive left temporal lobe neural systems than masking rotated speech, and thus potentially different perceptual processes. Rotated speech is unintelligible, but contains much of the acoustic structure of the original speech signal, in terms of spectrotemporal dynamics, although aspects of the amplitude modulation profile are likely to be different. Importantly, since harmonics are still represented as equally spaced spectral components (although typically no longer strictly harmonic), rotated speech maintains the pitch of speech (albeit with a weaker saliency), and hence preserves the intonation of the original signal. We have previously attributed common activation of the right STG to attended (and unmasked) speech and rotated speech as a result of the presence of pitch variation in both (Scott *et al.*, 2000), and several studies explicitly investigating pitch variation have demonstrated a clear preference in the right STG/STS for such signals (Zatorre and Belin 2001; Patterson *et al.*, 2002). In contrast, functional imaging studies of intelligible speech, which compare the activation produced by speech with that produced by rotated speech, reveal solely *left* STS regions (Scott *et al.*, 2000, 2006; Narain *et al.*, 2003). It seems likely, therefore, that masking speech and rotated masking speech activate right anterior STG because of the acoustic structure and pitch variation that both have in common, while masking speech activates the left STG as a function of its intelligibility—and therefore of its linguistic status. This may have some implications for approaches to informational masking. Studies have indicated that different masking effects can be seen when a talker is masked by speech and when the masker is comprised of reversed speech—e.g., there is a release from masking when reversed speech is used as a masker (Rhebergen *et al.*, 2005), and there is a developmental profile to this, with children showing greater masking from reversed speech than adults (Johnstone and Litovsky, 2006). There can be similar effects of masking speech and masking reversed speech (Hawley *et al.*, 2004), when there are multiple maskers. One interpretation of our data is that the linguistic component of informational masking is associated with left temporal lobe mechanisms, and some of the nonlinguistic aspects of masking from speech and reversed speech may be associated with right temporal lobe mechanisms.

In conclusion, neural activity associated with masking from speech (above masking from SMN) is seen in anterior, rather than posterior, auditory fields, in areas that represent auditory objects. We also suggest that part of the masking effects of speech and rotated speech result from their shared acoustic properties. However, speech as a masker activates the left STG to a far greater degree than rotated speech (Fig. 3) because, we suggest, of its intelligibility. This is consistent with behavioral findings suggesting that informational masking effects arise from competition for both acoustic and linguistic processing resources (Brungart, 2001; Hawley *et al.*, 2004; Rhebergen *et al.*, 2005; Johnstone and Litovsky, 2006). We identify two different types of central competition in the temporal lobes, one linguistically driven in the left anterior auditory association cortex, and one driven by acoustic properties in the right anterior auditory association cortex. Such an account allows for masking by speech to show both lexical and acoustic influences, and places the neural processing of "unattended" speech in the same framework as the neural responses to attended speech (Scott *et al.*, 2000). Attention is known to modulate neural processing of sound (Brechmann and Scheich, 2005) and our results do not contradict this distinction between the two modes of response, rather they indicate that attended and unattended (masking) speech enter the processing system via the same neural pathways. Furthermore, the unattended masking speech is processed within the anteriorly directed "what" pathway of processing, which is associated with the processing of intelligible attended, unmasked speech (Scott *et al.*, 2000). Further work will determine the extent to which the right temporal lobe response to both masking speech and masking rotated speech arises from aspects of their shared acoustic structure or their pitch variation. Future studies will also be able to determine whether the left temporal response to masking speech is driven by phonological or postlexical processing of the unattended speech.

ACKNOWLEDGMENTS

S.K.S. is funded by Grant No. WT074414MA from the Wellcome Trust. Many thanks to Sam Eaton-Rosen for help in creating Fig. 1.

Blesser, B. (**1972**). "Speech perception under conditions of spectral transformation: I. Phonetic characteristics," J. Speech Hear. Res. **15**, 5–41.

Brechmann, A., and Scheich, H. (**2005**). "Hemispheric shifts of sound representation in auditory cortex with conceptual listening," Cereb. Cortex **15**, 578–587.

Brungart, D. S. (**2001**). "Informational and energetic masking effect in the perception of two simultaneous talkers," J. Acoust. Soc. Am. **109**, 1101–1109.

Byrne, D., Dillon, H., Tran, K., Arlinger, S., Wilbraham, K., Cox, R. H. B., Hetu, R. K. J., Liu, C., Kiessling, J., Kotby, M. N., Nasser, N. H. A., Elkholy, W. A. H., Nakanishi, Y. O. H., Powell, R., Stephens, D., Merd-

edith, R., Sirimanna, T., Tavartkiladze, G. F., Westerman, S., and Ludvigsen, C. (**1994**). "An international comparison of long-term average speech spectra," J. Acoust. Soc. Am. **96**, 2108–2120.
Festen, J., and Plomp, R. (**1990**). "Effects of fluctuating noise and interfering speech on the speech reception threshold for impaired and normal hearing," J. Acoust. Soc. Am. **88**, 1725–1736.
Foster, J. R., Summerfield, A. Q., Marshall, D. H., Palmer, L., Ball, V., and Rosen, S. (**1993**). "Lip-reading the BKB sentence lists—Corrections for list and practice effects," Br. J. Audiol. **27**, 233–246.
Hawley, M. L., Litovsky, R. Y., and Culling, J. F. (**2004**). "The benefit of binaural hearing in a cocktail party: Effect of location and type of interferer," J. Acoust. Soc. Am. **115**, 833–843.
Huckvale, M. (**2007**). "Speech filing system (software package)," http://www.phon.ucl.ac.uk/resource/sfs/ (Last viewed May, 2008).
Jacquemot, C., Pallier, C., LeBihan, D., Dehaene, S., and Dupoux, E. (**2003**). "Phonological grammar shapes the auditory cortex: A functional magnetic resonance imaging study," J. Neurosci. **23**, 9541–9546.
Johnstone, P. M., and Litovsky, R. Y. (**2006**). "Effect of masker type and age on speech intelligibility and spatial release from masking in children and adults," J. Acoust. Soc. Am. **120**, 2177–2189.
MacLeod, A., and Summerfield, Q. (**1987**). "Quantifying the contribution of vision to speech perception in noise," Br. J. Audiol. **21**, 131–141.
Mummery, C. J., Ashburner, J., Scott, S. K., and Wise, R. J. S. (**1999**). "Functional neuroimaging of speech perception in six normal and two aphasic patients," J. Acoust. Soc. Am. **106**, 449–457.
Narain, C., Scott, S. K., Wise, R. J. S., Rosen, S., Leff, A. P., Iversen, S. D., and Matthews, P. M. (**2003**). "Defining a left-lateralized response specific to intelligible speech using fMRI," Cereb. Cortex **13**, 1362–1368.
Patterson, R. D., Uppenkamp, S., Johnsrude, I. S., and Griffiths, T. D. (**2002**). "The processing of temporal pitch and melody information in auditory cortex," Neuron **36**, 767–776.
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," J. Acoust. Soc. Am. **118**, 1274–1277.
Roucus, S., and Wilgus, A. W. (**1985**). "High quality time-scale modification for speech," in Proceedings of the IEEE International Conference of Acoustics, Speech and Signal Processing, pp. 493–496.
Scott, S. K., Rosen, S., Lang, H., and Wise, R. J. S. (**2006**). "Neural correlates of intelligibility in speech investigated with noise vocoded speech: A positron emission tomography study," J. Acoust. Soc. Am. **120**, 1075–1083.
Scott, S. K. (**2005**). "Auditory processing–speech, space and auditory objects," Curr. Opin. Neurobiol. **15**, 197–201.
Scott, S. K., Blank, S. C., Rosen, S., and Wise, R. J. S. (**2000**). "Identification of a pathway for intelligible speech in the left temporal lobe," Brain **123**, 2400–2406.
Scott, S. K., and Johnsrude, I. S. (**2003**). "The neuroanatomical and functional organization of speech perception," Trends Neurosci. **26**, 100–107.
Scott, S. K., Rosen, S., Wickham, L., and Wise, R. J. S. (**2004**). "A positron emission tomography study of the neural basis of informational and energetic masking effects in speech perception," J. Acoust. Soc. Am. **115**, 813–821.
Scott, S. K., and Wise, R. J. S. (**2003**). "Functional imaging and language: A critical guide to methodology and analysis," Speech Commun. **41**, 7–21.
Scott, S. K., and Wise, R. J. S. (**2004**). "The functional neuroanatomy of prelexical processing of speech," Cognition **92**, 13–45.
Wise, R. J. S., Scott, S. K., Blank, S. C., Mummery, C. J., and Warburton, E. (**2001**). "Identifying separate neural sub-systems within 'Wernicke's area,'" Brain **124**, 83–95.
Zatorre, R. J., and Belin, P. (**2001**). "Spectral and temporal processing in human auditory cortex," Cereb. Cortex **11**, 946–953.
Zatorre, R. J., Bouffard, M., and Belin, P. (**2004**). "Sensitivity to auditory object features in human temporal neocortex," J. Neurosci. **24**, 3637–3642.

Multisensory integration enhances phonemic restoration

Antoine J. Shahin[a)]
Center for Mind & Brain, University of California, 267 Cousteau Place, Davis, California 95618

Lee M. Miller
Center for Mind & Brain, University of California, 267 Cousteau Place, Davis, California 95618 and Department of Neurobiology, Physiology, & Behavior, University of California, One Shields Avenue, Davis, California 95616

(Received 7 July 2008; revised 16 December 2008; accepted 17 December 2008)

Phonemic restoration occurs when speech is perceived to be continuous through noisy interruptions, even when the speech signal is artificially removed from the interrupted epochs. This temporal filling-in *illusion* helps maintain robust comprehension in adverse environments and illustrates how contextual knowledge through the auditory modality (e.g., lexical) can improve perception. This study investigated how one important form of context, visual speech, affects phonemic restoration. The hypothesis was that audio-visual integration of speech should improve phonemic restoration, allowing the perceived continuity to span longer temporal gaps. Subjects listened to tri-syllabic words with a portion of each word replaced by white noise while watching lip-movement that was either *congruent*, temporally reversed (*incongruent*), or *static*. For each word, subjects judged whether the utterance sounded continuous or interrupted, where a "continuous" response indicated an illusory percept. Results showed that illusory filling-in of longer white noise durations (longer missing segments) occurred when the mouth movement was *congruent* with the spoken word compared to the other conditions, with no differences occurring between the *static* and *incongruent* conditions. Thus, phonemic restoration is enhanced when applying contextual knowledge through multisensory integration. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3075576]

PACS number(s): 43.71.Sy, 43.71.Rt [DOS] Pages: 1744–1750

I. INTRODUCTION

Phonemic restoration (PR) or the auditory continuity *illusion* (Warren, 1970; Samuel, 1981a; Bashford *et al.*, 1992; Shinn-Cunningham and Wang, 2008) occurs when interrupted speech is perceived to be continuous, provided the missing speech is replaced by another sound such as white noise (Warren, 1970; Warren *et al.*, 1997). In ecologically naturalistic, acoustically adverse environments, this illusory filling-in significantly increases speech comprehension (Bashford *et al.*, 1992). The PR *illusion* thus exemplifies how the brain uses context to repair degraded sensory information and thereby improve perception.

Phonemic restoration can be understood as a form of template matching, a negotiation between stimulus acoustics and prior knowledge. Naturally, stimulus acoustics during the noisy interruption influence the restoration of speech. PR tends to be enhanced when the replaced speech and the interrupting or masking sound share spectrotemporal characteristics (Warren, 1970; Samuel, 1981b, 1991; Petkov *et al.*, 2007). For example, white noise induces a stronger *illusion* when it replaces a fricative as opposed to a vowel (Samuel, 1981b), and the *illusion* is improved if amplitude modulations in the masking noise match the envelope of the missing utterance (Bashford *et al.*, 1996). However context, or the situation in which the interruption occurs, also profoundly influences PR by constraining the prior knowledge that may be used to restore the degraded acoustics. Context includes phonetic, lexical, semantic, and syntactic cues (Bashford *et al.*, 1992; Samuel, 1997; Friederici, 2002; Sivonen *et al.*, 2006b; Sivonen *et al.*, 2006b; Shahin *et al.*, 2008). For example, (1) *Semantic*: Sivonen *et al.* (2006a) reported that subjects recognize a sentence-final word with an obliterated phoneme faster when the word is highly expected. (2) *Lexical*: The more syllables present in a word, the easier the restoration (Samuel, 1981a), and words are more readily restored than phonologically legal pseudowords (Samuel, 1981a)—even engaging different neural networks during illusory restoration (Shahin *et al.*, 2008). PR is also influenced by cognitive status such as attentional state (Samuel and Ressler, 1986; Samuel, 1991). Each of these factors affects the level of matching between stimulus attributes and prior knowledge, and hence speech perception. Context determines what kind of information, and how much of it, the brain can apply from its experiential knowledge base to perceptually repair the degraded speech.

Most studies on PR have addressed the influence of stimulus attributes and context in only one sensory modality: audition. Here we examine the effect of visual speech, one of the most important forms of context in noisy environments. The premise is that visual speech, by providing more information about the degraded epoch, should facilitate template matching and allow the brain to repair or "fill-in" longer degradations. That is, audio-visual integration should enhance the phonemic restoration. This hypothesis is motivated

[a)]Author to whom correspondence should be addressed. Electronic mail: ajshahin@ucdavis.edu

 0001-4966/2009/125(3)/1744/7/$25.00

by numerous demonstrations that vision improves comprehension in noise (Sumby and Pollack, 1954; Grant *et al.*, 1998), influences speech perception at a basic phonetic level (McGurk and MacDonald, 1976; Repp *et al.*, 1992; Brancazio *et al.*, 2003; Helfer and Freyman, 2005; van Wassenhove *et al.*, 2005), and can even actively recalibrate speech sound categorization (Vroomen and Baart, 2008).

Only one prior study investigated how vision might affect phonemic restoration, by comparing an auditory-only condition with a *congruent* audio-visual condition (Trout and Poser, 1990). In that study, white noise either replaced or was added to an acoustic phoneme in a word. Subjects were asked whether they heard the word as intact (noise added) or interrupted (noise replaced). The authors hypothesized that since visual cues tend to enhance speech perception, PR should increase with vision. PR enhancement would then be reflected in reduced discriminability between added and replaced stimuli (reduced d'). The study found that visual input did not affect the discriminability of the speech signal. However, it found that the presence of visual cues reduced the bias to report replaced speech as continuous. In comparison to the Trout and Poser study (1990), the current design uses a different approach to test visual effects on PR. Here we use the duration of a white noise interruption to quantify the strength or efficacy of PR, adaptively in each subject under different visual conditions.

We examined whether mouth movement *congruent* with acoustic speech enhances phonemic restoration. Subjects listened to spoken words with white noise either replacing or added to a fricative/affricate. For each word, subjects identified whether the utterance sounded continuous through the noise or interrupted. Thus, a "continuous" response for a replaced phoneme indicated successful phonemic restoration. The duration of the white noise burst was adaptively adjusted to identify each subject's threshold for restoring interrupted speech. The auditory stimuli were presented with mouth movements either *congruent* or *incongruent* with the spoken word or with no mouth movement (*static condition). The incongruent condition was used to control for visual distraction: if PR enhancement is merely due to visual distraction, then the congruent and incongruent conditions should have equal white noise durations, but longer durations with respect to the static condition. Rather, we hypothesized that greater contextual information provided through audio-visual integration would result in illusory filling-in over longer gaps and hence longer white noise durations for the congruent condition with respect to the incongruent and static conditions.*

II. METHODS

A. Subjects

Twenty subjects were recruited for the study. Data from 14 subjects (aged 25.7 ± 5.7 years; 8 female, 1 left handed) who performed the task consistently (described below) were included in the final analysis. Informed consent was obtained from all subjects in accordance with the ethical guidelines of the University of California.

B. Stimuli

The stimuli were 230 tri-syllabic English nouns and adjectives compiled using the University of Western Australia MRC Psycholinguistic Database (http://www.psy.uwaedu.au/mrcdatabase/uwa_mrc.htm). Each word had a familiarity rating of 300–700 and contained between 1 and 3 fricatives/affricates, with at least one occurring between the first and last phonemes. The auditory and visual stimuli, recorded simultaneously, were produced by a professionally trained female vocalist (f_0 of 203 Hz). Informed consent, in accordance with the ethical guidelines of the University of California, was obtained from the professional vocalist prior to recording. The auditory signal was recorded using a Shure KSM studio microphone (http://www.shure.com) at a sampling rate of 48 kHz, in an acoustically shielded room. The visual stimuli were recorded using a Panasonic AG-DVX100A digital camera. The camera resolution was set to 30 frames per second.

Before further processing (see below), each audio-visual word was first centered in a 2.85 s (95 frames) segment using ADOBE PREMIERE PRO 2.0 (Adobe Systems Inc., San Jose, CA). This resulted in each segment's auditory signal containing about a second of silence at the beginning and end, with each video containing less than a second of motionless talker at the beginning and end (since lip-movement precedes acoustic onset and persists after acoustic offset).

1. Preparation of visual stimuli

The visual part of each audio-visual word (avi files) was extracted as a series of 30 ms frames, and processed into *congruent*, *incongruent*, and *static* visual stimuli as follows: naturally the video frames for the *congruent* condition words were perfectly aligned in time with the acoustic speech. To reduce the number of no-motion frames, the *congruent* condition was reduced to 85 frames, corresponding to a 2.55 s duration, of the original 95 frames. To prepare the visual stimulus of the *incongruent* condition, the videos for the *congruent* condition words were temporally reversed as follows: (1) the midpoint of the acoustical signal for each word was determined. (2) The video frame of the *congruent* condition corresponding to the acoustic midpoint of step 1 was determined. (3) Frames were extracted from this mid-utterance video frame minus 30 frames through the mid-utterance frame plus 30 frames of the *congruent* video. Notice that in this step, all sound and lip-movement spanning the onsets and offsets were included in the extracted segments. (4) The frame series from the third step was reversed in time to produce an *incongruent* frame series. (5) The *incongruent* frame series was padded by a variable number of frames (padded at the beginning with the first frame and at the end with last frame for each word) to comprise 85 frames. This procedure ensured that the mouth movement onset time and overall mouth movement energy during the speech were identical for each word between the *congruent* and *incongruent* conditions. This also ensured that visual discontinuities occurred only once at trial onset (approximately 1 s before acoustic onset). For the *static* condition, all of the 85 frames used were the same as the first still frame of each word with lips

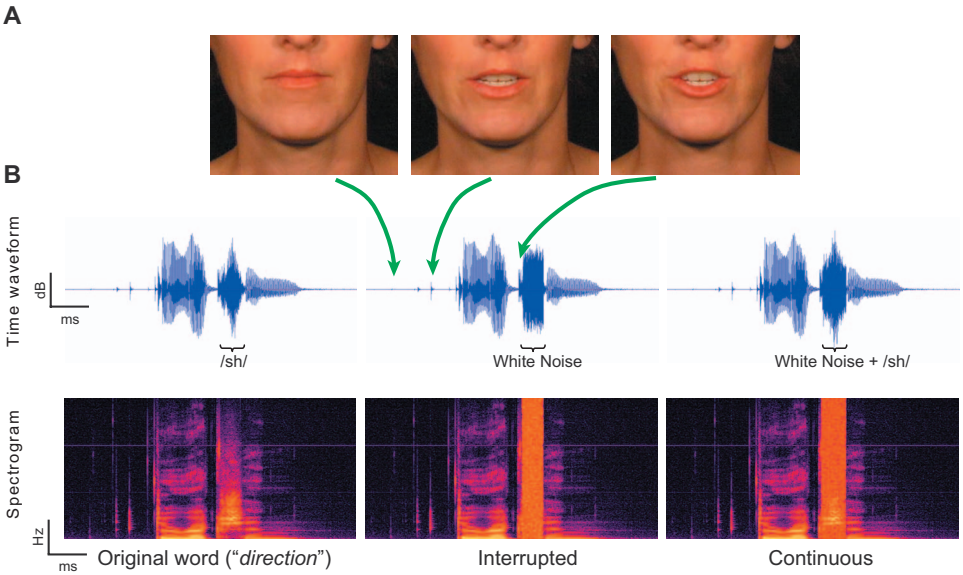


FIG. 1. Stimuli. (A) Three picture frames corresponding to mouth movements for three different times: 1—no sound; 2—beginning of mouth movement preceding the utterance; 3—mouth movement corresponding to the onset of the fricative /ʃ/. (B) The time waveforms (top) and corresponding spectrograms (bottom) for an example word ("direction"). The middle panel depicts a physically interrupted word, where white noise replaced part of the fricative /ʃ/. The right panel depicts a physically continuous word, where white noise was superimposed on the fricative. The left panel depicts the original word with no noise, which was not among the experimental stimuli.

closed. Accordingly, similar to the *congruent* and *incongruent* conditions, the *static* trial-onset visual discontinuities occurred only once at the beginning of each trial. Figure 1(A) shows three frames displaying part of the vocalist's face including the mouth. Each frame corresponded to a particular time in the acoustic presentation (Fig. 1(B)).

2. Preparation of auditory stimuli

Since the number of frames for each word was truncated to 85 frames, the auditory stimuli were trimmed accordingly. The acoustic part of each audio-visual word was extracted and processed as follows: Initially, a white noise segment was aligned with one fricative/affricate ([tʃ], [dʒ], [s], [ʃ], or [z]) between the first and last phonemes of each word. Fricatives/affricates were used to yield robust PR effects (Samuel, 1981a). The beginning and end of each fricative or affricate for each word were determined using ADOBE AUDITION 2.0 (Adobe Systems Inc., San Jose, CA). Each original word was then used to create both an "interrupted" and a continuous version. Interrupted words were those in which part of the word (centered on the fricative/affricate) was completely replaced by white noise, while continuous words were those in which white noise was superimposed on part of the word, centered on the fricative/affricate. To rule out effects due to physical differences of the acoustical stimuli across conditions, we first matched all words in sound level based on their A-weighted root mean square (RMS). Second, the A-weighted RMS sound level of the white noise segment was equalized to the A-weighted RMS sound level of the replaced/superimposed speech plus 3 dB. Because the white noise is uncorrelated with the fricative/affricate, this resulted in a RMS difference between the continuous and interrupted words of less than 1 dB for all stimuli. All words were identifiable and unambiguous, even when interrupted, thus minimizing differential semantic effects. Figure 1(B) shows the time waveforms and spectrograms for a representative word ("direction") with 100% of the fricative ([ʃ]) replaced with white noise (middle panel, interrupted) and with white noise superimposed (right panel, continuous). For comparison, the left panel shows the original waveform and spectrotemporal representation of the word.

C. Procedure and task

Subjects sat approximately 50 cm in front of a 24 in. LCD monitor and wore AKG 271 studio headphones (AKG Acoustics GmbH, Vienna, Austria). Visual stimuli spanned approximately 15°. Sound loudness was adjusted to the subject's comfort level and was kept constant across the entire experiment. The task involved nine blocks (three blocks for each condition, ~12 min per block).

In each of the nine blocks, subjects were presented with the 230 audio-visual words consecutively. The order of presentation of words was randomized by a genetic algorithm (Wager and Nichols, 2003) that ensured adequate counterbalancing between interrupted versus continuous stimulus conditions. The nine blocks were randomized across subjects, but the presentation order of stimuli within each block was the same for all subjects. Subjects took a few minute break between blocks. Following each audio-visual word presentation, there was a 0.6 s silent period combined with a still picture of the last displayed frame. Because sound onsets and offsets occurred at variable time from trial-to-trial, the sound interstimulus interval (ISI) ranged between 1 and 3 s (mean 2.15 ± 0.32 s). The ISI for lip-movements offset to onset was smaller. As rapidly as possible, subjects pressed their left index finger when they perceived the stimulus as continuous and their left middle finger when they perceived the stimulus as interrupted. They were instructed to focus their vision on the talker's face at all times while making their discrimination based on what they heard. They were also explicitly instructed to base their judgments on the continuity of the word and to ignore the white noise.

A two alternative forced choice adaptive procedure was used so that every subject experienced stimuli with white noise duration at or near their own psychophysical threshold for perceiving the restoration *illusion*. The psychophysical threshold is the white noise proportion where the subject perceives interrupted speech as continuous on about 50% of

trials. Examining differences at the subject's own threshold minimizes the physical differences of the acoustical signals within each of the three visual conditions (between continuous and interrupted percepts) and ensures that subjects perform at the same point in their psychometric range across conditions. This allows us to use noise proportion to measure the efficacy of phonemic restoration. If an actually interrupted word was identified as interrupted, the trial was labeled an *illusion-failure*. If an interrupted word was identified as continuous, the trial was labeled an *illusion* (the PR percept). If a continuous word was identified as continuous, the trial was labeled *natural*, and finally, if a continuous word was identified as interrupted, the trial was labeled a *miss*. At the start of the experiment, a white noise segment spanning a proportion of 100% of the fricative/affricate duration was used to either replace (physically interrupted word) the fricative/affricate or add to it (physically continuous word). The adaptation was done for each condition (*static*, *congruent*, and *incongruent*) separately. If, during the physically interrupted trials only, subjects identified the word as continuous (*illusion*), the gap and thus the white noise segment were increased on the next physically interrupted trial by 15% of the actual fricative/affricate (7.5% on either side of the fricative center) duration. Note that increased gap and hence white noise duration make it harder to identify the word as continuous. Conversely, the gap, and thus the white noise segment, was decreased by the same amount if they identified the word as interrupted (*illusion-failure*). The lengths of white noise segments between different words thus varied in proportion with respect to the replaced fricative/affricate rather than by a fixed duration. This allowed evaluation of illusory continuity based on representations contained in speech rather than information contained in fixed temporal values (Bashford *et al.*, 1988). It should be clear that the adaptive procedure allows the gap and white noise proportion to span part of or extend beyond the fricative/affricate. The noise duration algorithm only adapted for physically interrupted trials. For the physically continuous trials, the gap and thus the white noise proportion were the same as the most recent interrupted trial. This equalized the noise proportion distribution between physically continuous and physically interrupted trials. In each block, 70% of the words presented were physically interrupted and 30% were physically continuous. This was done deliberately to increase statistical power for the condition of interest (*illusion*), i.e., to attain a more reliable estimate of white noise proportion differences of the *illusion* percept between the three conditions. The remaining 30% of the stimuli were physically continuous (producing *natural* and *miss* percepts) and were included as a supplementary measure in the design to evaluate subjects' performance. First, a large *natural/miss* ratio indicates that subjects' responses were minimally influenced by guessing. For example, if the 35/35 response percentage seen for the *illusion/illusion-failure* were due to guessing, then we should see a similar ratio for *natural/miss*. As will be shown below, this was not the case. Of the 20 original subjects, only subjects with a *natural/miss* of 1.5 or larger were included. In addition, only subjects who had white noise proportion that deviated less than 50% between blocks were considered further. A 50% or larger deviation in white noise proportion indicated that those individuals changed strategies between blocks or did not understand the task initially. 14 of the initial 20 subjects satisfied these criteria. It should be noted, however, that including all 20 subjects in the analysis yields qualitatively identical results. Responses and white noise proportion were monitored and logged via PRESENTATION (Neurobehavioral Systems, Inc., Albany, CA) software.

TABLE I. Mean percentage and standard deviation of the *illusion*, *illusion-failure*, *natural*, *miss*, and trials with no response across the 14 subjects.

	Static	*Incongruent*	*Congruent*
Illusion	34 ± 1	34 ± 2	35 ± 1
Illusion-failure	34 ± 2	32 ± 3	33 ± 2
Natural	26 ± 4	26 ± 4	28 ± 3
Miss	3 ± 3	3 ± 3	3 ± 3
No response (NR)	2 ± 3	4 ± 5	2 ± 2

D. Analysis and statistics

Each subject's white noise proportions were obtained as the mean noise proportion across the *illusion* trials for each condition: *static*, *incongruent*, and *congruent*. Trials on which the subject gave no response were excluded from the analysis. Subject performance was based on the number and white noise proportion of *natural* stimuli—correctly identified continuous stimuli—for each condition.

Statistical tests were conducted using STATISTICA version 8.0 (StatSoft, Inc., Tulsa, OK). Differences in white noise proportion and *natural* response counts were calculated separately for each condition using analysis of variance (ANOVA), with condition as the only variable. *Post hoc* comparisons were made by paired t-tests.

III. RESULTS

Our adaptive design sought to yield an equal percentage of trials (35/35) between the *illusion* responses, which represent the PR percept, and the *illusion-failure* responses. The remaining 30% of physically continuous stimuli were incorporated to test performance. Table I shows the response mean percentage and standard deviation of the *illusion*, *illusion-failure*, *natural*, *miss*, and trials with no response across the 14 subjects. The equivalent numbers between *illusion* and *illusion-failure* across the three conditions speaks to our successful adaptive method. The large ratio of the *natural* to *miss* responses indicates that subjects performed the task as required—paying attention with minimal guessing. Since all these experimental criteria were met, we can therefore use white noise proportion to measure the strength or efficacy of PR and the *natural* response count as a measure of performance.

A. White noise proportion

According to our hypothesis, a greater white noise proportion indicates enhanced PR. Figure 2 shows the white noise proportion for each condition normalized to the within-

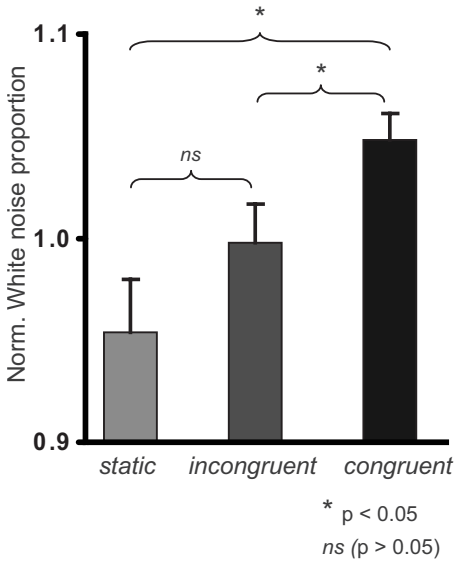


FIG. 2. Normalized white noise proportions with respect to the replaced fricative for (1) no mouth movement (*static*), (2) mouth movement *incongruent* with the spoken word (*incongruent*), and (3) mouth movement *congruent* with the spoken word (*congruent*).

subject mean of the three conditions, for the 14 subjects. Unnormalized, the mean and standard deviation of white noise proportion across the three conditions and all subjects were $134 \pm 48\%$. This corresponded to average noise duration of 172 ms. An ANOVA with condition as the only variable showed a main effect [$F_{(2,26)}=3.7; p<0.039$] for white noise proportion. Consistent with our hypothesis, *post hoc* t-tests revealed that white noise proportion was larger for words when the mouth movements were *congruent* with the auditory stimuli compared to the *static* and *incongruent* conditions [*congruent* versus *static*: $t_{(13)}=2.6$ and $p=0.022$; *congruent* versus *incongruent*: $t_{(13)}=2.5$; $p=0.027$]. Though the white noise proportion of the *incongruent* condition tended to be larger compared to the *static* condition, this difference was statistically weak ($p>0.25$).

B. *Natural* response count as a measure of performance

If audio-visual integration enhances PR and thereby improves speech perception, we might observe two related effects. First, as suggested in Sec. I and confirmed in Sec. III A, PR as measured by white noise duration should be enhanced. Alternately, for equivalent white noise durations, subjects should perform better with *congruent* audio-visual speech. Since our design adapted the noise duration to each subject's threshold within condition, we cannot compare performance across conditions with identical noise durations. However, our adaptive approach does allow a conservative test of performance differences. Specifically, subjects might perform better during audio-visual integration even though they endure longer noisy epochs (mean white noise proportion $\pm$SD normalized across the three conditions for the *congruent*, *incongruent*, and *static* conditions were 0.97 ± 0.09, 0.99 ± 0.07, and 1.04 ± 0.05, respectively). We therefore measured performance across conditions as the number of *natural* responses which represent correctly identified continuous stimuli. Normalized to the within-subject means of the three conditions, the number of *natural* responses was significantly larger (Table I) for the *congruent* than the other two conditions [t-tests, $t_{(14)}=2.7$; $p<0.02$ or better] with no differences between the *static* and *incongruent* conditions (t-tests, $p>0.6$). So, despite significantly longer noise bursts obscuring the underlying speech in the *congruent* condition, subjects performed better compared to the *static* and *incongruent* conditions.

IV. DISCUSSION

We have shown that audio-visual integration enhances the PR *illusion*, allowing the brain to fill-in longer epochs of degraded speech. This is consistent with our prediction that meaningful (e.g., *congruent*) mouth movement modulates our auditory experience to benefit speech perception in noise.

To perceptually restore speech in the absence of visual input, template matching could fit sensory input to closely related representations in memory, and may be lexical (Samuel, 1981a; Binder *et al.*, 2003), semantic (Sivonen *et al.*, 2006b), or syntactic (Friederici, 2002) in nature. In the current design, speech stimuli were equalized for lexicality across conditions but differed in their degree of auditory-visual congruency. A bimodal *congruent* sensory interaction therefore represents an additional domain for template matching to facilitate auditory perception, especially for lexically-mediated stimuli (Repp *et al.*, 1992). Such an auditory-visual template match may require mapping of visual inputs onto the auditory cortex or vice versa. This is supported by studies showing that meaningful visual inputs can induce activity in the auditory cortex in the absence of sounds (Sams *et al.*, 1991; Calvert *et al.*, 1997; Jancke and Shah, 2004). Also, ample evidence supports that this type of audio-visual convergence facilitates perception (Repp *et al.*, 1992; Helfer and Freyman, 2005; Martuzzi *et al.*, 2007; Romei *et al.*, 2007; Bishop and Miller, 2008). For instance, Helfer and Freyman (2005) showed that lip reading enhanced the ability to segregate one talker in a multi-talker environment, in accordance with our results of enhanced performance in noise for the *congruent* auditory-visual condition.

In addition to the linguistic information conveyed through vision, we should also emphasize that to be useful, visual cues must be temporally coherent with the auditory signal (Grant and Seitz, 2000; Grant, 2001; Yehia *et al.*, 2001). Grant and Seitz (2000) showed that when lip-movement for a sentence matches with the corresponding acoustical signal of the sentence, average speech detection threshold improves by 1.6 dB compared to auditory-only condition. Subsequent analysis indicated that speech detection threshold is related to the degree of temporal synchrony between the timing of mouth opening and acoustical variations reflected in the speech temporal envelope (Grant and Seitz, 2000). Our results concur with their conclusion. Of course in the present study, while mean visual energy spanning the period of auditory presentation was equivalent between the *audio-visual* conditions, mouth openings and acoustical energy variations were temporally related for the

congruent but not the *incongruent* condition. This temporal coherence is likely another reason why PR enhancement was only reflected in the *congruent* condition.

The *incongruent* condition, however, showed a trend toward enhanced PR compared to the *static* condition. This suggests that arbitrary visual events preceding or during listening could distract the listener and lead to an increased *illusion*, consistent with the findings of Kobayashi *et al.* (2007). Using a pure tone with gaps masked by white noise, they (Kobayashi *et al.*, 2007) showed that white noise proportion increased for the illusory response when a flash of light preceded white noise onset. The presence of visual cues unrelated to the auditory stimulus may shift the attentional state toward a global rather than a narrow scope resulting in, as posited by Samuel and Ressler (Samuel and Ressler, 1986; Samuel, 1991), a "failure to attend to low-level aspects of the stimulus because of global dominance." Nevertheless, PR for *congruent* audio-visual speech was still significantly enhanced relative to the *incongruent* condition, demonstrating that audio-visual integration dominates any attentional effects. Thus, the trend toward enhanced PR (longer white noise duration) in the *incongruent* condition was likely caused by attentional effects and not the visual information itself as in the *congruent* condition.

Context could enhance PR via at least two mechanisms: by reducing the perceptual discriminability of speech vs noise during the interruption (making the noise sound more like speech) or by increasing the post-perceptual bias of a listener to report restoration. Traditionally, *discriminability* and *bias* have been measured by the signal detection theory parameters d' and *beta*, respectively (Samuel, 1981a; Trout and Poser, 1990). For example, Trout and Poser's (1990) findings suggest that the bias, not discriminability, is the sensitive parameter for audio-visual PR. However, Samuel (1981a) found that word-level context affected discriminability, while sentence-level context affected subjects' bias for reporting restored speech. Our design differed from the traditional signal-detection approach by adapting the stimulus properties (noise duration) for each subject, in each condition, throughout the experiment. Though integral to our measure of PR enhancement, this adaptation essentially fixes the criterion value and precludes an unambiguous evaluation of d' and bias. Future studies could combine the adaptive approach with signal detection theory to assess the mechanisms underlying PR enhancement observed here.

Despite overwhelming evidence that vision influences speech perception, a few studies challenge or qualify this conclusion. For instance, Saldana and Rosenblum (1994) presented a case that did not support a visual influence on phonemic identification. In their design, phonemic boundaries for subjects were derived from a "baseline" condition in which syllables of different acoustical parameters were presented randomly, and also from an "adaptation" condition in which a syllable of specific acoustical parameters is repeated several times. The phonemic boundary shifted following the adaptation condition. The authors reported that this shift was not influenced by the visual adaptors; rather it was equally exhibited for audio-visual or auditory-only adaptors. One factor that may explain the differing accounts on audio-visual influence is lexicality. The Saldana and Rosenblum study did not take into account lexical information. Though PR occurs for both words and pseudowords, prior work suggests that lexicality plays an important role in audio-visual integration (Repp *et al.*, 1992), and thus may influence our observed visual effects on PR.

V. CONCLUSIONS

Our results show that PR is enhanced through audio-visual integration (lip-reading). Our findings offer a distinction between PR due to visual information and that likely due to attention. These results expand on earlier work of PR by examining how communication between auditory and visual modalities enhances top-down and bottom-up integration.

ACKNOWLEDGMENT

This research was funded (L.M.M.) by the National Institutes of Health (NIH): National Institute on Deafness and other Communication Disorders (NIH/NIDCD Grant No. R01-DC08171). The authors are grateful to Arthur Samuel and two anonymous reviewers for helpful comments on an earlier version of this manuscript.

Bashford, J. A., Jr., Meyers, M. D., Brubaker, B. S., and Warren, R. M. (**1988**). "Illusory continuity of interrupted speech: Speech rate determines durational limits," J. Acoust. Soc. Am. **84**, 1635–1638.

Bashford, Jr., J. A., Riener, K. R., and Warren, R. M. (**1992**). "Increasing the intelligibility of speech through multiple phonemic restorations," Percept. Psychophys. **51**, 211–217.

Bashford, Jr., J. A., Warren, R. M., and Brown, C. A. (**1996**). "Use of speech-modulated noise adds strong "bottom-up" cues for phonemic restoration," Percept. Psychophys. **58**, 342–350.

Binder, J. R., McKiernan, K. A., Parsons, M. E., Westbury, C. F., Possing, E. T., Kaufman, J. N., and Buchanan, L. (**2003**). "Neural correlates of lexical access during visual word recognition," J. Cogn. Neurosci. **15**, 372–393.

Bishop, C. W., and Miller, L. M. (**2009**). "A multisensory cortical network for understanding speech in noise," J. Cogn Neurosci. (in press).

Brancazio, L., Miller, J. L., and Pare, M. A. (**2003**). "Visual influences on the internal structure of phonetic categories," Percept. Psychophys. **65**, 591–601.

Calvert, G. A., Bullmore, E. T., Brammer, M. J., Campbell, R., Williams, S. C., McGuire, P. K., Woodruff, P. W., Iversen, S. D., and David, A. S. (**1997**). "Activation of auditory cortex during silent lipreading," Science **276**, 593–596.

Friederici, A. D. (**2002**). "Towards a neural basis of auditory sentence processing," Trends Cogn. Sci. **6**, 78–84.

Grant, K. W. (**2001**). "The effect of speechreading on masked detection thresholds for filtered speech," J. Acoust. Soc. Am. **109**, 2272–2275.

Grant, K. W., and Seitz, P. F. (**2000**). "The use of visible speech cues for improving auditory detection of spoken sentences," J. Acoust. Soc. Am. **108**, 1197–1208.

Grant, K. W., Walden, B. E., and Seitz, P. F. (**1998**). "Auditory-visual speech recognition by hearing-impaired subjects: Consonant recognition, sentence recognition, and auditory-visual integration," J. Acoust. Soc. Am. **103**, 2677–2690.

Helfer, K. S., and Freyman, R. L. (**2005**). "The role of visual speech cues in reducing energetic and informational masking," J. Acoust. Soc. Am. **117**, 842–849.

Jancke, L., and Shah, N. J. (**2004**). "Hearing syllables by seeing visual stimuli," Eur. J. Neurosci. **19**, 2603–2608.

Kobayashi, M., Osada, Y., and Kashino, M. (**2007**). "The effect of a flashing visual stimulus on the auditory continuity illusion," Percept. Psychophys. **69**, 393–399.

Martuzzi, R., Murray, M. M., Michel, C. M., Thiran, J. P., Maeder, P. P., Clarke, S., and Meuli, R. A. (**2007**). "Multisensory interactions within

human primary cortices revealed by BOLD dynamics," Cereb. Cortex **17**, 1672–1679.
McGurk, H., and MacDonald, J. (**1976**). "Hearing lips and seeing voices," Nature (London) **264**, 746–748.
Petkov, C. I., O'Connor, K. N., and Sutter, M. L. (**2007**). "Encoding of illusory continuity in primary auditory cortex," Neuron **54**, 153–165.
Repp, B. H., Frost, R., and Zsiga, E. (**1992**). "Lexical mediation between sight and sound in speechreading," Q. J. Exp. Psychol. **45**, 1–20.
Romei, V., Murray, M. M., Merabet, L. B., and Thut, G. (**2007**). "Occipital transcranial magnetic stimulation has opposing effects on visual and auditory stimulus detection: Implications for multisensory interactions," J. Neurosci. **27**, 11465–11472.
Saldana, H. M., and Rosenblum, L. D. (**1994**). "Selective adaptation in speech perception using a compelling audiovisual adaptor," J. Acoust. Soc. Am. **95**, 3658–3661.
Sams, M., Aulanko, R., Hamalainen, M., Hari, R., Lounasmaa, O. V., Lu, S. T., and Simola, J. (**1991**). "Seeing speech: Visual information from lip movements modifies activity in the human auditory cortex," Neurosci. Lett. **127**, 141–145.
Samuel, A. G. (**1981a**). "Phonemic restoration: Insights from a new methodology," J. Exp. Psychol. **110**, 474–494.
Samuel, A. G. (**1981b**). "The role of bottom-up confirmation in the phonemic restoration illusion," J. Exp. Psychol. Hum. Percept. Perform. **7**, 1124–1131.
Samuel, A. G. (**1991**). "A further examination of attentional effects in the phonemic restoration illusion," Q. J. Exp. Psychol. **43**, 679–699.
Samuel, A. G. (**1997**). "Lexical activation produces potent phonemic percepts," Cogn. Psychol. **32**, 97–127.
Samuel, A. G., and Ressler, W. H. (**1986**). "Attention within auditory word perception: insights from the phonemic restoration illusion," J. Exp. Psychol. Hum. Percept. Perform. **12**, 70–79.
Shahin, A. J., Bishop, C. W., and Miller, L. M. (**2009**). "Neural mechanisms for illusory filling-in of degraded speech," Neuroimage **44**, 1133–1143.
Shinn-Cunningham, B. G., and Wang, D. (**2008**). "Influences of auditory object formation on phonemic restoration," J. Acoust. Soc. Am. **123**, 295–301.
Sivonen, P., Maess, B., and Friederici, A. D. (**2006a**). "Semantic retrieval of spoken words with an obliterated initial phoneme in a sentence context," Neurosci. Lett. **408**, 220–225.
Sivonen, P., Maess, B., Lattner, S., and Friederici, A. D. (**2006b**). "Phonemic restoration in a sentence context: Evidence from early and late ERP effects," Brain Res. **1121**, 177–189.
Sumby, W. H., and Pollack, I. (**1954**). "Visual contribution to speech intelligibility in noise," J. Acoust. Soc. Am. **26**, 212–215.
Trout, J. D., and Poser, W. J.(**1990**). "Auditory and visual influences on phonemic restoration," Lang Speech **33**, 121–135.
van Wassenhove, V., Grant, K. W., and Poeppel, D. (**2005**). "Visual speech speeds up the neural processing of auditory speech," Proc. Natl. Acad. Sci. U.S.A. **102**, 1181–1186.
Vroomen, J., and Baart, M. (**2009**). "Recalibration of phonetic categories by lipread speech: Measuring aftereffects after a twenty-four hours delay," Lang. Speech (in press).
Wager, T. D., and Nichols, T. E. (**2003**). "Optimization of experimental design in fMRI: A general framework using a genetic algorithm," Neuroimage **18**, 293–309.
Warren, R. M. (**1970**). "Perceptual restoration of missing speech sounds," Science **167**, 392–393.
Warren, R. M., Hainsworth, K. R., Brubaker, B. S., Bashford, Jr., J. A., and Healy, E. W. (**1997**). "Spectral restoration of speech: Intelligibility is increased by inserting noise in spectral gaps," Percept. Psychophys. **59**, 275–283.
Yehia, H. C., Rubin, P. E., and Vatikiotis-Bateson, E. (**2001**). "Quantitative association of vocal-tract and facial behavior," Speech Commun. **26**, 23–43.

Simulation of elastic wave scattering in cells and tissues at the microscopic level[a)]

Timothy E. Doyle, Adam T. Tew, Keith H. Warnick, and Brent L. Carruth
Department of Physics, Utah State University, Logan, Utah 84322-4415

(Received 30 May 2008; revised 12 December 2008; accepted 17 December 2008)

The scattering of longitudinal and shear waves from spherical, nucleated cells and three-dimensional tissues with simple and hierarchical microstructures was numerically modeled at the microscopic level using an iterative multipole approach. The cells were modeled with a concentric core-shell (nucleus-cytoplasm) structure embedded in an extracellular matrix. Using vector multipole expansions and boundary conditions, scattering solutions were derived for single cells with either solid or fluid properties for each of the cell components. Tissues were modeled as structured packings of cells. Multiple scattering between cells was simulated using addition theorems to translate the multipole fields from cell to cell in an iterative process. Backscattering simulations of single cells indicated that changes in the shear properties and nuclear diameter had the greatest effect on the frequency spectra. Simulated wave field images and high-frequency spectra (15–75 MHz) from tissues containing 1211–2137 cells exhibited up to 20% enhancement of the field amplitudes at the plasma membrane, significant changes in spectral features due to neoplastic and other microstructural alterations, and a detection threshold of ~8.5% infiltration of tumor cells into normal tissue. These findings suggest that histology-based simulations may provide insight into fundamental ultrasound-tissue interactions and help in the development of new medical technologies. [DOI: 10.1121/1.3075569]

PACS number(s): 43.80.Cs, 43.80.Jz, 43.80.Qf, 43.20.Bi [CCC] Pages: 1751–1767

I. INTRODUCTION

Establishing a mechanistic link between ultrasonic measurements and histological changes associated with various diseases is critical in advancing diagnostic ultrasound. To date, empirical and analytical methods have been primarily used to produce this link. Empirical correlations provide strong evidence for histology-dependent changes in ultrasonic scattering for pathologies such as liver disease (Gaitini *et al.*, 2004; Yeh *et al.*, 2005; Meziri *et al.*, 2005; Liu *et al.*, 2006), breast cancer (Alacam *et al.*, 2004), and prostate cancer (Feleppa *et al.*, 1996), but are limited in use to well-characterized tissues. Analytical models have been developed to further our understanding of the relationships between the measured parameters, the cell and tissue structures, and the ultrasonic interactions. Analytical models simplify the scattering problem, however, by using approximations that exclude multiple scattering (MS), wave-mode conversion, and tissue heterogeneities. These factors have been found to be far from negligible (Lizzi *et al.*, 1997; Aubry *et al.*, 2008). More accurate, histology-based approaches for modeling ultrasound in tissues are therefore needed to develop new technologies for tissue characterization and improve the ultrasonic detection of cancer and other pathologies.

This paper presents the theory and results of a three-dimensional (3D) numerical model to address this goal. The ultrasonic interactions are simulated at the microscopic level by calculating the scattering from cells and nuclei modeled to a first approximation by spheres. The cell and nuclei sizes can be varied independently in the model, allowing the simulation of polydisperse cell and nuclei populations. The cell positions can additionally be configured to simulate a wide range of arbitrary microstructures including random, ordered, and complex hierarchical structures. Depending on the desired information, the model can compute wave field images, spectra, or macroscopic properties such as effective wave speed and attenuation.

Both empirical and analytical models have been developed to estimate tissue parameters such as cell sizes and distributions from ultrasonic backscatter measurements and to correlate them to histological modifications due to cancer and other diseases (Lizzi *et al.*, 1983, 1997; Kutay *et al.*, 2001; Oelze *et al.*, 2002; Bige *et al.*, 2006; Oelze and Zachary, 2006). Empirical models statistically define spectral parameters and their associated probability density functions from tissue measurements. Comparisons between empirical models and experimental results have shown excellent agreement for homogeneous liver and prostate tissues but discrepancies for heterogeneous tissues (Lizzi *et al.*, 1997). Analytical models average the scattering from uniformly distributed cells or cell nuclei, and employ simplifications such as the Rayleigh approximation for point scatterers, the Born approximation of weak (single) scattering, and fluid-like tissue properties with no shear waves (Oelze *et al.*, 2002).

[a)]Portions of this work were presented as "Simulation of elastic wave scattering in living tissue at the cellular level," at the Fourth Joint Meeting of the Acoustical Society of America and the Acoustical Society of Japan, Honolulu, HI, 2006, and as "Histology-based simulation of ultrasonic scattering in cells and tissues," 153rd Meeting of the Acoustical Society of America, Salt Lake City, UT, 2007.

To improve on these approximations, Baddour *et al.* (2005) and Baddour and Kolios (2007) modeled single-cell scattering with analytical expressions derived for scattering from simple fluid or elastic spheres with no internal structure. Further improvements have included modeling the scattering sites with Gaussian and exponential form factors (Oelze *et al.*, 2002; Mamou *et al.*, 2008). The form factors provide the scatterers, typically cells, with an approximate internal structure that has a higher density in the center and a gradually decreasing density with radial distance, but with no distinct boundary between tissue and scatterer/cell.

Cells have also been modeled with a more realistic internal structure comprised of a nucleus and cytoskeleton, and the results of these models were more consistent with ultrasonic data from tumors than those from cells modeled as fluid-filled spheres (Oelze and O'Brien, 2006; Oelze *et al.*, 2007). However, these models did not include MS, shear waves, or higher-order tissue structures. Mamou *et al.* (2008) concluded that ultrasonic scattering arises from both individual cells and their internal components such as nuclei, organelles, and cytoplasm. Therefore, a more accurate representation of the cells' internal structure is required for a histologically accurate, microscopic-level model of ultrasonic scattering in tissues. Analytical and semi-analytical models have additionally been developed to incorporate the effects of anisotropy in tissues (Insana, 1995), bone microstructure (Wear, 2004), and cell shape (Rose *et al.*, 1995; Santarelli and Landini, 1996; Savéry and Cloutier, 2007).

A simulation approach is attractive for modeling microstructural changes in tissues since it has the ability to include interactions that have been ignored by previous models such as MS and wave-mode conversion. MS becomes significant for strongly scattering or closely spaced scatterers, and can significantly affect the wave propagation properties of heterogeneous media. The role of MS in tissues has traditionally been assumed to be negligible, but recent experiments show that breast tissues produce appreciable MS at ~4.3 MHz (Aubry *et al.*, 2008). MS effects may also be important at high frequencies (>10 MHz) or in tumors where the close packing of cells, the specific densities of cells, and the nuclear diameters are typically higher than for normal tissues. Ultrasonic shear waves have also been considered insignificant in tissues due to high attenuation, but mode conversion of longitudinal waves to shear waves at the plasma membrane or nuclear envelope may be an important loss mechanism for ultrasonic energy, particularly for tissues with high shear stiffness such as solid tumors.

A simulation approach can also include heterogeneous distributions of cells with specific histological features and internal structures. By directly simulating changes in the structure and material properties of cells (e.g., malignant cells with pleomorphic nuclei) and tissues (e.g., loss of gland microstructure in breast malignancies), ultrasonic measurements can be mechanistically linked to the underlying histology of disease. Direct numerical simulation of ultrasonic wave propagation through multiphase heterogeneous materials such as tissues is challenging, however, since MS with mode conversion is a difficult problem to solve.

Histology-oriented, simulation-based methods developed thus far to model ultrasonic wave propagation through tissues have used the traditional assumptions of weak, single scattering (SS) and no shear waves in addition to other simplifications. Padilla *et al.* (2003) used tomographic reconstructions and a Fourier transform method to calculate the scattering from 3D bone microstructures, whereas Hunt *et al.* (2002) modeled tissues by limiting the microstructures to two-dimensional, semi-ordered cell arrangements. Mamou *et al.* (2005, 2008) used a sequence of stained histologic sections to construct a microscopic, 3D acoustic impedance map (3DZM) of a tissue region to identify ultrasonic scattering sites. Scattering from these sites was then modeled using form factors as previously discussed. The 3DZM model incorporated tissue structure into the scattering calculations by varying the size, number density, and acoustic impedance of the scatterers.

Acoustic scattering from a single, nucleated spherical cell has been modeled with a finite element method (Falou *et al.*, 2006), but the modeling of large numbers of cells is currently too computationally intensive for finite element, boundary element, or finite-difference time-domain methods. In contrast, multipole methods are more numerically efficient for modeling wave scattering from spheres since they encode the wave field information over all space into a relatively small set of multipole expansion coefficients. However, most multipole approaches have modeled only longitudinal wave propagation in random dispersions of simple spheres (Gumerov and Duraiswami, 2002, 2005) or elastic wave propagation in ordered lattices of spherical inclusions (Liu *et al.*, 2000).

Multipole expansions have been used to model the acoustic resonances of erythrocytes (Zinin, 1992) and bacterial cells (Zinin *et al.*, 2005) using a spherical shell model. These models were restricted to SS from non-nucleated cells, however, and employed a thin spherical shell for the cell membrane. Recently, an iterative multipole model was developed to simulate full elastic wave propagation through arbitrary suspensions of spherical particles (Doyle, 2005, 2006). This model was recently extended to simulate wave propagation in tissues (Doyle and Warnick, 2006) and hierarchical tissue structures modified by neoplastic processes (Doyle *et al.*, 2007a, 2007b). In contrast to the erythrocyte and bacterial cell models, the iterative-multipole-based tissue models simulated MS between nucleated cells, and used a thick shell and core structure to model the cytoplasm and nucleus.

This article presents the theory and preliminary results from the application of the iterative multipole approach to the simulation of elastic wave propagation in nucleated cells and in tissues with simple and complex structures. The approach models the cells as concentric spheres, with the inner sphere comprising the nucleus, the outer sphere comprising the cytoplasm, and the medium between the cells comprising the extracellular matrix (ECM). MS between subcellular, cellular, and tissue structures can therefore be modeled with the use of boundary conditions (BCs), matrix solutions, and iteration. Simulations of ordered, random, and hierarchical cell arrangements were performed to demonstrate the feasibility of the models and to determine which histological

changes in cell and tissue structure were discernible in ultrasonic spectra. Various morphologies were modeled, including tissues with nodules of different structure, cavities of varying size, and cells with different nuclei diameters. Each of these morphologies produced distinct spectral signatures that could potentially be used for diagnostic purposes. Extension of the models to cell and tissue structures of greater complexity is discussed. Possible applications and methods for implementation are also reviewed.

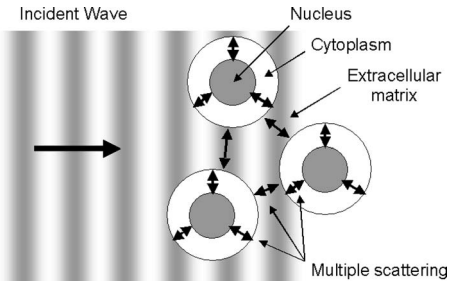


FIG. 1. Diagram of scattering processes and cellular structures simulated with the present work.

II. THEORY

A. Vector multipole functions

In the iterative multipole model each of the cell components—nucleus, cytoplasm, and ECM—is assumed to be a linear, homogeneous medium described by the Navier equation

$$\rho\frac{\partial^2\mathbf{u}}{\partial t^2} = (\lambda + 2\mu)\,\nabla\,(\nabla\cdot\mathbf{u}) - \mu\,\nabla\,\times(\nabla\times\mathbf{u}), \tag{1}$$

where λ and μ are Lamé's constants, ρ is the density, and $\mathbf{u}$ is the displacement vector. The Navier equation for elastic waves is solved by separating the displacement vector into longitudinal and transverse components. This yields two equations that can be reformulated as vector Helmholtz equations by imposing a harmonic time dependency to represent the wave behavior of the fields. Solutions to the vector Helmholtz equations in spherical coordinates are the vector multipole functions, which are complete and orthogonal multipole expansions for $n=0$ to ∞ and $m=-n$ to $+n$:

$$\mathbf{U} = \sum_{n=0}^{\infty}\sum_{m=-n}^{+n}\left(\sqrt{\frac{n}{2n+1}}z_{n-1}(kr)\mathbf{Y}_{nm}^{n-1}(\theta,\varphi) + \sqrt{\frac{n+1}{2n+1}}z_{n+1}(kr)\mathbf{Y}_{nm}^{n+1}(\theta,\varphi)\right), \tag{2}$$

$$\mathbf{V} = \sum_{n=0}^{\infty}\sum_{m=-n}^{+n}\left(i\sqrt{\frac{n+1}{2n+1}}z_{n-1}(kr)\mathbf{Y}_{nm}^{n-1}(\theta,\varphi) - i\sqrt{\frac{n}{2n+1}}z_{n+1}(kr)\mathbf{Y}_{nm}^{n+1}(\theta,\varphi)\right), \tag{3}$$

$$\mathbf{W} = \sum_{n=0}^{\infty}\sum_{m=-n}^{+n} z_n(kr)\mathbf{Y}_{nm}^{n}(\theta,\varphi). \tag{4}$$

U, **V**, and **W** correspond to the longitudinal, shear-electric, and shear-magnetic displacement fields, respectively. **V** and **W** are also the electric and magnetic multipole solutions to the Maxwell equations. These designations were used to distinguish the shear waves since the horizontal and vertical conventions normally used to differentiate the two types of shear fields are ambiguous in a spherical coordinate system. The radial function $z_n(kr)$ represents a spherical Bessel function $[j_n(kr)]$, a spherical Hankel function of the first kind $[h_n^{(1)}(kr)]$, or a spherical Hankel function of the second kind $[h_n^{(2)}(kr)]$. $\mathbf{Y}_{nm}^{l}(\theta,\varphi)$ is a pure-orbital vector spherical harmonic. Details of the Navier equation solutions, vector spherical harmonics, and vector multipole functions are given in Doyle (2006).

B. Single scattering from cells

The cells were approximated as two concentric spheres, with the inner sphere representing the cell nucleus, the outer sphere representing the cell cytoplasm, and the medium exterior to the spheres representing the ECM in the tissue. The simulations were performed by imposing an incident plane wave on the cells and then computing the MS both within and between the cells (Fig. 1). MS between cells was computed using an iterative method in conjunction with addition theorems, whereas scattering inside the cell was solved with the use of BCs that formed a set of matrix equations.

Since the ECM and cell components could be either fluid or solid (including viscoelastic), a total of eight different sets of BC solutions were required. The BCs were therefore first formulated for full elastic wave scattering (solid inner sphere, solid outer sphere, and solid matrix) since this comprised the most inclusive set of equations. The other seven sets of BCs were then obtained by subsequent modification of the full elastic wave solution. The BC solutions account for mode conversion of both longitudinal to shear waves and shear to longitudinal waves whenever a solid-solid, solid-fluid, or fluid-solid interface is present.

For scattering from a sphere comprised of a solid core and a solid shell in a solid matrix, there will be an incoming incident field in the matrix, an incident field within the shell, a refracted core field, an outgoing scattered field within the shell, and an outgoing scattered field in the matrix for each of the **U**, **V**, and **W** wave fields. Each wave field component will also have an associated amplitude coefficient. Table I shows the designations of the amplitudes for the corresponding fields. Figure 2 shows the relationship between each of the field components and the coefficients.

Given an arbitrary incident field, the amplitude coefficients of the interior and scattered fields are found by solving the BCs on the surface of the core and shell. The BCs provide a set of 12 linear equations for the 12 unknown coefficients. Six of these equations are obtained from continuity of the displacements:

$$\mathbf{u}^{\text{incident}} + \mathbf{u}^{\text{scattered}} = \mathbf{u}^{\text{interior}}. \tag{5}$$

The other six equations are derived from continuity of the stresses, where σ is the stress tensor:

TABLE I. Amplitude coefficients for wave fields in the single-cell scattering problem.

	Incident field	Incident shell field	Refracted core field	Scattered shell field	Scattered field
Longitudinal (**U**)	A_{nm}	As_{nm}	D_{nm}	Gs_{nm}	G_{nm}
Shear-electric (**V**)	B_{nm}	Bs_{nm}	E_{nm}	Hs_{nm}	H_{nm}
Shear-magnetic (**W**)	C_{nm}	Cs_{nm}	F_{nm}	Is_{nm}	I_{nm}

$$\sigma^{\text{incident}} + \sigma^{\text{scattered}} = \sigma^{\text{interior}}. \tag{6}$$

Equations for continuity of the radial stresses σ_{rr} and tangential stresses $\sigma_{r\theta}$ and $\sigma_{r\varphi}$ are solved separately. The stresses are related to the displacements by the following:

$$\sigma_{rr} = \lambda(\nabla \cdot \mathbf{u}) + 2\mu\frac{\partial u_r}{\partial r}, \tag{7}$$

$$\sigma_{r\theta} = \mu \cdot \left(\frac{1}{r}\frac{\partial u_r}{\partial \theta} + \frac{\partial u_\theta}{\partial r} - \frac{u_\theta}{r}\right), \tag{8}$$

$$\sigma_{r\varphi} = \mu \cdot \left(\frac{1}{r\sin\theta}\frac{\partial u_r}{\partial \varphi} - \frac{u_\varphi}{r} + \frac{\partial u_\varphi}{\partial r}\right). \tag{9}$$

Orthogonality conditions for the vector and scalar spherical harmonics are used to eliminate the angular dependence from the displacement and stress BCs, respectively:

$$\int_0^{2\pi} d\varphi \int_0^{\pi} d(\cos\theta)\mathbf{Y}_{NM}^{L} * (\theta,\varphi)\mathbf{Y}_{nm}^{l}(\theta,\varphi) = \delta_{Ll}\delta_{Nn}\delta_{Mm}, \tag{10}$$

$$\int_0^{2\pi} d\varphi \int_0^{\pi} d(\cos\theta)Y_{NM} * (\theta,\varphi)Y_{nm}(\theta,\varphi) = \delta_{Nn}\delta_{Mm}, \tag{11}$$

where $*$ denotes the complex conjugate of the function. The resulting equations are linear and can be cast into a matrix form that relates the 12 unknown coefficients (scattered and refracted fields) to the 3 known coefficients (incident fields).

The derivations for 3 of the 12 BC equations are presented in Appendix A as an example. For brevity, the derivations for the remaining equations are omitted, and only the results are presented. The displacement BC on the outer surface of the shell (outer sphere) yields three equations:

$$\begin{aligned} &As_{NM}\eta_1(g^{\text{shell}}) + Bs_{NM}\eta_2(g^{\text{shell}}) + Gs_{NM}\eta_1(h^{\text{shell}}) \\ &\quad + Hs_{NM}\eta_2(h^{\text{shell}}) - G_{NM}\eta_1(h^{\text{ext}}) - H_{NM}\eta_2(h^{\text{ext}}) \\ &\quad = A_{NM}\eta_1(g^{\text{ext}}) + B_{NM}\eta_2(g^{\text{ext}}), \end{aligned} \tag{12}$$

$$\begin{aligned} &As_{NM}\eta_4(g^{\text{shell}}) + Bs_{NM}\eta_5(g^{\text{shell}}) + Gs_{NM}\eta_4(h^{\text{shell}}) \\ &\quad + Hs_{NM}\eta_5(h^{\text{shell}}) - G_{NM}\eta_4(h^{\text{ext}}) - H_{NM}\eta_5(h^{\text{ext}}) \\ &\quad = A_{NM}\eta_4(g^{\text{ext}}) + B_{NM}\eta_5(g^{\text{ext}}), \end{aligned} \tag{13}$$

$$\begin{aligned} &Cs_{NM}\eta_3(g^{\text{shell}}) + Is_{NM}\eta_3(h^{\text{shell}}) - I_{NM}\eta_3(h^{\text{ext}}) \\ &\quad = C_{NM}\eta_3(g^{\text{ext}}). \end{aligned} \tag{14}$$

The η-symbols are functions of spherical Bessel functions $[j=j_n(kr)]$ and spherical Hankel functions $[h=h_n^{(1)}(kr)$ and $g=h_n^{(2)}(kr)]$, and are presented in Appendix B. The displacement BC on the inner surface of the shell (outer surface of the core) also yields three equations:

$$\begin{aligned} &As_{NM}\eta_{11}(g^{\text{shell}}) + Bs_{NM}\eta_{12}(g^{\text{shell}}) + Gs_{NM}\eta_{11}(h^{\text{shell}}) \\ &\quad + Hs_{NM}\eta_{12}(h^{\text{shell}}) - D_{NM}\eta_{11}(j^{\text{core}}) - E_{NM}\eta_{12}(j^{\text{core}}) \\ &\quad = 0, \end{aligned} \tag{15}$$

$$\begin{aligned} &As_{NM}\eta_{14}(g^{\text{shell}}) + Bs_{NM}\eta_{15}(g^{\text{shell}}) + Gs_{NM}\eta_{14}(h^{\text{shell}}) \\ &\quad + Hs_{NM}\eta_{15}(h^{\text{shell}}) - D_{NM}\eta_{14}(j^{\text{core}}) - E_{NM}\eta_{15}(j^{\text{core}}) \\ &\quad = 0, \end{aligned} \tag{16}$$

$$Cs_{NM}\eta_{13}(g^{\text{shell}}) + Is_{NM}\eta_{13}(h^{\text{shell}}) - F_{NM}\eta_{13}(j^{\text{core}}) = 0. \tag{17}$$

Two equations result from the radial stress BC, one from the outer surface of the shell [Eq. (18)] and one from the surface of the core [Eq. (19)]:

$$\begin{aligned} &As_{NM}\eta_6(g^{\text{shell}}) + Bs_{NM}\eta_7(g^{\text{shell}}) + Gs_{NM}\eta_6(h^{\text{shell}}) \\ &\quad + Hs_{NM}\eta_7(h^{\text{shell}}) - G_{NM}\eta_6(h^{\text{ext}}) - H_{NM}\eta_7(h^{\text{ext}}) \\ &\quad = A_{NM}\eta_6(g^{\text{ext}}) + B_{NM}\eta_7(g^{\text{ext}}), \end{aligned} \tag{18}$$

$$\begin{aligned} &As_{NM}\eta_{16}(g^{\text{shell}}) + Bs_{NM}\eta_{17}(g^{\text{shell}}) + Gs_{NM}\eta_{16}(h^{\text{shell}}) \\ &\quad + Hs_{NM}\eta_{17}(h^{\text{shell}}) - D_{NM}\eta_{16}(j^{\text{core}}) - E_{NM}\eta_{17}(j^{\text{core}}) \\ &\quad = 0 \end{aligned} \tag{19}$$

Two equations result from the tangential stress BC on the outer surface of the shell:

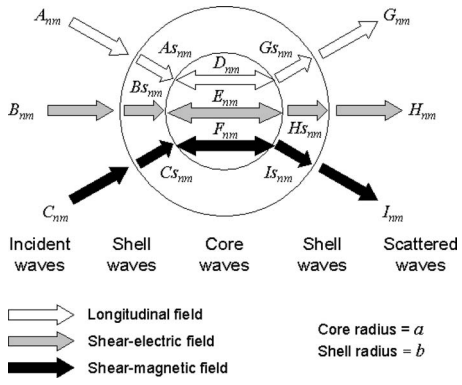


FIG. 2. Diagram of incident, refracted, and scattered elastic waves for scattering from a cell with a shell-core (cytoplasm-nucleus) structure with associated amplitude coefficients for the (n,m) multipole moment.

TABLE II. Sizes of matrix equations (number of BCs or unknown coefficients) for various cell configurations and an arbitrary incident wave. The matrix system for the **W** field is listed separately since it is decoupled from the matrix system for the **U** and **V** fields.

Matrix	Shell	Core	System size: **U** and **V** fields	System size: **W** field
Fluid	Fluid	Fluid	4	0
Fluid	Fluid	Solid	5	0
Fluid	Solid	Fluid	6	0
Solid	Fluid	Fluid	5	1
Fluid	Solid	Solid	7	0
Solid	Fluid	Solid	6	1
Solid	Solid	Fluid	7	3
Solid	Solid	Solid	8	4

$$As_{NM}\eta_9(g^{\text{shell}}) + Bs_{NM}\eta_{10}(g^{\text{shell}}) + Gs_{NM}\eta_9(h^{\text{shell}}) + Hs_{NM}\eta_{10}(h^{\text{shell}}) - G_{NM}\eta_9(h^{\text{ext}}) - H_{NM}\eta_{10}(h^{\text{ext}}) = A_{NM}\eta_9(g^{\text{ext}}) + B_{NM}\eta_{10}(g^{\text{ext}}), \quad (20)$$

$$Cs_{NM}\eta_8(g^{\text{shell}}) + Is_{NM}\eta_8(h^{\text{shell}}) - I_{NM}\eta_8(h^{\text{ext}}) = C_{NM}\eta_8(g^{\text{ext}}). \quad (21)$$

Finally, two equations result from the tangential stress BC on the surface of the core:

$$As_{NM}\eta_{19}(g^{\text{shell}}) + Bs_{NM}\eta_{20}(g^{\text{shell}}) + Gs_{NM}\eta_{19}(h^{\text{shell}}) + Hs_{NM}\eta_{20}(h^{\text{shell}}) - D_{NM}\eta_{19}(j^{\text{core}}) - E_{NM}\eta_{20}(j^{\text{core}}) = 0, \quad (22)$$

$$Cs_{NM}\eta_{18}(g^{\text{shell}}) + Is_{NM}\eta_{18}(h^{\text{shell}}) - F_{NM}\eta_{18}(j^{\text{core}}) = 0. \quad (23)$$

Equations (12)–(23) form a system of 12 linear equations that can be solved as a matrix equation for the 12 unknown coefficients of multipole order N, M for the solid core-solid shell-solid matrix configuration. Note, however, that the four equations for the shear-magnetic field [Eqs. (14), (17), (21), and (23)] are decoupled from the longitudinal and shear-electric equations, allowing them to be solved as separate matrix equations. The sizes of the matrix equations are dependent on the cell and matrix material properties, and are listed in Table II.

C. Multiple scattering between cells

To simulate MS in tissues, the scattered waves from each cell propagate to the other cells and modify the other cells' scattering interactions (Fig. 1). Since the vector multipole functions are specific to each cell's coordinate system, the scattered wave fields from each cell must be transformed into the coordinate systems of the other cells in order to compute the modified interactions. These transformations are also known as the scalar and vector addition theorems for spherical wave functions, and have been previously described for the **U**, **V**, and **W** wave fields (Doyle, 2005, 2006; Doyle *et al.*, 2007c).

The iterative multipole model first calculates the scattered wave fields for each cell in the simulated tissue arising from an initial plane wave. The scattered wave fields are then translated between all cell pairs and summed at each cell. The new incident wave field (initial plane wave+scattered wave fields) is then used to compute a revised set of scattered wave fields from each cell. This process is repeated iteratively until the scattered wave fields converge (i.e., no change in wave field amplitudes between consecutive iterations). Each iteration represents a successive order of scattering, where the first iteration is first-order MS, the second iteration is second-order MS, and so forth.

After the iterations converge, the Cartesian components (x, y, and z amplitudes) of the **U**, **V**, and **W** multipole fields are then calculated for a two-dimensional array of grid points either inside or outside the tissue. Inside the tissue, these grid point values are used to generate cross-sectional images of the ultrasonic wave field distribution at the microscopic level. Outside the tissue, the grid point values are averaged across the grid to simulate the face of a virtual transducer and yield the wave amplitude and phase of the transmitted or reflected ultrasound. Frequency spectra are generated by sampling the amplitudes and phases across a frequency band. In this study, ultrasonic backscatter spectra were computed up to 100 MHz with frequency steps of 0.25 MHz per data point for single-cell simulations and 0.50 MHz per data point for multiple-cell simulations.

Pulse velocity and attenuation can also be modeled by using a Fourier transform method to convert wavelet functions in the frequency domain to pulse functions in the spatial domain (Doyle, 2006). Although these parameters were not calculated in this study, the iterative multipole model has been shown to be sensitive to ultrasonic velocity in dispersions containing up to 736 glass particles in a viscoelastic polymer (Doyle, 2006). The ability to simulate ultrasonic velocity and attenuation as a function of microscopic tissue structure would potentially be very useful for new diagnostic methods such as ultrasonic tomography and intravascular ultrasound.

An incident plane wave was used for the simulations since the modeled tissue region was small (280 μm across) compared to typical focal spot sizes for ultrasonic beams ($\geq$1 mm). More complex incident wave fronts, such as spherical or cylindrical wave fronts, can be simulated with the iterative multipole model by calculating a different set of multipole expansion coefficients for the incident waves with the use of an appropriate partial wave expansion. The change in the moment distribution for the incident waves would be expected to change the scattering processes in the tissues since the multipole moments (and thus the resulting wave properties) have strong radial and angular dependencies. New ultrasound-tissue responses may therefore be revealed due to, for example, changes in dipole scattering relative to monopole scattering, and would include the appearance of cell or nucleus resonances at new frequencies.

Computation limits for the simulations include the maximum multipole expansion order ($n_{\max}$) for computing the wave field and addition theorem expansions, and the conver-

TABLE III. Typical tissue and cell properties used for multiple-cell simulations. Values in parentheses for the shear wave speed are the ranges used in the single-cell tests.

Property	Extracellular matrix	Cytoplasm	Nucleus
Longitudinal wave speed (m/s)	1570	1483	1509
Shear wave speed (m/s)	106 (0–200)	0 (0–200)	0 (0–560)
Density (g/cm^3)	1.06	0.998	1.43

gence criteria (precision limit) for stopping the iterations. Computer algorithms for the model were written, debugged, and compiled in FORTRAN 90 with double precision. Simulations were performed on a single-CPU desktop computer with 1.0 Gbyte RAM and a 1.66-GHz processor. Computation times varied from 1 to 54 h. Single-cell simulations were performed to a multipole expansion order of $n_{max}=10$, whereas multiple-cell simulations were primarily conducted at $n_{max}=3$. Convergence studies with the multiple-cell models were consistent with previous simulation results from glass particles in a viscoelastic polymer, which showed convergence of the ultrasonic velocity by $n_{max}=3$ (Doyle, 2006).

Calculation of addition theorem coefficients for all cell pairs in the virtual tissue is computationally intensive and exceeds the dynamic memory capabilities of most desktop computers. The MS computations were therefore made more efficient by using a method described by Doyle *et al.* (2007c) that creates a spatially discretized template for the translation vectors between cells. This template allows the approximation of addition theorem coefficients for a large number of cell-to-cell scattering geometries without having to directly calculate the coefficients for each cell pair. Testing showed that this method provided sufficiently accurate results without compromising the complex nature of the microstructures.

Tissue structures were generated from spherical cells with a monomodal size distribution (20 μm in diameter) but with a variable nucleus size (8–14 μm in diameter). Both ordered [hexagonal-close-packed (hcp)] and disordered (random) tissue structures were created with cell fractions of 0.30–0.60. The random tissue structures were constructed from a Monte Carlo method developed to compare the structures of simple liquids to random close packings of particles (Finney, 1970). The simulated tissues were configured into a cylindrical shape of fixed volume (0.0172 mm^3), with the cylinder's diameter (0.28 mm or 280 μm) equal to its length and its axis aligned with the direction of the wave propagation (z axis). The tissues contained from 1211 to 2137 nucleated cells. Hierarchical tissue structures were additionally constructed from the random cell packings that modeled structures with either tissue cavities or nodes of cells differing from the surrounding cells. An artifact created by the simulation geometry was the presence of strong interference fringes in the ultrasonic backscatter spectra. These were produced by the separation between the virtual transducer grid and tissue cylinder, and were removed from the spectra by low-pass filtering.

Table III presents the elastic wave properties used for the ECM, cytoplasm, and nucleus. Estimated values were used since the properties of the cell components are not well known and vary widely in literature. The properties of freshly excised myocardium were chosen for the ECM since they are intermediate between those of stiff biomaterials such as collagen and those of softer tissues in the body (Yang *et al.*, 2006). The cytoplasm was given water-like properties with a room-temperature wave speed to correspond to published measurements of cell cultures, tissue cultures, and excised tissues. The properties of the nucleus were those used by Baddour *et al.* (2005) to model ultrasonic scattering from single cells, and were obtained from typical eukaryotic nuclear properties and measurements from condensed cell nuclei. In the single-cell simulations the shear wave speed of all three cell components was varied one at a time to determine the influence of shear properties on the ultrasonic scattering. The effects of varying the shear wave speeds of the cytoplasm and nucleus were quite large for single-cell scattering, but became insignificant in multiple-cell simulations. They were therefore set to zero for the multiple-cell computations. Although viscoelastic (attenuation) properties are an important consideration for ultrasonic propagation in tissues, they were not included in these initial trials.

III. RESULTS

A. Single-cell simulations

Baseline properties for the single-cell simulations included a cell diameter of 20 μm, a nucleus diameter of 10 μm, longitudinal wave speeds and densities as given in Table III, and shear wave speeds of 106, 0, and 560 m/s for the ECM, cytoplasm, and nucleus, respectively. Wave field images were generated for various incident wave frequencies to visualize the ultrasonic scattering from a single cell with baseline properties. Figure 3 shows the scattered longitudinal wave with the incident plane wave removed for clarity. Ultrasonic backscatter dominates in the image for 65 MHz [Fig. 3(a)], but forward scattering dominates in the image for 70 MHz [Fig. 3(b)]. These images correspond to the third peak (enhanced backscatter) and second trough (enhanced forward scatter) observed in the ultrasonic backscatter spectrum in Fig. 4(a) for a cell with the same baseline properties (cytoplasm $c_{shear}=0$ m/s).

Ultrasonic backscatter spectra were additionally simulated to study the effects of parameter variations on ultrasonic scattering from single cells. The varied parameters included nucleus diameter, cell diameter, cytoplasm density, and the shear wave speed of all three cell components. Varying the cell diameter and cytoplasm density had very little effect on the spectra. In contrast, variations in the shear wave speeds from baseline produced striking changes in the spectra [Fig. 4(a)]. Sharp peaks and troughs due to resonances were observed in the simulated spectra whenever a fluid-solid interface was present. The spectra associated with a fluid ECM differed from those of a solid ECM by showing an abrupt frequency shift in the resonance structure. This shift can be ascribed to a change in the fluid-solid interface that is dominant in the scattering process. When the ECM becomes fluid ($c_{shear}=0$ m/s), this interface changes from the ECM-cytoplasm boundary to the cytoplasm-nucleus

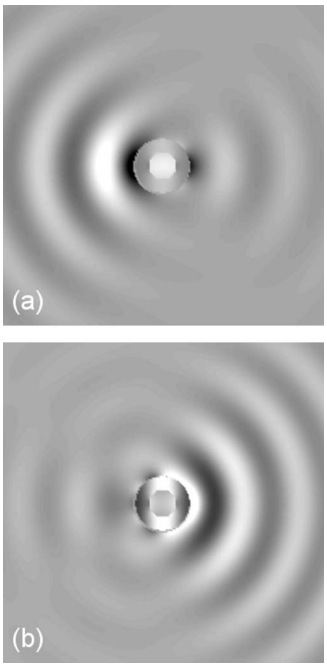


FIG. 3. Scattered wave field images for a longitudinal wave incident on a 20-μm-diameter cell with a 10-μm-diameter nucleus and embedded in an ECM. Strong backscattering is observed at 65 MHz (a) in contrast to strong forward scattering at 70 MHz (b). Wave propagation is from left to right. Wave fields inside the cell have been attenuated in the images to visualize the external scattered wave fields.

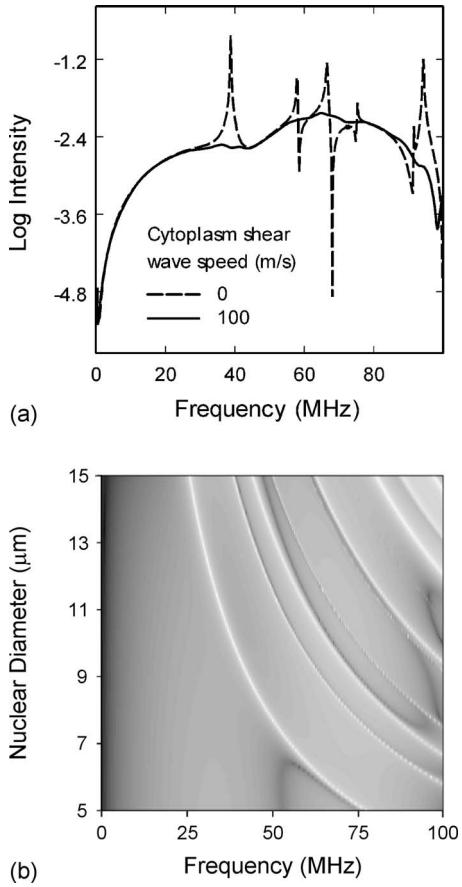


FIG. 4. (a) Ultrasonic backscatter spectra for a single 20-μm-diameter cell with a 10-μm-diameter nucleus, where the ECM and nucleus have elastic properties and the cytoplasm has shear wave speeds of 0 and 100 m/s. (b) Contour plot of frequency versus nuclear diameter for ultrasonic backscatter spectra of a 20-μm-diameter cell embedded in an ECM.

boundary. Variations in the ECM shear wave speed above 0 m/s produced very little effect in the spectra, however.

In contrast to the ECM, increasing the shear wave speed of the cytoplasm caused the sharp peaks and troughs in the spectra to vanish [Fig. 4(a)] by eliminating both the inner and outer fluid-solid interfaces. Varying the shear wave speed of the nucleus had less effect on the spectra than varying cytoplasm shear wave speed. Although the sharp peaks disappear for a solid cytoplasm and nucleus, they reappear when the nucleus becomes fluid. An important conclusion from these results is that the resonance peaks disappear when sufficient shear moduli are introduced to all three cell components. Varying the nucleus diameter also produced significant changes in the ultrasonic spectra [Fig. 4(b)]. The spectra in Fig. 4(b) were modeled using a cell with baseline properties except for a variable nucleus diameter, and show that the spectral peaks decrease in frequency with increasing nucleus diameter. This response is consistent with the peaks originating from a resonance process.

The single-cell simulations were also used to validate the theory and algorithms for the tissue model. By setting the wave speeds and densities of the cytoplasm and nucleus (or ECM and cytoplasm) equal to each other, a simple fluid or elastic sphere can be simulated in a fluid or elastic matrix. These simulations were performed and the model was verified by comparing the results to the analytical solutions of Anderson (1950), Faran (1951), and Hickling (1962).

B. Multiple-cell simulations

All of the multiple-cell simulations were performed with a solid ECM (c_{shear}=106 m/s). Although baseline shear wave properties were used that gave a solid-fluid interface between the ECM and cytoplasm (cytoplasm c_{shear}=0 m/s), no sharp resonance-peak structures were observed in the ultrasonic backscatter spectra. A range of shear wave velocities for the cell cytoplasm and nucleus was therefore tested to determine the sensitivity of the simulations to these cell properties. Contrary to the single-cell results, no change was observed in the ultrasonic backscatter spectra resulting from increasing the shear wave velocity from 0 to 50 m/s. The shear wave velocities were therefore left at 0 m/s for both the cytoplasm and nucleus.

The numerical simulation of a small random cluster of 327 cells has been previously published (Doyle *et al.*, 2007b), and the results show mode conversion of longitudinal waves to shear waves upon scattering at the plasma membrane. Although the shear waves will quickly attenuate in tissue due to the viscoelasticity of biopolymers and organic structures, the mode conversion process indicates a loss mechanism for ultrasonic energy in tissue that has not been properly accounted for with a high level of detail and accuracy. Mode conversion additionally provides a mechanism for generating shear forces at the plasma membrane that may not be possible to generate if only fluid constituents were considered. In addition to mode conversion, images of the random MS show localized areas between cells and close to

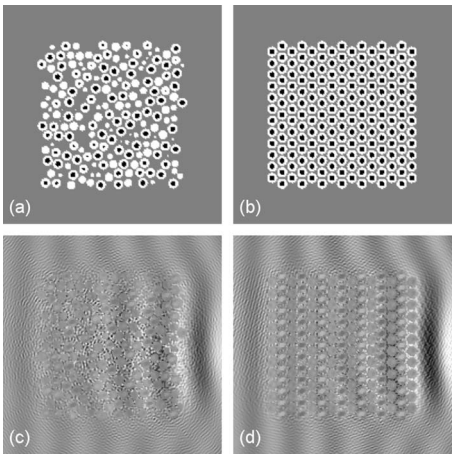


FIG. 5. Top: Two-dimensional slices through simulated tissues of approximately 2100 cells with random (a) and hcp (b) cell packings at 0.50 cell volume fraction. Gray, white, and black represent the ECM, cytoplasm, and nuclei, respectively. Bottom: Scattered wave field images for an incident 20-MHz longitudinal wave in random (c) and hcp (d) cell packings as shown in (a) and (b), respectively. Wave propagation is from left to right.

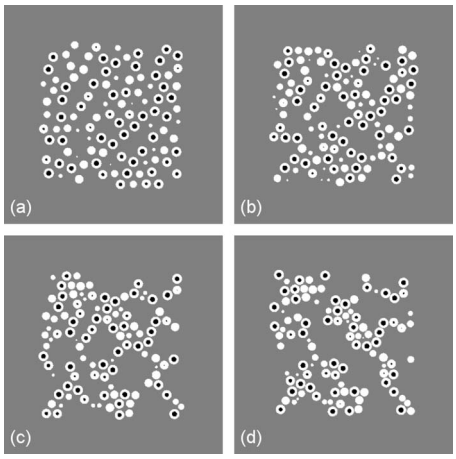


FIG. 7. Simulated tissue structures where the cells display various degrees of aggregation into a foam-like structure with spherical cavities. Local cell densities are (a) 30% (no aggregation), (b) 40%, (c) 50%, and (d) 60% (maximum aggregation).

the plasma membrane that display enhanced ultrasonic amplitudes up to 20% greater than the incident wave.

Comparative simulations were performed on random and hcp cell packings in order to determine the effects of cell order and tissue structure on ultrasonic backscatter spectra. Such a comparison is not unrealistic since the hcp structure could approximate tissues with hexagonal cell packings such as found in the liver. Each packing had approximately 2100 cells with a 20 μm cell diameter, a 10 μm nuclear diameter, and a 0.50 cell volume fraction [Figs. 5(a) and 5(b)]. The 20-MHz wave field images in Figs. 5(c) and 5(d) display significant forward scattering for both structures. Although the scattered waves appear similar for both the random and hcp structures, the hcp structure shows slightly more coherent scattering.

The ultrasonic backscatter spectra of the random and hcp structures, Fig. 6, show significant disparities. The greatest differences arise from the peaks observed in the hcp spectrum at 39 and 74.5 MHz and in the random spectrum at 18 and 68 MHz. The hcp peaks can be ascribed to acoustic band gaps formed by the cell lattice. Assuming a tissue wave speed averaged over the ECM, cytoplasm, and nucleus, the 39- and 74.5-MHz peaks correspond to wavelengths of 39.2 and 20.5 μm, respectively. These wavelengths are approximately four times the diameter of the nucleus (10 μm) and four times the thickness of the cytoplasm layer between the nucleus and plasma membrane (5 μm), which form periodic layers stacked along the z (horizontal) axis. These layers act as quarter-wave plates that preferentially reflect the acoustic waves propagating in the z direction. The 18- and 68-MHz peaks in the random spectrum may arise from single-cell resonance effects that are suppressed by coherent lattice interactions in the hcp structure.

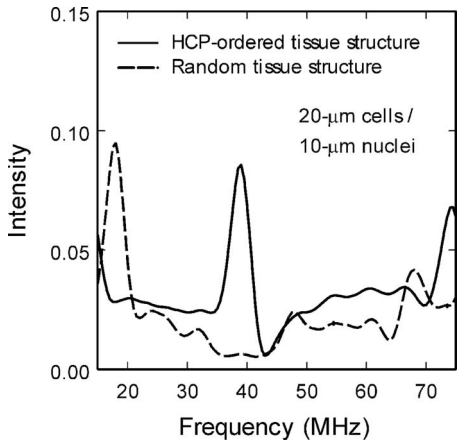
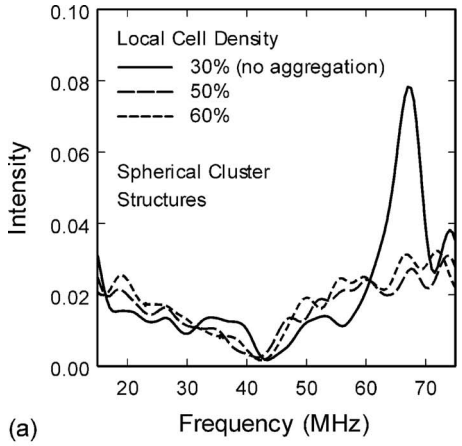


FIG. 6. Comparison of ultrasonic backscatter spectra for simulated random and hcp cell packings. The hcp cell packing displays enhanced backscattering peaks at 39 and 74.5 MHz that can be associated with acoustic band gaps.

Comparative simulations were performed on tissue structures that exhibited more complex variations in microstructure than merely ordered versus disordered states. These simulations examined the effects of cell aggregation on the backscatter spectra, where initially uniform but random distributions of cells coalesce into either spherical clusters or foam-like structures surrounding spherical cavities. This aggregation forms a hierarchical microstructure since the derivative clusters or cavities represent a higher level of structure in tissue. Figure 7 shows representative cell microstructures with increasing aggregation into a foam-like structure with each panel. Although the total cell fraction (and therefore cell number) is kept constant, the local cell packing density increases as the cells coalesce. The local cell densities in Fig. 7 are (a) 30% (no aggregation), (b) 40%, (c) 50%, and (d) 60% (maximum aggregation or local cell packing).

Simulations were performed on five different microstructures for each local cell density (aggregation level) to provide averaged spectra and standard deviations for determining the statistical significance of the results. The microstructures varied by the locations of individual cells and clusters/cavities. The degree of randomness in the cell and cluster/cavity locations was kept constant, however (i.e., same radial distribution functions), as were the cell diam-



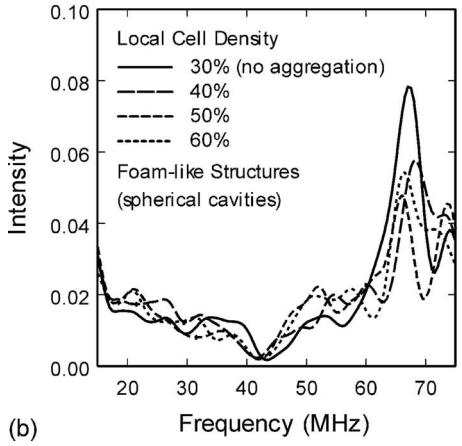
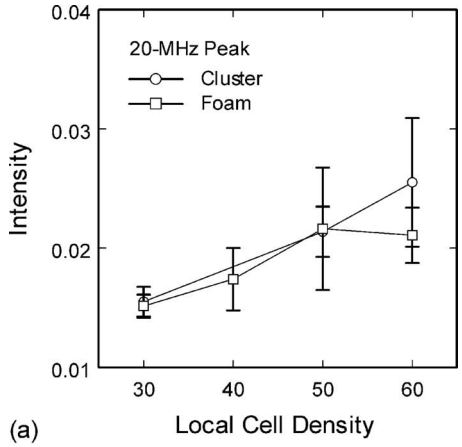


FIG. 8. Averaged ultrasonic backscatter spectra for simulated cell packings with progressively aggregated microstructures for (a) aggregation into spherical clusters and (b) aggregation into foam-like structures with spherical cavities. Each curve was averaged from simulations of five tissue structures where the cell and cluster/cavity locations were varied.



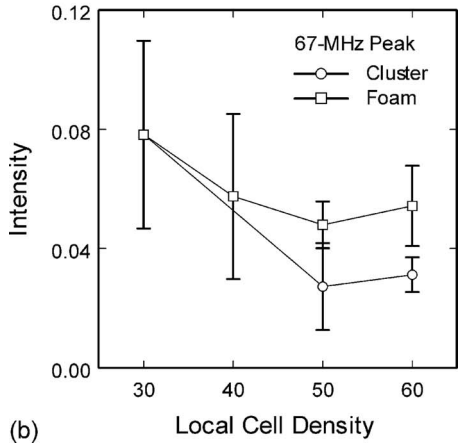


FIG. 9. Trends of peaks near (a) 20 MHz and (b) 67 MHz as a function of local cell density (level of aggregation) for aggregation into spherical clusters and aggregation into foam-like structures with spherical cavities. The error bars are the standard deviations in the spectra arising from the five structural variations.

eters, nuclear diameters, and cluster/cavity diameters. Figure 8 compares the average spectra for various types and levels of aggregation. The only feature that appeared to change substantially was the peak at 67 MHz, which decreased partially for foam aggregation and more fully for cluster aggregation. Figure 9 displays the trends for the 67-MHz peak as a function of local cell density. The trends for a smaller peak near 20 MHz (19 and 21 MHz for cluster and foam aggregation, respectively) are also shown for comparison. Although both peaks showed changes with aggregation, the uncertainties in the trends were on the same order as the trend changes.

The effects of cell structure on ultrasonic backscatter spectra were simulated by varying the size of the nucleus for all of the cells in a random tissue structure such as shown in Fig. 5(a) (Doyle *et al.*, 2007b). Such nuclear size variations can be used to model nuclear pleomorphism in tumors. The results showed statistically significant spectral changes between 20 and 56 MHz, with the development of a large peak at 47 MHz as the nucleus diameter increased from 8 to 14 μm (cell diameter=20 μm). The results showed a negative linear trend of spectral intensity with nuclear diameter in the 20–35-MHz band and a positive linear trend in the 40–56-MHz band.

The sensitivity of the numerical model to pathological processes such as lobular carcinoma *in situ* (LCIS), a type of breast cancer, was also explored by simulating the *in situ* growth of cancer cells in cavities within a tissue (Doyle *et al.*, 2007b). These processes combine changes in cell structure such as nuclear pleomorphism, changes in tissue structure such as the progressive filling of cavities, and hierarchical structures such as random cavities in a random distribution of cells. An initial cell structure with spherical cavities, as shown in Fig. 7(d), was used for the normal tissue structure. Microstructures were then generated to simulate the gradual accumulation of cancer cells within the cavities. The cancer cells were differentiated from the normal cells by having 40% larger nuclei (nuclear pleomorphism). The spectral results were similar to those of the nuclear pleomorphism results, showing the development of a large peak at 47 MHz. However, the LCIS simulation results differed from the nuclear pleomorphism results by showing no spectral changes in the 20–35-MHz band.

Another pathological process modeled was that of small clusters of cancer cells growing within a random, uniform tissue. This process is similar to invasive carcinoma where the growth of microtumors pushes the normal cells closer together and disrupts the normal histology. Figure 10 dis-

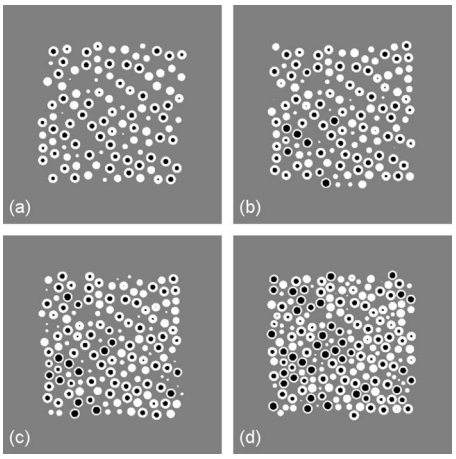


FIG. 10. Simulations of malignant cell invasion in a random tissue with 1231–2081 cells and 0% (a), 14% (b), 25% (c), and 40% (d) invasion of malignant cells growing as approximately spherical, randomly distributed clusters in the tissue. Note that the malignant cells displace the normal cells, increasing the total cell volume fraction from 0.30 (a) to 0.35 (b), 0.40 (c), and 0.50 (d). Malignant cells are differentiated from the normal cells by having 40% larger nuclei (nuclear pleomorphism).

plays one of the modeled tissue structures where each panel displays progressive growth of cancerous cell clusters and the resulting increase in cell density for the normal cells. As with previous simulations, five different microstructures were simulated for each stage of invasion where the cell and microtumor locations were varied. Figure 11 shows averaged spectra with standard deviations for 0% [Fig. 11(a)] and 40% [Fig. 11(b)] malignant cell invasion. Significant differences were observed in the spectra that were larger than the standard deviations.

The size of the standard deviations in Figs. 11(a) and 11(b) varies substantially with frequency. Although the standard deviations are computed from microstructures with variations in the microtumor and cell positions, the sizes of the microtumors and cells are kept constant, as are the sizes for the normal and pleomorphic nuclei. Therefore, spectral regions with small standard deviations most likely arise from resonances or other coherent scattering interactions with these structures at these frequencies. Spectral regions with large standard deviations can be attributed to frequencies not dominated by mesoscopic (microtumor) or microscopic (cell and nucleus) size interactions, and are therefore influenced mostly by random parameters such as the cell and microtumor arrangement. The standard deviations will most likely increase in more accurate simulations with polydisperse nuclei, cells, and microtumors. However, even spectral regions with relatively large standard deviations such as the 47-MHz peak in Fig. 11(b) may produce observable trends such as seen in Fig. 12(b).

Figure 12(a) compares the averaged spectra for all invasive stages modeled. Standard deviation bars have been omitted for clarity. Spectral peaks at 18 and 47 MHz appeared with increasing microtumor growth. In contrast, the peak at 67 MHz present in the normal tissue spectrum (0% malignant cell invasion) dropped in intensity between 14% and 25% invasion like a step function. The behavior of the 67-MHz peak was similar to that of the 67-MHz peak in the

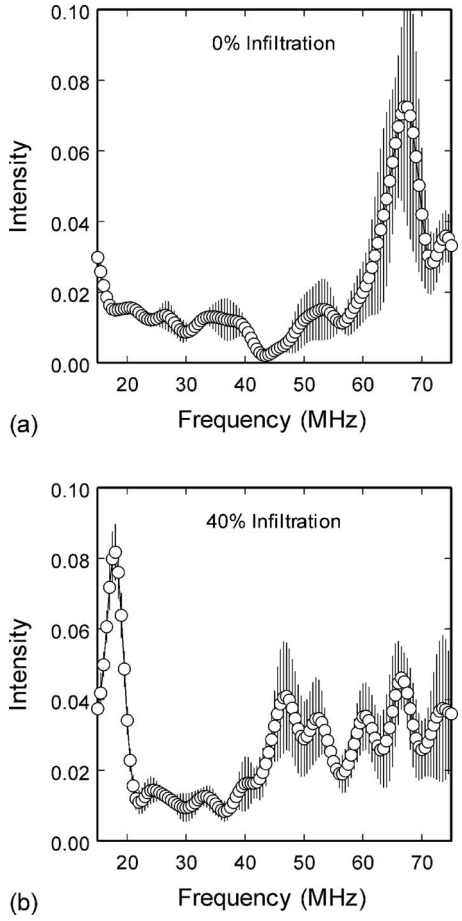


FIG. 11. Simulated spectra for (a) 0% and (b) 40% malignant cell invasion corresponding to the tissue structures in Figs. 10(a) and 10(d), respectively. Plots show average intensities (circles) and standard deviations (bars) for the simulation of five tissue structures where the cell and malignant cluster (microtumor) locations were varied.

aggregation simulations (Fig. 9), and had comparable standard deviations that were as large as the observed step change. The correlations of the 18- and 47-MHz peak intensities with malignant cell invasion were significant, however, as shown in Fig. 12(b). The 47-MHz peak displayed a mostly linear trend, whereas the 18-MHz peak trend appeared nonlinear. Using a linear fit and averaged standard deviation for the 47-MHz peak, a detection limit of approximately 8.5% tumor cell infiltration was estimated for the simulated region. Since ultrasound can be focused to regions as small as 2 mm (and smaller for higher frequencies), this detection limit would correspond to sub-millimeter tumor detection.

To determine the effects of MS in the multiple-cell simulations, SS simulations were performed on the 30% uniform cell distributions [Fig. 7(a)] and the 60% aggregated cell distributions. Figure 13 compares the averaged spectra between the MS and SS simulations. Differences between MS and SS become significant above 60 MHz for the 30% uniform tissues and above 45 MHz for the 60% aggregated tissues. Surprisingly, the less dense, less structured uniform tissues show a greater effect due to MS than the denser, more structured aggregate tissues.

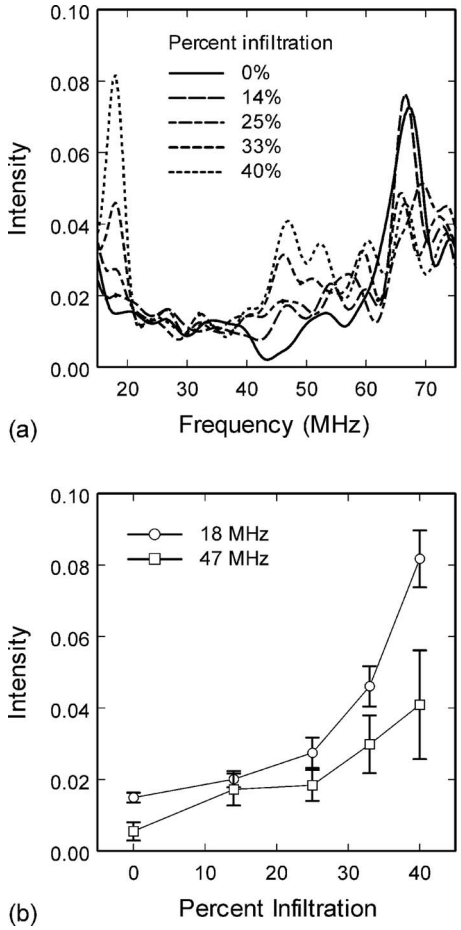


FIG. 12. (a) Model ultrasonic backscatter spectra for malignant cell invasion in random tissue averaged from simulations of five tissue structures where the cell and malignant cell clump locations were varied. (b) Intensity trends at 18 and 47 MHz from the spectra in Fig. 11(a). The error bars are the standard deviations in the spectra arising from the five structural variations.

IV. DISCUSSION

A. Single-cell simulations

The backscatter spectra from the single-cell simulations showed sharp peaks and troughs for cell structures that have a fluid-solid interface between the ECM and cytoplasm or cytoplasm and nucleus. The resonance structure vanishes when either the shear wave speed is increased above 0 m/s for the fluid component or decreased to 0 m/s for the solid component. The shear properties of the cytoplasm had the greatest effect on the fine structure in the spectra, probably because the cytoplasm lies between the ECM and nucleus. Variations in the nuclear diameter had a significant effect on the backscatter spectra as well by shifting the frequency position of the peaks and troughs. Additionally, for spectra with no resonance structure (i.e., for cells with either all solid or all fluid cell components), the shape of the spectrum shifted in frequency with nuclear diameter. The sensitivity of the model to nucleus size as opposed to overall cell properties is probably due to the density of the nucleus in the computations.

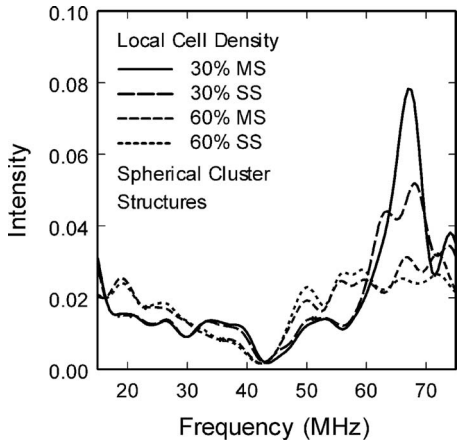


FIG. 13. Comparison between the averaged ultrasonic backscatter spectra for MS and SS simulations of uniform tissue microstructures with 30% local cell density and aggregated tissue microstructures with 60% local cell density.

Baddour *et al.* (2005) and Baddour and Kolios (2007) obtained experimental spectra from polystyrene spheres and whole cells in saline solution, and compared them with theoretical spectra from simple elastic and fluid spheres. They assumed that the nucleus would be the primary scattering center due to its higher density and wave speeds (Baddour *et al.*, 2005), and concluded that the influence of the nucleus on total cell scattering depends on the nucleus size (Baddour and Kolios, 2007). The simulated spectra of this work compare fairly well with their measurements. The experimentally measured dip observed in the backscatter transfer function at approximately 40 MHz for 12-μm-diameter polystyrene spheres (Baddour *et al.*, 2005) and whole cells (Baddour and Kolios, 2007) roughly corresponds to the dip at 45 MHz in the simulated spectra of Fig. 4(a). The details of the spectra differ, however, since the simulated spectra were for cells embedded in an ECM with a shear component of $c_{\text{shear}} = 106$ m/s. Significant spectral differences are also observed when the simulated spectra are compared against the simple sphere models of Anderson (1950), Faran (1951), and Hickling (1962). These discrepancies can be attributed to the more complex concentric structure of the simulated cells (Fig. 1). However, the resonance structure arising from a fluid-solid interface is seen in both the simpler theories and in the numerical simulations.

Other experimental data from centrifuged pellets of single cells and isolated nuclei also showed strong correlations between backscatter amplitudes and nuclear diameter at 20–60 MHz (Taggart *et al.*, 2007). However, as in previous work, the experimental spectra did not show the sharp peaks as predicted by the models. This has been attributed to a wide distribution of nuclear diameters and shapes for the cell types investigated (Baddour *et al.*, 2005; Taggart *et al.*, 2007). The results from this work show that variations in nuclear diameter will shift the positions of the resonance peaks. Since the peaks are narrow and often followed by troughs, only small size variations are needed to average out the peaks. Other explanations for the lack of resonance structure in single-cell spectra could include, however, variations in cell properties and overall cell structure. The mechanical properties of cells have been shown to be highly variable due to the adaptive nature of the cell and of the cytoskeleton

TABLE IV. Trends for significant peaks and bands in simulated ultrasonic backscatter spectra for four types of histological changes in tissues. I, D, and N denote increasing trend, decreasing trend, and no trend, respectively. Values in parentheses are coefficients of linear correlation. The trends for the 20–35-MHz band were obtained at 23.5 MHz.

Microstructural modification	18–20-MHz peak	20–35-MHz band	47-MHz peak	67-MHz peak
Cell aggregation (cluster; foam)	I (0.77; 0.60)	N (0.44; 0.13)	N (0.46; 0.46)	D (−0.69; −0.36)
Uniform nuclear pleomorphism	D (−0.91)	D (−0.99)	I (0.92)	N (−0.10)
In situ growth of malignant cells in lobules	I (0.91)	N (−0.30)	I (0.87)	N (0.020)
Malignant cell invasion into uniformly random tissue	I (0.85)	N (0.056)	I (0.79)	D (−0.37)

(Discher *et al.*, 2005). The cytoplasm will take on a range of shear properties depending on its environment, and the cytoplasmic shear properties of cells *in vivo* may differ from those of cells in suspension due to increased cytoskeletal organization resulting from cell-cell and cell-matrix contacts.

The simulation results for single cells indicate that a more realistic model of cell structure that includes shear properties and that differentiates the nucleus, cytoplasm, and ECM is needed since the shear properties of the cytoplasm and diameter of the nucleus have the strongest effects on ultrasonic backscatter spectra. Additionally, although the multiple-cell simulations appeared insensitive to cytoplasm shear properties, they were sensitive to the shear wave speed in the ECM. The results therefore indicate that tissues cannot be considered *a priori* to have the acoustic properties of a fluid since changes in the spectra will occur based on the shear properties given to the ECM or cytoplasm, and many tissues will have elastic matrix proteins such as collagen and elastin. The shear properties of the ECM and cytoplasm are also important for mode conversion between longitudinal and shear waves at the plasma membrane. This process may be important for the generation of shear forces at the plasma membrane and possible sonoporation or membrane disruption at high ultrasonic intensities.

B. Multiple-cell simulations

Ultrasonic backscatter spectra from the multiple-cell simulations were substantially different from single-cell spectra, even where all of the cells and nuclei had the same size and shape (e.g., Figs. 6 and 8). Although the ECM and cytoplasm provided a solid-fluid interface in the tissues, sharp peaks were not observed as they were in the single-cell spectra. The absence of single-cell resonance structure in both MS and SS simulations (Fig. 13) indicates that it cannot be attributed to MS in the tissue. Rather, it is most likely that backscattered waves from all of the cells sum incoherently at the virtual transducer. This incoherent summation would average over the single-cell resonance responses and eliminate the sharp peaks in the tissue spectra.

Broader peaks were observed in the multiple-cell simulations that showed a strong dependence on tissue microstructure. For ordered cell arrangements, the cell lattice produces conditions favorable for acoustic band gaps, and the simulations correctly predicted band-gap frequencies consistent with theory (Fig. 6). Backscatter spectral peaks were also produced or attenuated due to structural changes in random and hierarchical microstructures (Fig. 8). Comparisons between MS and SS, Fig. 13, showed that aggregation of cells produced spectral changes that were independent of MS effects. The uniform and aggregated packs produced substantially different spectra whether MS was present or not. The spectral differences therefore most probably arise from the differential superposition of the backscattered waves at the virtual transducer.

Figure 13 shows that MS effects become significant above 60 MHz for tissue microstructures with a uniform, low packing density (30%) and above 45 MHz for tissue microstructures with an aggregated, high packing density (60%). These results are consistent with the conclusion that MS in tissues will be dependent on the cell density of the tissue structure. Tissues with denser cell packings may show MS effects at lower frequencies, and recent experiments suggest that actual tissues may exhibit MS at frequencies as low as 4.3 MHz (Aubry *et al.*, 2008). Other factors may be important as well, such as the material properties of the ECM and cell components. Important information on tissue microstructure may therefore be lost in high-frequency tissue characterization models that do not include MS.

Internal cell structure also had a dramatic effect on the simulated ultrasonic spectra. Changes in the sizes of nuclei produced a 47-MHz peak that showed strong correlations with either nucleus size or number of enlarged nuclei introduced into the microstructure. Both homogeneous changes, where all of the nuclei diameters were the same and increased uniformly (Doyle *et al.*, 2007b), and heterogeneous changes, where new cells with larger nuclei were introduced into tissue cavities (Doyle *et al.*, 2007b) or uniformly random tissues [Fig. 12(b)], displayed similar linear trends in the 47-MHz peak. Other spectral changes were observed as well that were more specific to the simulated structural change. For example, uniform changes in nuclei sizes produced spectral intensity decreases in the 18-MHz peak and 20–35-MHz region that were not observed for other microstructural modifications such as cell aggregation and malignant cell invasion. Table IV summarizes the simulation results for the most significant spectral changes. Experimental evidence for a dominant 47-MHz peak arising from nuclear size effects comes from Taggart *et al.* (2007), who show a peak in the backscatter coefficient at 47 MHz for multinucleated human epithelial kidney cells.

The spectral peaks predicted by the multiple-cell calculations and their trends as a function of microstructure are robust and statistically significant. The error bars in Fig. 12(b) represent the standard deviation of results obtained from the simulation of five different particle packs where the particle positions and larger microstructures were varied but

the overall histological change—percent malignant cell infiltration—was kept constant. The predicted trends exceed the uncertainty introduced by the structural variations and correlate well with malignant cell invasion. These results and those by Doyle *et al.* (2007b) show that even though the microstructure can exhibit complex permutations, general trends may still be predictable based on common microstructural features. Moreover, Table IV shows that different structural processes in tissues may produce different spectral responses. Use of pattern recognition approaches such as principal components analysis or artificial neural networks could in principle be used to classify these spectral signatures in a database of simulation spectra. The findings suggest that high-frequency (10–100 MHz) ultrasonic spectroscopy would be useful for determining histology in a clinical setting and that histology-based simulation tools may be able to detect basic histological features such as microscopic cancer in spite of a large degree of tissue complexity.

The histology-based ultrasonic model developed in this work can also be used to predict effective propagation properties such as longitudinal velocity, shear velocity, and attenuation (Doyle, 2006). Correlating these properties to tissue structure is important in ultrasonic tomography, ultrasound elastography, and other ultrasonic imaging modalities. The model can also simulate ultrasonic pulses with arbitrary spatial and temporal characteristics by altering the spherical-wave expansion coefficients for the incident plane wave and by using a Fourier transform method to simulate a time-domain pulse in the frequency domain (Doyle, 2006).

C. Future research

Although the focus of this article is on modeling and simulation, experimental validations with real tissues are essential for assessing the accuracy of the models and their utility in clinical practice. Further development of the model will therefore include simulations of larger and more complex tissue structures, parallelization on a multiprocessor system to improve the computational efficiency, and validation using ultrasonic testing of tissue samples. Although the simulation algorithm has the capability to model viscoelastic behavior in tissues, viscoelastic properties were not included in this study since the objective was to demonstrate the feasibility of the approach and not to obtain detailed predictions. Viscoelastic properties are known to preferentially attenuate shear and high-frequency waves in tissues, and will be included in future work. Comparison of model predictions with measurements from well-characterized tissues and cell cultures will additionally be performed to refine the tissue and cell properties.

The present work is currently being expanded to simulate non-centered nuclei as well as an additional layer in the cell to represent a plasma membrane with nonzero thickness and properties. Multipole expansion methods can also be used to model spheroidal cells (Rose *et al.*, 1995; Ciric and Cooray, 2000; Li *et al.*, 2002) as well as multiple organelles within a cell (Fuller and Mackowski, 2000). Such capabilities would allow the simulation of more structured and histologically accurate tissues. For example, columnar and squamous cells could be more closely approximated using prolate and oblate spheroids, respectively, instead of spheres.

D. Medical applications

The results from this study indicate that a simulation approach can directly model the effects of histological changes on ultrasonic backscattering in a variety of tissue microstructures. Since fewer approximations are used than with previous approaches, the models can account for MS, shear waves, mode conversion, and intracellular structure in tissues with random, semi-random, ordered, and hierarchical microstructures. By incorporating these structural details and higher-order scattering processes, numerical approaches can build upon the capabilities of previous analytical models for tissue characterization.

For diagnostic purposes, histology-based numerical models could be used to generate a database of ultrasonic properties and spectra for a range of pathologies in tissues. As demonstrated in Figs. 11 and 12, a number of simulations could be used to extract robust signatures from the data for specific diseases. Classifier methods such as principal components analysis could then be used to evaluate the clinical measurements and diagnose the disease state. This approach promises to provide real-time histopathology during ultrasonic examinations.

A potential limitation of this application is that the simulation results to date indicate that the greatest changes in the ultrasonic spectra will occur at higher frequencies (>15 MHz) and across broader bandwidths than those routinely used for conventional ultrasonic imaging. However, for many new applications these frequencies and bandwidths are not proving to be impediments. Huang *et al.* (2007) acquired high-frequency measurements for the lens of the eye using a bandwidth covering 25–45 MHz, and Taggart *et al.* (2007) characterized centrifuged pellets of leukemia, epithelial kidney, and breast cancer cells from 20 to 60 MHz using a single, broadband transducer. The use of high-frequency ultrasound has also become more common with the development of shallow, high-resolution scans and intravascular ultrasound. The numerical models developed in this work would therefore be applicable to these applications as well as to new ultrasonic methods for biopsies, endoscopies, surgical margin assessments, and oral and skin examinations. In particular, the results indicate that the models may permit the *in vivo* detection, grading, and staging of microscopic or microinvasive cancer in heterogeneous tissues (Doyle *et al.*, 2007b). The shell-core multipole solutions used for the cell models are also well suited for modeling the ultrasonic response of microbubble contrast agents in tissues.

For therapeutic applications, histology-based models could be used to predict the spatial distribution of ultrasonic energy in heterogeneous tissues at the microscopic level and understand the roles of elasticity, viscoelasticity, shear forces, and the tissue environment on ultrasound-cell interactions. These factors are important for a fundamental understanding of sonoporation, cell heating, and other processes that alter the cell state for medical treatment. Incorporating heating, cavitation, and nonlinear bubble or cell oscillations

into histology-based simulations will require further developments, however, since the iterative multipole approach is a linear, small-deformation model.

V. CONCLUSIONS

Forward models that simulate ultrasonic interactions at the microscopic level were developed to provide a more detailed, physics-based understanding of elastic wave scattering in cells and tissues. The models used an iterative multipole method to simulate MS, cells with internal structure, arbitrary configurations of scatterers, and both shear and longitudinal waves with associated mode conversion at the plasma membrane and nuclear envelope. Wave field images and ultrasonic backscatter spectra were calculated for simulated tissues with random, ordered, and hierarchical microstructures. The preliminary results indicate that MS and internal cell structure may have significant effects in high-frequency ultrasonic spectra. The results also suggest that common microstructural changes associated with neoplasms such as nuclear pleomorphism, malignant cell infiltration, and tissue remodeling will show robust, diagnostic changes in the predicted ultrasonic properties and spectra regardless of background microstructural variations.

The prediction of spectral features for single cells and of acoustic band gaps for hcp tissue lattices demonstrated the consistency of the model with acoustic theory. However, experimental verification of the model predictions for heterogeneous tissues is required to determine the value of applying a direct simulation approach to the interpretation of clinical ultrasound data. Although distributions in the sizes, shapes, and structures of cells and nuclei will broaden the spectral features produced by the "ideal" tissue structures presented here, the iterative multipole method has the capability to simulate such distributions and therefore more realistic tissue structures.

Histology-based numerical models hold promise for understanding the fundamental interactions of ultrasound in tissues at the microscopic level. This understanding is vital in extracting detailed histological information from ultrasonic data, advancing ultrasonic diagnostic capabilities, and providing simulation-based tools for the early and accurate diagnosis of diseases and disease states such as cancer, premalignant conditions, coronary heart disease, inflammation, fibrosis, liver steatosis, and tissue degradation. In addition to *in vivo* tissue characterization, such an understanding would benefit ultrasound-based therapies by providing predictive models for determining the mechanisms and effects of ultrasonic treatments from the cell level upward. Such treatments may include ultrasound-enhanced drug delivery, gene transfection, and tissue healing.

ACKNOWLEDGMENTS

This work was supported by a Community/University Research Initiative (CURI) grant from Utah State University. The authors thank David Robinson, Raymond DeVito, Leigh Neumayer, Carol Achziger, Greg Podgorski, and the Breast Interdisciplinary Group at the Huntsman Cancer Institute for their continuing help and encouragement.

APPENDIX A: DERIVATION OF BOUNDARY CONDITION EQUATIONS

The first three BC equations for the concentric-sphere cell model [Eqs. (12)–(14)] are derived from the continuity of the displacements [Eq. (5)] at the matrix-shell boundary. The vector equation for the displacement BC is expressed as an expansion of vector multipole functions [Eqs. (2)–(4)] of n,m moments with corresponding amplitude coefficients for each wave field:

$$\begin{aligned}
&\sum_{n=0}^{\infty}\sum_{m=-n}^{n} A_{nm}\mathbf{U}_{nm}(r,\theta,\varphi)+\sum_{n=0}^{\infty}\sum_{m=-n}^{n} B_{nm}\mathbf{V}_{nm}(r,\theta,\varphi)\\
&\quad+\sum_{n=0}^{\infty}\sum_{m=-n}^{n} C_{nm}\mathbf{W}_{nm}(r,\theta,\varphi)\\
&\quad+\sum_{n=0}^{\infty}\sum_{m=-n}^{n} G_{nm}\mathbf{U}_{nm}(r,\theta,\varphi)\\
&\quad+\sum_{n=0}^{\infty}\sum_{m=-n}^{n} H_{nm}\mathbf{V}_{nm}(r,\theta,\varphi)+\sum_{n=0}^{\infty}\sum_{m=-n}^{n} I_{nm}\mathbf{W}_{nm}(r,\theta,\varphi)\\
&=\sum_{n=0}^{\infty}\sum_{m=-n}^{n} As_{nm}\mathbf{U}_{nm}(r,\theta,\varphi)+\sum_{n=0}^{\infty}\sum_{m=-n}^{n} Bs_{nm}\mathbf{V}_{nm}(r,\theta,\varphi)\\
&\quad+\sum_{n=0}^{\infty}\sum_{m=-n}^{n} Cs_{nm}\mathbf{W}_{nm}(r,\theta,\varphi)\\
&\quad+\sum_{n=0}^{\infty}\sum_{m=-n}^{n} Gs_{nm}\mathbf{U}_{nm}(r,\theta,\varphi)\\
&\quad+\sum_{n=0}^{\infty}\sum_{m=-n}^{n} Hs_{nm}\mathbf{V}_{nm}(r,\theta,\varphi)\\
&\quad+\sum_{n=0}^{\infty}\sum_{m=-n}^{n} Is_{nm}\mathbf{W}_{nm}(r,\theta,\varphi). \qquad \text{(A1)}
\end{aligned}$$

The vector equation for the displacement BC has traditionally been separated into three scalar equations. The angular dependencies have then been removed and the moments isolated using Eq. (11). However, this orthogonality condition does not reduce the scalar equations to a closed-form solution for the coefficients (Doyle, 2004). Rather, the vector equation must be solved in its entirety using the orthogonality of the vector spherical harmonics [Eq. (10)]. To achieve this, Eq. (A1) is multiplied by $\mathbf{Y}_{NM}^{N-1}*(\theta,\varphi)$ and integrated over the shell at $r=b$ to generate the first coefficient equation. The second and third coefficient equations are generated in a similar manner by multiplying with $\mathbf{Y}_{NM}^{N}*(\theta,\varphi)$ and $\mathbf{Y}_{NM}^{N+1}*(\theta,\varphi)$, respectively, and integrating.

With multiplication by $\mathbf{Y}_{NM}^{N-1}*(\theta,\varphi)$ and integration over θ and φ, only terms with $\mathbf{Y}_{nm}^{n-1}(\theta,\varphi)$ in the vector multipole functions survive due to the delta functions over both N,n and L,l in Eq. (10). This operation results in the following equation:

$$\sqrt{\frac{N}{2N+1}}[A_{NM}h_{N-1}^{(2)}(k_L^{\text{ext}}b)+G_{NM}h_{N-1}^{(1)}(k_L^{\text{ext}}b)]$$
$$+i\sqrt{\frac{N+1}{2N+1}}[B_{NM}h_{N-1}^{(2)}(k_S^{\text{ext}}b)+H_{NM}h_{N-1}^{(1)}(k_S^{\text{ext}}b)]$$
$$=\sqrt{\frac{N}{2N+1}}[As_{NM}h_{N-1}^{(2)}(k_L^{\text{shell}}b)+Gs_{NM}h_{N-1}^{(1)}(k_L^{\text{shell}}b)]$$
$$+i\sqrt{\frac{N+1}{2N+1}}[Bs_{NM}h_{N-1}^{(2)}(k_S^{\text{shell}}b)$$
$$+Hs_{NM}h_{N-1}^{(1)}(k_S^{\text{shell}}b)]. \quad \text{(A2)}$$

Similarly, multiplication with $\mathbf{Y}_{NM}^{N}*(\theta,\varphi)$ and $\mathbf{Y}_{NM}^{N+1}*(\theta,\varphi)$ and integration over θ and φ yields the following, respectively:

$$C_{NM}h_N^{(2)}(k_S^{\text{ext}}b)+I_{NM}h_N^{(1)}(k_S^{\text{ext}}b)-Cs_{NM}h_N^{(2)}(k_S^{\text{shell}}b)$$
$$-Is_{NM}h_N^{(1)}(k_S^{\text{shell}}b)=0. \quad \text{(A3)}$$

$$\sqrt{\frac{N+1}{2N+1}}[A_{NM}h_{N+1}^{(2)}(k_L^{\text{ext}}b)+G_{NM}h_{N+1}^{(1)}(k_L^{\text{ext}}b)]$$
$$-i\sqrt{\frac{N}{2N+1}}[B_{NM}h_{N+1}^{(2)}(k_S^{\text{ext}}b)+H_{NM}h_{N+1}^{(1)}(k_S^{\text{ext}}b)]$$
$$=\sqrt{\frac{N+1}{2N+1}}[As_{NM}h_{N+1}^{(2)}(k_L^{\text{shell}}b)+Gs_{NM}h_{N+1}^{(1)}(k_L^{\text{shell}}b)]$$
$$-i\sqrt{\frac{N}{2N+1}}[Bs_{NM}h_{N+1}^{(2)}(k_S^{\text{shell}}b)$$
$$+Hs_{NM}h_{N+1}^{(1)}(k_S^{\text{shell}}b)]. \quad \text{(A4)}$$

Equations (A2) and (A4) are reformulated by dividing out common factors, multiplying Eq. (A2) by $\sqrt{N}$, multiplying Eq. (A4) by $-\sqrt{N+1}$, and adding the two equations:

$$A_{NM}[(N)h_{N-1}^{(2)}(k_L^{\text{ext}}b)-(N+1)h_{N+1}^{(2)}(k_L^{\text{ext}}b)]$$
$$+G_{NM}[(N)h_{N-1}^{(1)}(k_L^{\text{ext}}b)-(N+1)h_{N+1}^{(1)}(k_L^{\text{ext}}b)]$$
$$-As_{NM}[(N)h_{N-1}^{(2)}(k_L^{\text{shell}}b)-(N+1)h_{N+1}^{(2)}(k_L^{\text{shell}}b)]$$
$$-Gs_{NM}[(N)h_{N-1}^{(1)}(k_L^{\text{shell}}b)-(N+1)h_{N+1}^{(1)}(k_L^{\text{shell}}b)]$$
$$+i\sqrt{(N)(N+1)}\{B_{NM}[h_{N-1}^{(2)}(k_S^{\text{ext}}b)+h_{N+1}^{(2)}(k_S^{\text{ext}}b)]$$
$$+H_{NM}[h_{N-1}^{(1)}(k_S^{\text{ext}}b)+h_{N+1}^{(1)}(k_S^{\text{ext}}b)]$$
$$-Bs_{NM}[h_{N-1}^{(2)}(k_S^{\text{shell}}b)+h_{N+1}^{(2)}(k_S^{\text{shell}}b)]$$
$$-Hs_{NM}[h_{N-1}^{(1)}(k_S^{\text{shell}}b)+h_{N+1}^{(1)}(k_S^{\text{shell}}b)]\}=0. \quad \text{(A5)}$$

The following spherical radial function identities are next used to simplify Eq. (A5):

$$\frac{(N)z_{N-1}(kr)-(N+1)z_{N+1}(kr)}{2N+1}=N\frac{z_N(kr)}{kr}-z_{N+1}(kr), \quad \text{(A6)}$$

$$\frac{z_{N-1}(kr)+z_{N+1}(kr)}{2N+1}=\frac{z_N(kr)}{kr}. \quad \text{(A7)}$$

With substitution of the symbols $\eta_1(z)$ and $\eta_2(z)$ (see Appendix B) and solving for the unknown coefficients (As_{NM}, Bs_{NM}, Gs_{NM}, Hs_{NM}, G_{NM}, and H_{NM}), the first BC equation becomes

$$As_{NM}\eta_1(g^{\text{shell}})+Bs_{NM}\eta_2(g^{\text{shell}})+Gs_{NM}\eta_1(h^{\text{shell}})$$
$$+Hs_{NM}\eta_2(h^{\text{shell}})-G_{NM}\eta_1(h^{\text{ext}})-H_{NM}\eta_2(h^{\text{ext}})$$
$$=A_{NM}\eta_1(g^{\text{ext}})+B_{NM}\eta_2(g^{\text{ext}}). \quad \text{(A8)}$$

The second BC equation is derived by dividing Eq. (A2) by $\sqrt{N}$, dividing Eq. (A4) by $\sqrt{N+1}$, adding the two equations, and simplifying algebraically:

$$A_{NM}[h_{N-1}^{(2)}(k_L^{\text{ext}}b)+h_{N+1}^{(2)}(k_L^{\text{ext}}b)]+G_{NM}[h_{N-1}^{(1)}(k_L^{\text{ext}}b)$$
$$+h_{N+1}^{(1)}(k_L^{\text{ext}}b)]-As_{NM}[h_{N-1}^{(2)}(k_L^{\text{shell}}b)+h_{N+1}^{(2)}(k_L^{\text{shell}}b)]$$
$$-Gs_{NM}[h_{N-1}^{(1)}(k_L^{\text{shell}}b)+h_{N+1}^{(1)}(k_L^{\text{shell}}b)]$$
$$+\frac{i}{\sqrt{(N)(N+1)}}\{B_{NM}[(N+1)h_{N-1}^{(2)}(k_S^{\text{ext}}b)$$
$$-(N)h_{N+1}^{(2)}(k_S^{\text{ext}}b)]+H_{NM}[(N+1)h_{N-1}^{(1)}(k_S^{\text{ext}}b)$$
$$-(N)h_{N+1}^{(1)}(k_S^{\text{ext}}b)]-Bs_{NM}[(N+1)h_{N-1}^{(2)}(k_S^{\text{shell}}b)$$
$$-(N)h_{N+1}^{(2)}(k_S^{\text{shell}}b)]-Hs_{NM}[(N+1)h_{N-1}^{(1)}(k_S^{\text{shell}}b)$$
$$-(N)h_{N+1}^{(1)}(k_S^{\text{shell}}b)]\}=0. \quad \text{(A9)}$$

Along with Eq. (A7), an additional spherical radial function identity is used to simplify Eq. (A9):

$$\frac{(N+1)z_{N-1}(kr)-(N)z_{N+1}(kr)}{2N+1}$$
$$=(N+1)\frac{z_N(kr)}{kr}-z_{N+1}(kr). \quad \text{(A10)}$$

With substitution of the symbols $\eta_4(z)$ and $\eta_5(z)$ and solving for the unknown coefficients, the second BC equation becomes

$$As_{NM}\eta_4(g^{\text{shell}})+Bs_{NM}\eta_5(g^{\text{shell}})+Gs_{NM}\eta_4(h^{\text{shell}})$$
$$+Hs_{NM}\eta_5(h^{\text{shell}})-G_{NM}\eta_4(h^{\text{ext}})-H_{NM}\eta_5(h^{\text{ext}})$$
$$=A_{NM}\eta_4(g^{\text{ext}})+B_{NM}\eta_5(g^{\text{ext}}). \quad \text{(A11)}$$

The third coefficient equation, Eq. (A3), is rewritten with $\eta_3(z)$ to yield the third BC equation:

$$Cs_{NM}\eta_3(g^{\text{shell}})+Is_{NM}\eta_3(h^{\text{shell}})-I_{NM}\eta_3(h^{\text{ext}})$$
$$=C_{NM}\eta_3(g^{\text{ext}}). \quad \text{(A12)}$$

The displacement BC equations at the shell-core interface [Eqs. (15)–(17)] are derived in the same manner. The stress BC equations [Eqs. (18)–(23)] are generated in a similar manner as well, except that the scalar orthogonality condition [Eq. (11)] is sufficient for removing the angular dependencies and isolating the moments since the BC equation [Eq. (6)] and its constituent terms [Eqs. (7)–(9)] are scalar in character and produce linearly separable coefficient equations.

APPENDIX B: SYMBOL DEFINITIONS FOR RADIAL FUNCTION TERMS

The η-symbols from the matrix-shell and shell-core BC equations, Eqs. (12)–(23), are presented below. The symbols are functions of the multipole order N, longitudinal wave vector k_L, shear wave vector k_S, inner sphere radius a, and outer sphere radius b. The spherical radial functions are denoted by $z_n(kr)$, and vary according to the radial function designated by the subscripts in Eqs. (12)–(23) as $j=j_n(kr)$, $h=h_n^{(1)}(kr)$, or $g=h_n^{(2)}(kr)$.

$$\eta_1(z) = N\frac{z_N(k_L b)}{k_L b} - z_{N+1}(k_L b), \tag{B1}$$

$$\eta_2(z) = i\sqrt{(N)(N+1)}\frac{z_N(k_S b)}{k_S b}, \tag{B2}$$

$$\eta_3(z) = z_N(k_S b), \tag{B3}$$

$$\eta_4(z) = \frac{z_N(k_L b)}{k_L b}, \tag{B4}$$

$$\eta_5(z) = \frac{i}{\sqrt{(N)(N+1)}}\left[(N+1)\frac{z_N(k_S b)}{k_S b} - z_{N+1}(k_S b)\right], \tag{B5}$$

$$\eta_6(z) = k_L\left[-\lambda z_N(k_L b) + 2\mu\left\{\left[\frac{(N-1)(N)}{(k_L b)^2} - 1\right]z_N(k_L b) + 2\frac{z_{N+1}(k_L b)}{k_L b}\right\}\right], \tag{B6}$$

$$\eta_7(z) = 2i\mu k_S\sqrt{(N)(N+1)}\left[(N-1)\frac{z_N(k_S b)}{(k_S b)^2} - \frac{z_{N+1}(k_S b)}{k_S b}\right], \tag{B7}$$

$$\eta_8(z) = \mu k_S\left[(N-1)\frac{z_N(k_S b)}{k_S b} - z_{N+1}(k_S b)\right], \tag{B8}$$

$$\eta_9(z) = 2\mu k_L\left[(N-1)\frac{z_N(k_L b)}{(k_L b)^2} - \frac{z_{N+1}(k_L b)}{k_L b}\right], \tag{B9}$$

$$\eta_{10}(z) = \frac{i\mu k_S}{\sqrt{(N)(N+1)}}\left\{\left[\frac{2(N^2-1)}{(k_S b)^2} - 1\right]z_N(k_S b) + 2\frac{z_{N+1}(k_S b)}{k_S b}\right\}, \tag{B10}$$

$$\eta_{11}(z) = N\frac{z_N(k_L a)}{k_L a} - z_{N+1}(k_L a), \tag{B11}$$

$$\eta_{12}(z) = i\sqrt{(N)(N+1)}\frac{z_N(k_S a)}{k_S a}, \tag{B12}$$

$$\eta_{13}(z) = z_N(k_S a), \tag{B13}$$

$$\eta_{14}(z) = \frac{z_N(k_L a)}{k_L a}, \tag{B14}$$

$$\eta_{15}(z) = \frac{i}{\sqrt{(N)(N+1)}}\left[(N+1)\frac{z_N(k_S a)}{k_S a} - z_{N+1}(k_S a)\right], \tag{B15}$$

$$\eta_{16}(z) = k_L\left[-\lambda z_N(k_L a) + 2\mu\left\{\left[\frac{(N-1)(N)}{(k_L a)^2} - 1\right]z_N(k_L a) + 2\frac{z_{N+1}(k_L a)}{k_L a}\right\}\right], \tag{B16}$$

$$\eta_{17}(z) = 2i\mu k_S\sqrt{(N)(N+1)}\left[(N-1)\frac{z_N(k_S a)}{(k_S a)^2} - \frac{z_{N+1}(k_S a)}{k_S a}\right], \tag{B17}$$

$$\eta_{18}(z) = \mu k_S\left[(N-1)\frac{z_N(k_S a)}{k_S a} - z_{N+1}(k_S a)\right], \tag{B18}$$

$$\eta_{19}(z) = 2\mu k_L\left[(N-1)\frac{z_N(k_L a)}{(k_L a)^2} - \frac{z_{N+1}(k_L a)}{k_L a}\right], \tag{B19}$$

$$\eta_{20}(z) = \frac{i\mu k_S}{\sqrt{(N)(N+1)}}\left\{\left[\frac{2(N^2-1)}{(k_S a)^2} - 1\right]z_N(k_S a) + 2\frac{z_{N+1}(k_S a)}{k_S a}\right\}. \tag{B20}$$

Alacam, B., Yazici, B., Bilgutay, N., Forsberg, F., and Piccoli, C. (**2004**). "Breast tissue characterization using FARMA modeling of ultrasonic RF echo," Ultrasound Med. Biol. **30**, 1397–1407.

Anderson, V. C. (**1950**). "Sound scattering from a fluid sphere," J. Acoust. Soc. Am. **22**, 426–431.

Aubry, A., Derode, A., and Tanter, M. (**2008**). "Extraction of the multiple scattering contribution in weakly scattering media: Application to human soft tissue," J. Acoust. Soc. Am. **123**, 3001.

Baddour, R. E., and Kolios, M. C. (**2007**). "The fluid and elastic nature of nucleated cells: Implications from the cellular backscatter response," J. Acoust. Soc. Am. **121**, EL16–EL22.

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**, 934–943.

Bige, Y., Hanfeng, Z., and Rong, W. (**2006**). "Analysis of microstructural alterations of normal and pathological breast tissue *in vivo* using the AR cepstrum," Ultrasonics **44**, 211–215.

Ciric, I. R., and Cooray, F. R., **2000**, in *Light Scattering by Nonspherical Particles*, edited by M. I. Mishchenko, J. W. Hovenier, and L. D. Travis (Academic, San Diego), pp. 89–130.

Discher, D. E., Janmey, P., and Wang, Y.-L. (**2005**). "Tissue cells feel and respond to the stiffness of their substrate," Science **310**, 1139–1143.

Doyle, T. E. (**2004**) "Computational scattering models for elastic and electromagnetic waves in particulate media," Ph.D. dissertation, Utah State University, Logan, UT.

Doyle, T. E., **2005**, in *Review of Progress in Quantitative Nondestructive Evaluation*, edited by D. O. Thompson and D. E. Chimenti (American Institute of Physics, Melville, NY), Vol. **24**, pp. 1121–1128.

Doyle, T. E. (**2006**). "Iterative simulation of elastic wave scattering in arbitrary dispersions of spherical particles," J. Acoust. Soc. Am. **119**, 2599–2610.

Doyle, T. E., Robinson, D. A., Jones, S. B., Warnick, K. H., and Carruth, B. L. (**2007c**). "Modeling the permittivity of two-phase media containing

monodisperse spheres: Effects of microstructure and multiple scattering," Phys. Rev. B **76**, 054203.
Doyle, T. E., and Warnick, K. H. (**2006**). "Simulation of elastic wave scattering in living tissue at the cellular level," J. Acoust. Soc. Am. **120**, 3283.
Doyle, T. E., Warnick, K. H., and Carruth, B. L. (**2007a**). "Histology-based simulation of ultrasonic scattering in cells and tissues," J. Acoust. Soc. Am. **121**, 3111.
Doyle, T. E., Warnick, K. H., and Carruth, B. L. (**2007b**). "Histology-based simulations for the ultrasonic detection of microscopic cancer *in vivo*," J. Acoust. Soc. Am. **122**, EL210–EL216.
Falou, O., Kumaradas, J. C., and Kolios, M. C. (**2006**). "Finite element modeling of ultrasound scattering by spherical objects and cells," 2006 IEEE Ultrasonics Symposium, Vancouver, Canada, Oct. 2–6, pp. 2072–2075.
Faran, J. J. (**1951**). "Sound scattering by solid cylinders and spheres," J. Acoust. Soc. Am. **23**, 405–418.
Feleppa, E. J., Kalisz, A., Sokil-Melgar, J. B., Lizzi, F. L., Liu, T., Rosado, A. L., Shao, M. C., Fair, W. R., Wang, Y., Cookson, M. S., Reuter, V. E., and Heston, W. D. W. (**1996**). "Typing of prostate tissue by ultrasonic spectrum analysis," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **43**, 609–619.
Finney, J. L. (**1970**). "Random packings and the structure of simple liquids: I. The geometry of random close packing," Proc. R. Soc. London, Ser. A **319**, 479–493.
Fuller, K. A., and Mackowski, D. W., **2000**, in *Light Scattering by Nonspherical Particles*, edited by M. I. Mishchenko, J. W. Hovenier, and L. D. Travis (Academic, San Diego), pp. 225–272.
Gaitini, D., Baruch, Y., Ghersin, E., Veitsman, E., Kerner, H., Shalem, B., Yaniv, G., Sarfaty, C., and Azhari, H. (**2004**). "Feasibility study of ultrasonic fatty liver biopsy: Texture vs. attenuation and backscatter," Ultrasound Med. Biol. **30**, 1321–1327.
Gumerov, N. A., and Duraiswami, R. (**2002**). "Computation of scattering from N spheres using multipole reexpansion," J. Acoust. Soc. Am. **112**, 2688–2701.
Gumerov, N. A., and Duraiswami, R. (**2005**). "Computation of scattering from clusters of spheres using the fast multipole method," J. Acoust. Soc. Am. **117**, 1744–1761.
Hickling, R. (**1962**). "Analysis of echoes from a solid elastic sphere in water," J. Acoust. Soc. Am. **34**, 1582–1592.
Huang, C.-C., Ameri, H., DeBoer, C., Rowley, A. P., Xu, X., Sun, L., Wang, S.-H., Humayun, M. S., and Shung, K. K. (**2007**). "Evaluation of lens hardness in cataract surgery using high-frequency ultrasonic parameters *in vitro*," Ultrasound Med. Biol. **33**, 1609–1616.
Hunt, J. W., Worthington, A. E., Xuan, A., Kolios, M. C., Czarnota, G. J., and Sherar, M. D. (**2002**). "A model based upon pseudo regular spacing of cells combined with the randomization of the nuclei can explain the significant changes in high-frequency ultrasound signals during apoptosis," Ultrasound Med. Biol. **28**, 217–226.
Insana, M. F. (**1995**). "Modeling acoustic backscatter from kidney microstructure using an anisotropic correlation function," J. Acoust. Soc. Am. **97**, 649–655.
Kutay, M. A., Petropulu, A. P., and Piccoli, C. W. (**2001**). "On modeling biomedical ultrasound RF echoes using a power-law shot-noise model," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **48**, 953–968.
Li, L.-W., Kang, X.-K., and Leong, M.-S., **2002**, *Spheroidal Wave Functions in Electromagnetic Theory* (Wiley, New York).
Liu, X.-Z., Gong, X.-F., Zhang, D., Ye, S.-G., and Rui, B. (**2006**). "Ultrasonic characterization of porcine liver tissue at frequency between 25 to 50 MHz," World J. Gastroenterol. **12**, 2276–2279.
Liu, Z., Chan, C. T., Sheng, P., Goertzen, A. L., and Page, J. H. (**2000**). "Elastic wave scattering by periodic structures of spherical objects: Theory and experiment," Phys. Rev. B **62**, 2446–2457.
Lizzi, F. L., Astor, M., Feleppa, E. J., Shao, M., and Kalisz, A. (**1997**). "Statistical framework for ultrasonic spectral parameter imaging," Ultrasound Med. Biol. **23**, 1371–1382.
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**, 1366–1373.
Mamou, J., Oelze, M. L., O'Brien, W. D., Jr., and Zachary, J. F. (**2005**). "Identifying ultrasonic scattering sites from three-dimensional impedance maps," J. Acoust. Soc. Am. **117**, 413–423.
Mamou, J., Oelze, M. L., O'Brien, W. D., Jr.., and Zachary, J. F. (**2008**). "Extended three-dimensional impedance map methods for identifying ultrasonic scattering sites," J. Acoust. Soc. Am. **123**, 1195–1208.
Meziri, M., Pereira, W. C. A., Abdelwahab, A., Degott, C., and Laugier, P. (**2005**). "*In vitro* chronic hepatic disease characterization with a multiparametric ultrasonic approach," Ultrasonics **43**, 305–313.
Oelze, M. L., and O'Brien, W. D., Jr. (**2006**). "Application of three scattering models to characterization of solid tumors in mice," Ultrason. Imaging **28**, 83–96.
Oelze, M. L., O'Brien, W. D., Jr., and Zachary, J. F. (**2007**). "Quantitative ultrasound assessment of breast cancer using a multiparameter approach," 2007 IEEE Ultrasonics Symposium, pp. 981–984.
Oelze, M. L., and Zachary, J. F. (**2006**). "Examination of cancer in mouse models using high-frequency quantitative ultrasound," Ultrasound Med. Biol. **32**, 1639–1648.
Oelze, M. L., Zachary, J. F., and O'Brien, W. D., Jr. (**2002**). "Characterization of tissue microstructure using ultrasonic backscatter: Theory and technique for optimization using a Gaussian form factor," J. Acoust. Soc. Am. **112**, 1202–1211.
Padilla, F., Peyrin, F., and Laugier, P. (**2003**). "Prediction of backscatter coefficient in trabecular bones using a numerical model of three-dimensional microstructure," J. Acoust. Soc. Am. **113**, 1122–1129.
Rose, J. H., Kaufmann, M. R., Wickline, S. A., Hall, C. S., and Miller, J. G. (**1995**). "A proposed microscopic elastic wave theory for ultrasonic backscatter from myocardial tissue," J. Acoust. Soc. Am. **97**, 656–668.
Santarelli, M. F., and Landini, L. (**1996**). "A model of ultrasound backscatter for the assessment of myocardial tissue structure and architecture," IEEE Trans. Biomed. Eng. **43**, 901–911.
Savéry, D., and Cloutier, G. (**2007**). "High-frequency ultrasound backscattering by blood: Analytical and semianalytical models of the erythrocyte cross section," J. Acoust. Soc. Am. **121**, 3963–3971.
Taggart, L. R., Baddour, R. E., Giles, A., Czarnota, G. J., and Kolios, M. C. (**2007**). "Ultrasonic characterization of whole cells and isolated nuclei," Ultrasound Med. Biol. **33**, 389–401.
Wear, K. A. (**2004**). "Measurement of dependence of backscatter coefficient from cylinders on frequency and diameter using focused transducers—with applications in trabecular bone," J. Acoust. Soc. Am. **115**, 66–72.
Yang, M., Baldwin, S. L., Marutyan, K. R., Wallace, K. D., Holland, M. R., and Miller, J. G. (**2006**). "Elastic stiffness coefficients (c_{11}, c_{33}, and c_{13}) for freshly excised and formalin-fixed myocardium from ultrasonic velocity measurements," J. Acoust. Soc. Am. **119**, 1880–1887.
Yeh, W.-C., Jeng, Y.-M., Li, C.-H., Lee, P.-H., and Li, P.-C. (**2005**). "Liver steatosis classification using high-frequency ultrasound," Ultrasound Med. Biol. **31**, 599–605.
Zinin, P. V. (**1992**). "A theoretical analysis of sound attenuation mechanisms in blood and in the erythrocyte suspensions," Ultrasonics **30**, 26–34.
Zinin, P. V., Allen, J. S., and Levin, V. M. (**2005**). "Mechanical resonances of bacterial cells," Phys. Rev. E **72**, 061907.

Off-axis effects on the multi-pulse structure of sperm whale coda clicks

Tyler M. Schulz[a] and Hal Whitehead
Department of Biology, Dalhousie University, Halifax, Nova Scotia B3H 4J1, Canada

Luke Rendell
School of Biology, University of St. Andrews, St. Andrews, Fife KY16 9TS, Scotland

(Received 9 March 2008; revised 13 December 2008; accepted 29 December 2008)

Sperm whale (*Physeter macrocephalus*) clicks have a multi-pulse structure, a result of the reflection of sound energy between air sacs in the spermaceti organ. Although previous research revealed that usual clicks (used for echolocation) recorded away from a vocalizing whale's longitudinal axis have waveforms with poorly defined pulse structures, it has been unknown whether sperm whale coda clicks (used for communication) show similar off-axis effects. To address this knowledge gap, a hydrophone array was used to localize vocalizing sperm whales, and the waveforms of coda clicks recorded from different aspects were examined. Coda clicks recorded close to the whale's acoustic axis showed well-defined multi-pulsed waveforms, while those recorded off-axis did not. As for usual clicks, this suggests that sound energy radiates directly into the water upon reflection off the frontal sac. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3075598]

PACS number(s): 43.80.Ka [WWA] Pages: 1768–1773

I. INTRODUCTION

The largest toothed whale, the sperm whale (*Physeter macrocephalus*), demonstrates extreme differences in morphology and life-history, both within the species and in comparison to other cetaceans. However, no feature of the sperm whale is perhaps as noticeably extreme and fascinating as its large nasal complex. Its large nose, which contains the spermaceti organ, junk bodies, and other organs associated with sound production (Fig. 1), makes up approximately 1/3 of the sperm whale's total body weight and body length (Rice, 1989), giving this species the claim to the "biggest nose on record" (Raven and Gregory, 1933).

Although other functions had previously been proposed for the hypertrophied nasal complex (see Clarke, 1970, 1978; Carrier *et al.*, 2002), Norris and Harvey (1972) were the first to advance a sound generating function. They suggested that an initial sound pulse generated by the forcing of air through the museau de singe (or phonic lips) (Fig. 1) is reflected between air sacs at the anterior and posterior ends of the spermaceti organ (Norris and Harvey, 1972), resulting in the observed multi-pulsed structure of sperm whale clicks. This initial theory was revised as the "bent horn" theory (Møhl *et al.*, 2003) to explain the weak initial pulse (p0) and powerful subsequent pulse (p1) obvious in the recordings of usual clicks (echolocation clicks) recorded from in front of the vocalizing whale [Fig. 2(a)].

Recent research confirms the bent-horn theory; sound in a usual click is produced at the museau de singe using a pressure differential (Madsen *et al.*, 2003), and a fraction of the initial sound energy leaks directly into the water as the weak initial pulse (Møhl, 2001). The majority of the sound energy, however, is reflected backward into the spermaceti organ (Fig. 1) (Zimmer *et al.*, 2005b) and is subsequently reflected off the air-filled frontal sac at the posterior of the spermaceti organ and focused in the junk complex before emission into the water as the powerful p1 pulse (Fig. 1; Cranford, 1999; Møhl *et al.*, 2000, 2003; Zimmer *et al.*, 2005a). The multi-pulsed structure of usual clicks is related to the two-way travel time between the air sacs (Møhl, 2001; Møhl *et al.*, 2003).

Just as sound energy is leaked at the anterior end of the spermaceti organ when the initial p0 pulse is produced, the reflection of usual click sound energy on the frontal sac at the posterior end of the nasal complex also involves the leakage of sound energy into the water, resulting in the emission of a p1/2 pulse (Zimmer *et al.*, 2005a). When a usual click is recorded on-axis directly in front of a vocalizing whale, the p1/2 pulse merges with the p1 pulse (Zimmer *et al.*, 2005a). Conversely, when a usual click is recorded on-axis directly behind the vocalizing whale, the p1/2 pulse merges with the p0 pulse [Fig. 2(b)] (Zimmer *et al.*, 2005a).

In either case, the recorded waveform contains distinct and regular inter-pulse intervals (IPIs) that can be measured to estimate the length of the vocalizing whale (Gordon, 1991b; Rhinelander and Dawson, 2004). However, when usual clicks are recorded off-axis, the p1/2 pulse will appear in the far-field with a delay between 0 and the two-way travel time of the spermaceti organ, sometimes resulting in IPIs that cannot be distinguished [see Fig. 2(c)] and thus the incorrect or impossible estimation of whale length (Zimmer *et al.*, 2005a).

Unlike usual clicks, which exhibit a highly directional p1 pulse with relatively few and weak subsequent pulses [Figs. 2(a) and 2(b)], clicks produced in short stereotyped communication patterns termed "codas" (Watkins and Schevill, 1977) typically exhibit many successive pulses

[a]Present address: 301-60 Homewood Avenue, Toronto, ON M4Y 2X4, Canada. Electronic mail: tmschulz@dal.ca

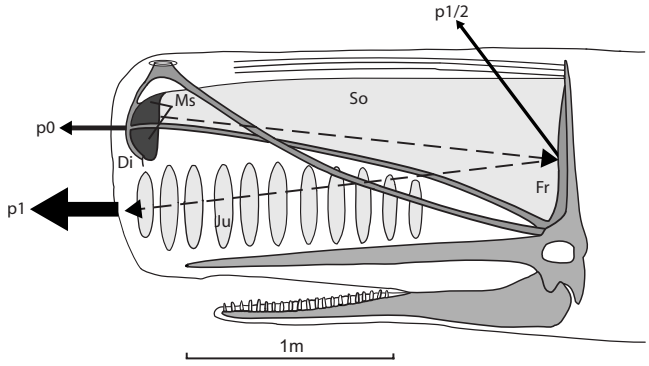


FIG. 1. Schematic view of the head of a sperm whale depicting the bent-horn model of usual click sound generation (adapted from Zimmer *et al.*, 2005a). The dashed arrows indicate the primary sound path within the nasal complex according to the modified Norris and Harvey (1972) theory. The solid arrows indicate the emission of the weak pulse (**p0**) from the phonic lips/museau de singe (**Ms**), the emission of the highly directional sonar pulse (**p1**) from the junk (**Ju**), and the leakage of sound energy as the p1/2 pulse (**p1/2**) from the frontal air sac (**Fr**). **D**, distal air sac; **So**, spermaceti organ.

(Fig. 3) and thus a longer overall click duration (Madsen *et al.*, 2002). The lower decay rate of coda clicks suggests that coda click sound energy is retained within the spermaceti organ to reverberate repeatedly between the air sacs rather than redirected into the junk complex to be released as a powerful and directional pulse (Madsen *et al.*, 2002).

The observed differences in pulse structure between usual and coda clicks must result from internal differences in the structure of the sound production apparatus. One obvious way this could be achieved is through changes in the amount of air in the frontal and distal air sacs; for example, the introduction of air into the frontal air sac could keep higher levels of energy reverberating within the spermaceti organ, consistent with the extended pulse structure of coda clicks. If

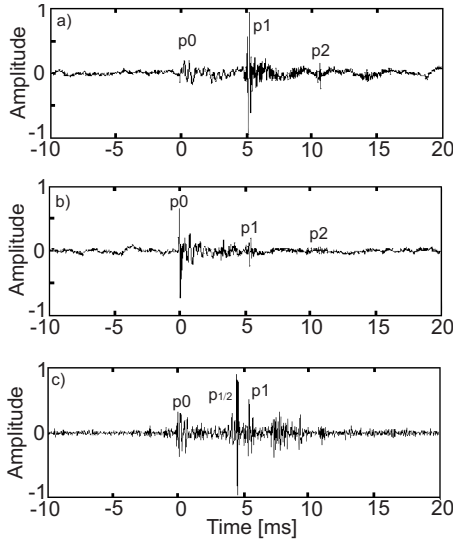


FIG. 2. Usual sperm whale clicks recorded from a remote receiver from (a) in front, (b) behind, (c) and off the acoustic axis of a vocalizing whale (adapted from Zimmer *et al.*, 2005a). The different component pulses in the clicks are denoted by p0, p1, and p2. Note that in (a) and (b), a single pulse by far dominates the energy content of the click. In the waveform recorded off-axis (c), p1/2 denotes the click energy leaked from the spermaceti organ at the frontal sac.

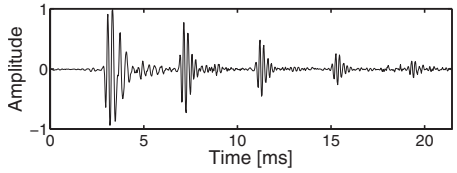


FIG. 3. Waveform of a coda click recorded from behind a vocalizing whale. Note the multiple pulses with a low rate of decay compared to the nearly mono-pulsed waveform of usual clicks [see Figs. 2(a) and 2(b)].

this were the case, we would also expect to observe significant sound leakage from the frontal sac area. Although the waveforms of usual clicks have been examined from different recording aspects (Zimmer *et al.*, 2005a; Madsen *et al.*, 2002), it is unknown whether the pulse structure of coda clicks also demonstrates off-axis effects indicative of sound energy leakage at the frontal sac. Most codas are recorded from an unknown recording aspect or from behind the vocalizing whale (see Marcoux *et al.*, 2006), prohibiting the observation of off-axis effects. To address this research gap, we inspected and compared the waveforms of coda clicks recorded on different hydrophones in a passive dynamic acoustic array.

II. METHODS

A. Field methods

We conducted fieldwork from a 40-ft sailboat, *Balaena*, between May 5 and June 20, 2004 (38 days effort) in international waters between Bermuda and the east coast of the United States in the Sargasso Sea. Encountered sperm whales were tracked visually during the day and acoustically at night using a directional hydrophone (see Whitehead and Gordon, 1986). During the day, if whales at the surface were moving slowly (<1 knot) and the weather conditions were favorable, we deployed a dynamic acoustic array.

B. Acoustic array

This localization system consisted of several small battery-powered remotely-piloted vessels (RPVs) as well as the primary research platform from which they were launched. From the side of each recording platform was suspended a hydrophone (Vemco VHLF; frequency response: 200 Hz–20 kHz ± 3 dB; midband sensitivity: 147 dB re 1 V/μPa) at approximately 80 cm below the water surface. On each RPV, acoustic signals from the hydrophone were amplified, high-pass filtered at 1 kHz, and broadcast by a FM transmitter (NRG Kits PLL PRO III). This signal was then received by a digital AM/FM PLL synthesized radio (SONY ICF-M260) onboard the deployment platform and digitally recorded on a multi-track recorder (FOSTEX VF-160; sampling rate: 44.1 kHz), which simultaneously recorded the acoustic signals detected by each of the hydrophones in the array. On each recording platform, a global positioning system (GPS) unit (Garmin GPS25-HVS) logged positional data to a flashcard. A frequency shift keying (FSK) modulator transformed the stream of ASCII sentences from the GPS unit onboard the research platform to an amplitude-modulated tonal signal (see Møhl *et al.*, 2001), which was

recorded as an acoustic track on the multi-track recorder in synchrony with the hydrophone signals. Subsequent demodulation of the FSK timestamp during analysis allowed for synchronization of the acoustic and positional data (Møhl *et al.*, 2001). The same hydrophone depth and filtering were used on the deployment platform. Recording sessions were labeled numerically according to month, day, and session of the day (e.g., 051403 was the third recording session on May 14).

During array deployment, the locations of the RPVs and whales relative to the primary research platform were recorded on a digital camcorder (SONY DCR-PC 105) from the sailboat's crow's nest. Sea surface temperature was measured using an onboard electronic thermometer, and sea salinity was estimated using a refractometer.

C. Localization analysis

The binary GPS file logged on each recording platform was converted to a RINEX file and submitted to an online Precise Point Positioning processor (Canadian Geodetic Service) to improve the accuracy of the positions. Further exclusion of erroneous noise in GPS positions was achieved by discarding fixes obtained using less than seven satellites and by smoothing the x-coordinates and y-coordinates for each GPS receiver by fitting quadratic equations to time segments spanning several seconds before and after each epoch in the record (see Christal and Whitehead, 2001).

Acoustic recordings were inspected for codas that were detected on at least three of the four hydrophones in the array. Clicks were marked in these codas using a dedicated software package, RAINBOW CLICK (see Gillespie, 1997; Jaquet *et al.*, 2001), and the click data from each recording were output to a custom-written routine in MATLAB® (Mathworks) for the calculation of time of arrival differences (TOADs) between each pair of hydrophone receivers (see Wahlberg *et al.*, 2001). Because sperm whales produce loud, abrupt broadband clicks, TOADs were calculated as time differences between hydrophones in the click onset. Since the sperm whales were observed from the sailboat, the whales were assumed to be at or near the surface when vocalizing, and we localized recorded clicks in two dimensions.

For each click in each analyzed coda, an equal time-difference two-dimensional hyperbola was calculated for each TOAD using the relative locations of the receivers and the speed of sound in water (calculated using sea surface temperature and salinity in the Leroy equation; Urick, 1983) (see Wahlberg *et al.*, 2001). The intersections of these hyperbolae were averaged to estimate the location of the sound source for each click. The localization method used was the same as the MINNA (minimum number of receiver array) method described by Wahlberg *et al.* (2001) except that it repeated the MINNA method for each pair of intersecting hyperbolae and averaged the intersections to give a solution that accounts for measurement error (see Janik, 2000; Laurinolli *et al.*, 2003). The location of each coda was estimated as the average of the locations of each of its clicks.

For clicks localized using four hydrophones, the error in each click's location was estimated from the standard deviation of the hyperbola intersections in the zonal (ε_x) and the meridional (ε_y) directions (as in Laurinolli *et al.*, 2003). The error bars for each coda localized with four hydrophones were then calculated by taking the mean of each of these errors (zonal and meridional) over the clicks in the coda. A calibration of this system estimated the precision of estimated locations as approximately 0.5 m within the array (Schulz *et al.*, 2006).

D. IPI assignment of codas

To estimate the recording angle between hydrophones and the acoustic axis of vocalizing whales, whale trajectories were estimated by localizing successive codas with similar IPIs. The IPIs of localized clicks were calculated using a modified version of a previously described IPI analysis method (see Schulz *et al.*, 2008). This modified method extracts the maximum cross-correlation peak, rather than the absolute cross-correlation peak (used by Gordon, 1991a), between pulses for clicks with well-defined pulse structures, allowing the user to discard clicks with distorted pulse structures (see Schulz *et al.*, 2008). The IPI for each coda was calculated by taking the mode of the clicks in that coda. Codas with modal IPIs within 0.05 ms of one another were assumed to have been produced by the same whale (Schulz *et al.*, 2008). This conservative criterion was used since other analyses indicated that within recordings, the IPIs of codas produced by the same whale (and analyzed using these methods) differ by no more than 0.05 ms (see Schulz *et al.*, 2008). Because the clarity of the pulse structure of coda clicks sometimes varied between acoustic channels, IPI analysis was repeated for each hydrophone in the array, thereby increasing the number of codas for which an IPI could be obtained. The assignment of IPI values to codas was consistent between hydrophone receivers.

E. Determination of whale trajectory and orientation relative to receivers

Successive codas likely produced by the same whale (as determined by IPI similarity) were localized and plotted to estimate the whale's approximate trajectory and orientation. We reasoned that if successive codas are produced every few seconds and are localized in a relatively straight line, fitting a line through the estimated locations should give a reasonable approximation of the vocalizing whale's acoustic axis. Using the GPS positions of the hydrophone receivers and the estimated trajectory and orientation of the vocalizing whale, the cosine rule was used to calculate the approximate angle between the whale and each hydrophone receiver relative to the whale's presumed body axis. The waveforms of clicks in these localized codas were then visually inspected in a standard sound-editing program (COOL EDIT, Syntrillium) and compared between hydrophone receivers to determine whether there were differences in pulse structure as a result of recording aspect.

To quantitatively describe the clarity in the multi-pulse structure of recorded coda click waveforms, we calculated the coefficient of variation (CV) in the squared amplitude over each sample in each recorded click within a 5 ms time

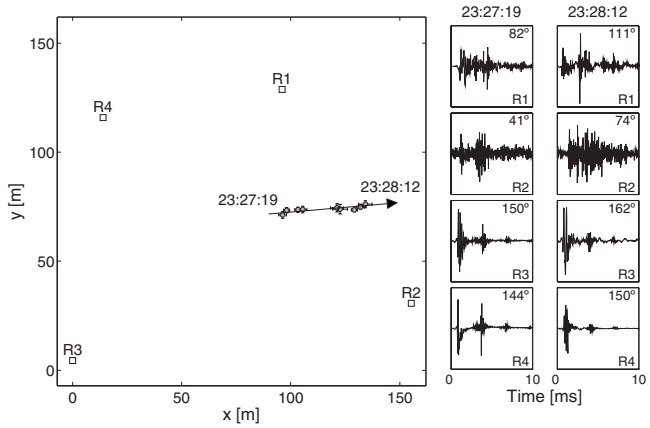


FIG. 4. The GPS positions of four hydrophone receivers (R1–R4; □) at 23:27:19 UTC and the estimated location solutions (with error bars of two standard deviations in the zonal and meridional directions) for codas with IPIs of 2.95 ms (●) produced throughout recording session 051403. The arrow indicates the estimated heading of the vocalizing whale based on codas localized between 23:27:19 and 23:28:12. The waveform of coda clicks as received on each of the four hydrophone receivers at 23:27:19 and 23:28:12 are presented along the right side of the figure together with the calculated angle between each hydrophone receiver and the estimated acoustic axis of the whale. The hydrophone deployed from the primary research platform is designated as R1. Note: the initial pulse of the click depicted in the hydrophone R4 waveform is a result of an anomalous occurrence where the FM receiver also picked up the hydrophone signal transmitted by another FM transmitter.

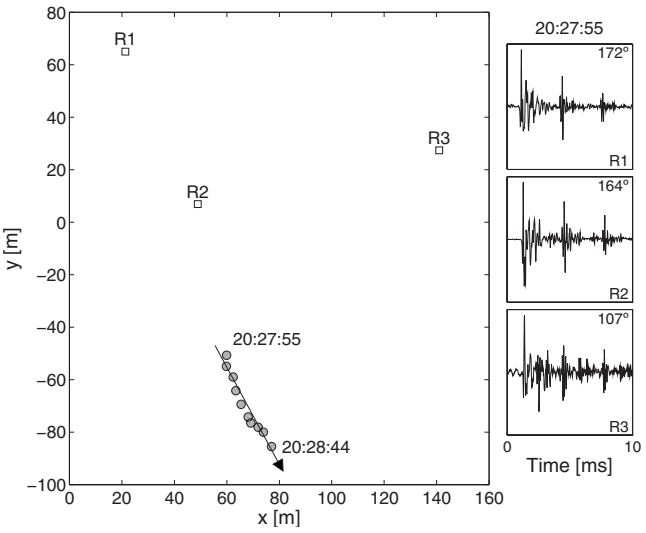


FIG. 5. The GPS positions of three hydrophone receivers (R1–R3; □) at 20:27:55 and the estimated location solutions for codas with IPIs of 3.24–3.27 ms (●) produced throughout recording session 061002. The arrow indicates the estimated heading of the vocalizing whale based on codas localized between 20:27:55 and 20:28:44. The waveform of a coda click as received on each of the three hydrophone receivers at 20:27:55 is presented along the right side of the figure together with the calculated angle between the hydrophone receiver and the estimated acoustic axis of the whale. The hydrophone deployed from the primary research platform is designated as R1.

interval beginning at click onset. A 5 ms (221 sample) time interval was used because the IPIs of all localized whales were estimated as less than 5 ms and because click durations were variable but generally longer than 5 ms. Because clicks with distinct initial pulses possessed high CVs and clicks with poorly defined pulse structures possessed low CVs, the CV provided a general measure of the clarity of the click structure while standardizing for the relative amplitude of the recorded click.

III. RESULTS

In recording session 051403, a whale with an IPI of 2.95 ms was localized as it moved within the array toward the periphery (Fig. 4). The structure of coda clicks produced while the whale was near the center of the array was clearly multi-pulsed in recordings made on hydrophones positioned behind the vocalizing whale (R3 and R4; Fig. 4). However, the pulse structure of the same clicks but recorded on a hydrophone receiver (R1) in an off-axis aspect was poorly defined (Fig. 4). Moreover, the pulse structure of the same clicks but recorded slightly more on-axis in front of the vocalizing whale (hydrophone receiver R2) demonstrated a clear initial pulse but a less-defined succeeding pulse (Fig. 4).

Several seconds later at 23:28:12, after the whale had moved approximately 38 m toward the periphery of the array, the waveforms of coda clicks recorded on hydrophones from behind the whale still demonstrated a clear multi-pulsed structure (Fig. 4). For hydrophone R1, which was now 111° behind the vocalizing whale, the waveforms of recorded coda clicks were slightly more multi-pulsed (Fig. 4) than for the coda clicks recorded 53 s earlier when this hydrophone was at an angle of 82° in front of the animal (Fig. 4). Conversely, for hydrophone R2, which was now 74° in front of the animal and thus at a more off-axis angle than 53 s earlier, the multi-pulsed structure of the click waveforms was much less discernible (Fig. 4).

In another recording session, 061002, a whale with an IPI of 3.24 ms was localized moving away from the 3-receiver array, nearly inline with two hydrophone receivers (R1 and R2) while off-axis to the third hydrophone receiver (R3) (Fig. 5). In the two recordings made from behind the vocalizing whale (R1: 172°; R2: 164°), the waveforms of coda clicks recorded at 20:27:55 possessed well-defined pulse structures, although the first pulse in the clicks was more elongated and less distinct than the subsequent pulses (Fig. 5). In contrast, in the recording made from an off-axis aspect (107°) to the acoustic axis of the vocalizing whale, the waveforms of the same clicks demonstrated additional pulses between the primary pulses (Fig. 5). We note, however, that since this whale was localized outside the array where errors can increase markedly (see Watkins and Schevill, 1972), these results should be considered with caution. Nonetheless, similar differences in waveforms on different hydrophone receivers were observed for another whale in session 061002 with an IPI of 3.51 ms also localized moving away from the 3-receiver array.

To examine the effect of recording angle on the clarity of pulse structure, we also plotted the CV of the squared amplitude for each localized click waveform on each hydrophone receiver against the estimated angle between the location of that hydrophone and the acoustic axis of the whale at the time at which the click was produced. Figure 6 quantitatively illustrates the qualitative observation described above

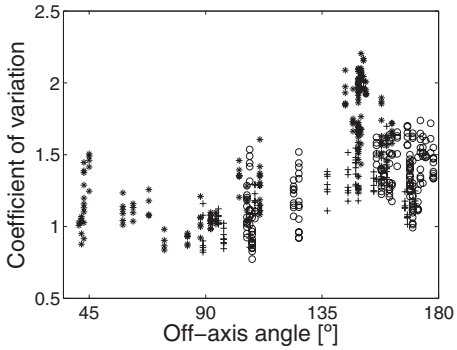


FIG. 6. Scatterplot of the CV (in the squared amplitude of the initial 221 samples) in coda click waveforms vs the estimated angle between the hydrophone receiver and the acoustic axis of the whale. The CV was calculated for 652 click waveforms from 33 different codas from three whales in two recording sessions (051403 and 061002). Each whale is represented by a different symbol.

that coda clicks recorded off-axis tended to possess less well-defined waveforms than those recorded closer to the acoustic axis. Although some click waveforms that were recorded on-axis possessed low CVs (see Fig. 6), such waveforms possessed clear, well-defined pulse structures but poorly defined initial pulses (e.g., see Fig. 7), thereby resulting in a low CV during the initial 5 ms of the click. Poorly defined initial pulses in otherwise well-defined clicks may have been a result of the initial release of the sound energy into the water or the distortion of the initial pulse by the p1/2 pulse.

IV. DISCUSSION

Examination of coda click waveforms recorded from different aspects indicates that sperm whale coda clicks, like usual clicks, are affected by recording orientation. For several different localized whales, waveforms of coda clicks recorded on or near the whale's estimated acoustic axis visually demonstrated a much more well-defined pulse structure than the same clicks recorded off-axis. Furthermore, plotting the CV in amplitude of localized clicks against the estimated angle of recording indicated that clicks recorded off-axis (~90°) tended to have more poorly defined pulse structures than those recorded closer to the acoustic axis.

Differences in click waveforms between hydrophones are clearly due to differences in recording aspect and not an artifact of variation in recording quality between different hydrophones since recordings from receivers R1 and R2 in session 051403 demonstrated poor click waveforms for the clicks of one whale recorded off-axis within the array (Fig. 4) but demonstrated distinct multi-pulsed waveforms for the

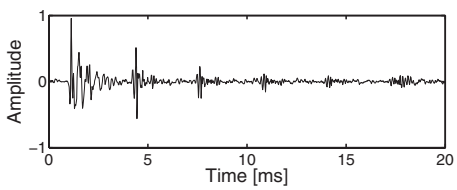


FIG. 7. Waveform of a coda click recorded at an estimated angle of 172° between the hydrophone receiver and the acoustic axis of the whale. Although the pulse structure of the waveform is well-defined, it possesses a relatively low CV in squared amplitude due to the indistinct initial pulse.

clicks of another whale recorded on-axis in session 061002 out of the array (Fig. 5). Moreover, there was also consistency within recording sessions in the clarity of pulse structure between recordings made from similar recording aspects, again indicating that the waveform of recorded coda clicks is dependent on the angle between the hydrophone receiver and the orientation of the vocalizing whale rather than the quality of the hydrophone recording.

If the sound energy of coda clicks were emitted only from the anterior end of the spermaceti organ, one would find the structure of recorded clicks to be similar at all recording aspects. The results presented here that sperm whale coda clicks recorded off-axis are much less defined in pulse structure than clicks recorded on-axis indicate that the sound energy does not exit solely from the front of the spermaceti organ and must also be leaked as a p1/2 pulse from some other point, most likely upon reflection of the frontal sac as in usual clicks.

Zimmer *et al.* (2005a) mistakenly stated that field observations by Rendell and Whitehead (2004) suggest that codas recorded in the far-field have stable IPIs. The finding here of off-axis effects clearly indicates otherwise and partially explains the considerable number of coda clicks discarded in previous IPI analyses of coda clicks (Rendell and Whitehead, 2004; Marcoux *et al.*, 2006). Moreover, the results presented here should encourage researchers to record sperm whale codas on-axis, most likely from behind the whale, and exclude from IPI analysis coda clicks recorded off-axis with poor pulse structure. Although ensuring on-axis recordings with a single hydrophone can be difficult, using a number of hydrophones in an acoustic array can increase the likelihood of recording coda clicks from an on-axis aspect and thus of obtaining clear IPIs for the estimation of body length (Marcoux *et al.*, 2006) or assignment of codas to specific whales (Schulz *et al.*, 2008).

ACKNOWLEDGMENTS

The authors wish to acknowledge the valuable assistance and suggestions given by M. Wahlberg concerning acoustic localization. We also wish to thank R. Lang, T. Beran, and D. Heffler for their advice and assistance with GPS processing. We are grateful to the International Fund for Animal Welfare and, in particular, to D. Gillespie for allowing us to use RAINBOW CLICK. We also appreciate the assistance of S. Gero, M. Marcoux, and E. Johnson with calibration tests and fieldwork. The equipment and fieldwork for this project were funded by Natural Sciences and Engineering Research Council of Canada grants to H.W. T.M.S. was supported by a NSERC pre-doctoral scholarship and an Izaak Walton Killam Memorial Scholarship. L.R. was supported by a NERC Postdoctoral Fellowship (No. NER/I/S/2002/00632)

Carrier, D. R., Deban, S. M., and Otterstrom, J. (**2002**). "The face that sunk the *Essex*: Potential function of the spermaceti organ in aggression," J. Exp. Biol. **205**, 1755–1763.

Christal, J., and Whitehead, H. (**2001**). "Social affiliations within sperm whale (*Physeter macrocephalus*) groups," Ethology **107**, 323–340.

Clarke, M. R. (**1970**). "The function of the spermaceti organ of the sperm whale," Nature (London) **228**, 873–874.

Clarke, M. R. (**1978**). "Buoyancy control as a function of the spermaceti

organ in the sperm whale," J. Mar. Biol. Assoc. U.K. **58**, 27–71.
Cranford, T. W. (**1999**). "The sperm whale's nose: Sexual selection on a grand scale?" Marine Mammal Sci. **15**, 1133–1157.
Gillespie, D. (**1997**). "An acoustic survey for sperm whales in the Southern Ocean sanctuary conducted from the R/V Aurora Australia," Rep. Int. Whal. Comm. **47**, 897–908.
Gordon, J. C. D. (**1991a**). "Evaluation of a method for determining the length of sperm whales (*Physeter catodon*) from their vocalizations," J. Zool., (Lond.) **224**, 301–314.
Gordon, J. C. D., (**1991b**). "The World Wildlife Fund's Indian Ocean sperm whale project: An example of cetacean research within the Indian Ocean Sanctuary," in *Cetaceans and cetacean research in the Indian Ocean Sanctuary, UNEP Technical Report Number 3*, edited by S. Leatherwood and G. P. Donovan (United Nations Environment Programme, Nairobi, Kenya), pp. 219–239.
Janik, V. M. (**2000**). "Whistle matching in wild bottlenose dolphins (*Tursiops truncatus*)," Science **289**, 1355–1357.
Jaquet, N., Dawson, S., and Douglas, L. (**2001**). "Vocal behavior of male sperm whales: Why do they click?" J. Acoust. Soc. Am. **109**, 2254–2259.
Laurinolli, M. H., Hay, A. E., Desharnais, F., and Taggart, C. T. (**2003**). "Localization of North Atlantic right whale sounds in the Bay of Fundy using a sonobuoy array," Marine Mammal Sci. **19**, 708–723.
Madsen, P. T., Carder, D. A., Au, W. W., Nachtigall, P. E., Møhl, B., and Ridgway, S. H. (**2003**). "Sound production in neonate sperm whales," J. Acoust. Soc. Am. **113**, 2988–2991.
Madsen, P. T., Payne, R., Kristiansen, N. U., Wahlberg, M., Kerr, I., and Møhl, B. (**2002**). "Sperm whale sound production studied with ultrasound time/depth-recording tags," J. Exp. Biol. **205**, 1899–1906.
Marcoux, M., Whitehead, H., and Rendell, L. (**2006**). "Coda vocalizations recorded in breeding areas are almost entirely produced by mature female sperm whales (*Physeter macrocephalus*)," Can. J. Zool. **84**, 609–614.
Møhl, B. (**2001**). "Sound transmission in the nose of the sperm whale *Physeter catodon*. A post mortem study," J. Comp. Physiol. [A] **187**, 335–340.
Møhl, B., Wahlberg, M., and Heerfordt, A. (**2001**). "A large-aperture array of nonlinked receivers for acoustic positioning of biological sound sources," J. Acoust. Soc. Am. **109**, 434–437.
Møhl, B., Wahlberg, M., Madsen, P. T., Heerfordt, A., and Lund, A. (**2003**). "The monopulsed nature of sperm whale clicks," J. Acoust. Soc. Am. **114**, 1143–1154.
Møhl, B., Wahlberg, M., Madsen, P. T., Miller, L. A., and Surlykke, A. (**2000**). "Sperm whale clicks: Directionality and source level revisited," J. Acoust. Soc. Am. **107**, 638–648.
Norris, K. S., and Harvey, G. W., (**1972**). "A theory for the function of the spermaceti organ of the sperm whale (*Physeter catodon* L.)," in *Animal Orientation and Navigation*, edited by S. R. Galler, K. Schmidt-Koenig, G. J. Jacobs, and R. E. Belleville (NASA, Washington, D.C., pp. 397–417.
Raven, H. C., and Gregory, W. K. (**1933**). "The spermaceti organ and nasal passages of the sperm whale (*Physeter catodon*) and other odontocetes," American Museum Novitates **677**, 1–18.
Rendell, L. E., and Whitehead, H. (**2004**). "Do sperm whales share coda vocalizations? Insights into coda usage from acoustic size measurements," Anim. Behav. **67**, 865–874.
Rhinelander, M. Q., and Dawson, S. M. (**2004**). "Measuring sperm whales from theirclicks: Stability of inter-pulse intervals and validation that they indicate whale length," J. Acoust. Soc. Am. **115**, 1826–1831.
Rice, D. W., (**1989**). "The sperm whale (*Physeter macrocephalus* Linnaeus, 1758)," in *Handbook of Marine Mammals*, edited by S. H. Ridgway and R. Harrison (Academic, London), Vol. **4**, pp. 177–233.
Schulz, T. M., Whitehead, H., Gero, S., and Rendell, L. (**2008**). "Overlapping and matching of codas in vocal interactions between sperm whales: Insights into communication function," Anim. Behav. **76**, 1977–1988.
Schulz, T. M., Whitehead, H., and Rendell, L. (**2006**). "A remotely-piloted acoustic array for studying sperm whale vocal behaviour," Can. Acoust. **34**, 54–55.
Urick, R. J. (**1983**). *Principles of Underwater Sound*, 3rd ed. (McGraw-Hill, New York).
Wahlberg, M., Møhl, B., and Madsen, P. T. (**2001**). "Estimating source position accuracy of a large-aperture hydrophone array for bioacoustics," J. Acoust. Soc. Am. **109**, 397–406.
Watkins, W. A., and Schevill, W. E. (**1972**). "Sound source location by arrival times on a non-rigid three-dimensional hydrophone array," Deep-Sea Res. **19**, 691–706.
Watkins, W. A., and Schevill, W. E. (**1977**). "Sperm whale codas," J. Acoust. Soc. Am. **62**, 1485–1490.
Whitehead, H., and Gordon, J. (**1986**). "Methods of obtaining data for assessing and modelling sperm whale populations which do not depend on catches," Rep. Int. Whal. Comm. **8**, 149–166.
Zimmer, W. M. X., Madsen, P. T., Teloni, V., Johnson, M. P., and Tyack, P. L. (**2005a**). "Off-axis effects on the multi-pulse structure of sperm whale usual clicks with implications for sound production," J. Acoust. Soc. Am. **118**, 3337–3345.
Zimmer, W. M. X., Tyack, P. L., Johnson, M. P., and Madsen, P. T. (**2005b**). "Three-dimensional beam pattern of regular sperm whale clicks confirms bent-horn hypothesis," J. Acoust. Soc. Am. **117**, 1473–1485.

Geographic variation in Northwest Atlantic fin whale (*Balaenoptera physalus*) song: Implications for stock structure assessment

Julien Delarue[a)]
College of the Atlantic, 3078 George Dauphinee Avenue, Halifax, Nova Scotia B3L3S9, Canada

Sean K. Todd
College of the Atlantic, 105 Eden Street, Bar Harbor, Maine 04609

Sofie M. Van Parijs
Northeast Fisheries Science Center, 166 Water Street, Woods Hole, Massachusetts 02543-1026

Lucia Di Iorio
Zoologisches Institut, Universität Zurich, Rämistrasse 71, 8006 Zurich, Switzerland

(Received 12 June 2008; revised 23 November 2008; accepted 12 December 2008)

Passive acoustic data are increasingly being used as a tool for helping to define marine mammal populations and stocks. Fin whale (*Balaenoptera physalus*) songs present a unique opportunity to determine interstock differences. Their highly stereotyped interpulse interval has been shown to vary between geographic areas and to remain stable over time in some areas. In this study the structure of songs recorded at two geographically close feeding aggregations in the Gulf of St. Lawrence (GSL) and Gulf of Maine (GoM) was compared. Recordings were made from September 2005 through February 2006 in the GSL and intermittently between January 2006 and September 2007 at two locations in the GoM. 6257 pulse intervals corresponding to 19 GSL and 29 GoM songs were measured to characterize songs from both areas. Classification trees showed that GSL songs differ significantly from those in the GoM. The results are consistent with those derived from other stock structure assessment methodologies, such as chemical signature and photoidentification analysis, suggesting that fin whales in these areas may form separate management stocks. Song structure analysis could therefore provide a useful and cost-efficient tool for defining conservation units over temporal and geographical scales relevant to management objectives in fin whales. [DOI: 10.1121/1.3068454]

PACS number(s): 43.80.Ka, 43.30.Sf [WWA] Pages: 1774–1782

I. INTRODUCTION

Effective management of marine mammals requires a working knowledge of their stock structure. In this paper, the word "stock" is used in a management context to refer to groups of individuals of the same species that are demographically, but not necessarily genetically, isolated (Taylor, 1999; Taylor, 2005; Clapham *et al.*, 2008). Dispersal rates between theorized stocks, though high enough to mask genetic differences, may not be sufficient for enough recruits to migrate from an area experiencing little to no anthropogenic removal to an area subject to high mortality to prevent the decline or extirpation of the latter. Accordingly, it has been suggested that management of large whales should be based on units that would not recover within a given time frame if extirpated (Clapham *et al.*, 2008). The identification and delineation of such units, whose recruitment is primarily affected by local births and deaths in comparison to immigration and emigration, are a key to successful management.

Stock structure can be assessed using population trends, tagging, photoidentification, differences in parasites and contaminant loads, or morphological and demographic data collected on harvested animals (Taylor, 1999; Dizon *et al.*, 1992; Gorbics and Bodkin, 2001). Genetics is also a valued assessment method because genetic differences indicate reproductive isolation, which is in turn often proof of demographic isolation. Considered individually, each of these methods can lack the power to detect differences of importance to management, and a number of authors have called for a multidisciplinary approach to assess stock structure (Donovan, 1991; Mellinger and Barlow, 2003; Clapham *et al.*, 2008). For instance, a combination of photoidentification studies and genetic data was necessary to define an accurate population model for Northwest Atlantic humpback whales (*Megaptera novaeangliae*) (Smith *et al.*, 1999). One possible explanation for the failure of population assessments based solely on genetics is that genetic data represent how feeding aggregations interacted historically and are thus only partially relevant to assessing present day differentiation within this species (Clapham and Palsbøll, 1999; Taylor, 1999). This example demonstrates one of the main weaknesses of genetic-based stock assessment, specifically that genetic data reflect millennia of evolution and may not evolve quickly enough to track structural changes that are temporally meaningful to management, whose time frame is

[a)]Electronic mail: juldel30@hotmail.com

often decadal at most (McDonald *et al.*, 2006; Clapham *et al.*, 2008).

Among the alternatives recommended to conduct or improve population structure assessments, acoustic methods are among those most commonly cited as worthy of further investigation (Mellinger and Barlow, 2003; McDonald *et al.*, 2006). For example, Mellinger and Barlow (2003) suggested assessing population structure using acoustics and then using a secondary method—such as genetics—to ground-truth assessment. More importantly, they noted the importance of examining the relationship between acoustic and genetic differences in cases where acoustic metrics are used to assert population structure.

The case of using baleen whale song as a proxy for stock identity has been elegantly made by Hatch and Clark (2004) and McDonald *et al.* (2006). It relies essentially on the growing body of evidence indicating regional variations in vocalizations in a number of terrestrial and marine species, including several marine mammal species (see McDonald *et al.*, 2006). Although some researchers argue that acoustic signals may be too susceptible to environmental or social factors to be stable and, thus, reliable indicators of stock structure (Harvey and Pagel, 1991), they may reflect intraspecific variation that is relevant to management but too recent to be apparent in the genome (Hatch and Clark, 2004). More particularly, in species such as fin whales where genetic assessment of population structure is hindered by logistic difficulties in collecting sufficiently large sample sizes of tissue, geographic variations in songs may offer a good alternative, provided they track meaningful biological units from a management point of view.

Broad-scale acoustic population assessments, largely without genetic ground-truthing, have already been conducted or contemplated for several species. McDonald *et al.* (2006) reported on the existence of nine distinctive blue whale (*Balaenoptera musculus*) songs worldwide and suggested using this acoustic population framework for management purposes. Humpback whale songs also vary between, and to a limited extent within, oceanic basins (Winn *et al.*, 1981) and have been suggested as a stock indicator (Payne and Guinee, 1983). Sperm whale (*Physeter macrocephalus*) populations may be best managed at the clan level, and clans can be easily identified by their codas (Rendell and Whitehead, 2003). Killer whale (*Orcinus orca*) dialects have also been proposed as stock indicators (Ford and Fisher, 1982). Finally, fin whales (*B. physalus*) are known to produce geographically distinctive songs, and several authors have hypothesized using this to identify populations (Watkins *et al.*, 1987; Edds, 1988; Thompson *et al.*, 1992; Clark *et al.*, 2002; Hatch and Clark, 2004). However, their observations have typically targeted individuals at distant sites, thereby maximizing the possibility of comparing geographically isolated populations. To date, no studies have compared the structure of songs produced in proximate areas such as adjacent feeding grounds.

Central questions regarding fin whale stock structure remain unanswered (Donovan, 1991; Coakes *et al.*, 2005), primarily as a result of their pelagic distribution and their typically large, sometimes overlapping, and still partly unknown ranges (Kellogg, 1929; Mizroch *et al.*, 1984). Historically, fin whale stock assessment in the North Atlantic has relied heavily on biological and demographic data from harvested animals (Bérubé *et al.*, 2006). More recently, genetic studies have revealed a stock structure characterized by isolation-by-distance (Bérubé *et al.*, 1998) but failed to detect genetic divergence at distances smaller than across oceanic basins. Within this context, the investigation of stock identity using acoustic signatures in fin whales offers great promise: fin whale low-frequency (LF) calls have the potential to propagate over long distances (Payne and Webb, 1971), thus maximizing their detection, and fin whales are the most commonly recorded species in the North Atlantic (Clark, 1995). In addition, the structure of their songs—repetitive sequence of pulses centered around 20 Hz and characterized by a highly stereotyped pulse interval (PI) (e.g., Watkins *et al.*, 1987)—varies geographically (Watkins *et al.*, 1987; Thompson *et al.*, 1992) and yet appears to remain stable over time in some areas (Hatch and Clark, 2004).

In the Northwest Atlantic, the International Whaling Commission (IWC) recognizes two stocks centered around Nova Scotia (NS), and Newfoundland-Labrador (NL) (Sergeant, 1977) (Fig. 1). Body size differences and differential rates of population decline during a short period of exploitation (1965–1972) off the Canadian Atlantic coast provided support for some degree of population structure between whales in NS and NL (Mitchell, 1974; Sergeant, 1977). Although concurrent tagging experiments indicated movements between NS and NL (representing ~10% of tag recaptures), these movements were not sufficient to offset the higher rate of decline around NS, thereby leading the IWC to consider two stocks (Mitchell, 1974; Sergeant, 1977). More recently, Breiwick (1993) reanalyzed catch and effort data from the same whaling period as well as biological parameters of the catch and confirmed Mitchell's hypothesis (1974); the current stock definition has not been reexamined since that time.

In this paper, we reassess the stock structure of Northwest Atlantic fin whales using the structure of their song with a special emphasis on the connection of Gulf of St. Lawrence (GSL) fin whales to those in adjoining areas, particularly in the Gulf of Maine (GoM). Previous studies comparing contaminant loads (Hobbs *et al.*, 2001) and residency and movement patterns (Coakes *et al.*, 2005; Robbins *et al.*, 2007) provided evidence that GSL fin whales are relatively isolated from whales in the GoM, NS, or NL, even though limited movements occur between these areas. On the other hand, Bérubé *et al.* (1998) found no evidence of genetic differentiation between the GSL and the GoM, which could be explained by gene flow, recent divergence, or interbreeding. These fin whales are thus best described as belonging to distinct feeding aggregations linked by limited genetic interchange. Here, we test if song structure can provide a way to identify these aggregations that are not populations as commonly defined, i.e., genetically isolated, but may be isolated enough to be considered separately for management purposes. The results of the song comparative analysis are compared to other stock assessment methods and discussed with

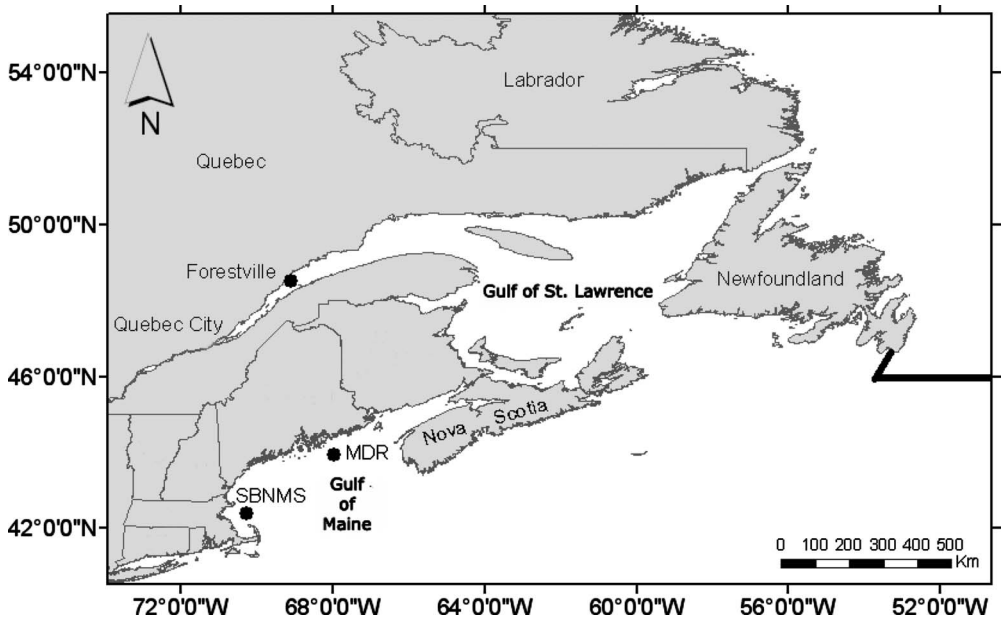


FIG. 1. Map of the study area showing the three recording locations. The solid line indicates the current IWC boundary between the NL and NS fin whale stocks.

respect to fin whale management. Finally the influence of male dispersal and vocal learning on geographic variations in song is also discussed.

II. METHODS

A. Recording equipment and schedule

Passive acoustic data were obtained using autonomous recording units (ARUs) developed by Cornell's Laboratory of Ornithology's Bioacoustics Research Program (http://www.birds.cornell.edu/brp/hardware/pop-ups). Each unit contains batteries, a hydrophone, and a computer hard drive encased inside a 17-in. pressurized sphere. Recordings were obtained at three sites (Fig. 1). In the GSL, recordings were obtained from September 18, 2005 to February 4, 2006 on a 1 h on/1 h off duty cycle near Forestville, QC (48°30′ N, 69°06′ W) in the St. Lawrence Estuary. In the GoM, recordings were obtained at two locations: near Mount Desert Rock (MDR) (44°04′ N, 67°55′ W) from June 28 to September 17, 2006 and from September 1 to October 1, 2007 and within the Stellwagen Bank National Marine Sanctuary (SBNMS; 42°15′ N, 70°15′ W) from January 1 to March 26, 2006, September 7 to November 23, 2006, and December 5, 2006 to February 18, 2007 (Table I). The ARUs were programed to record at a sampling rate of 2000 Hz for SBNMS and MDR and at 1000 Hz in the GSL. An effective frequency response range for each system can be calculated as half the value of the relevant sampling rate (that is, ⩽1000 Hz and ⩽500 Hz respectively).

B. Fin whale song and definition

A fin whale's song is a stereotyped repetitive sequence of LF downsweeping pulses approximately 1 s in duration and centered around 20 Hz (Watkins, 1981; Watkins *et al.*, 1987; Thompson *et al.*, 1992) (Fig. 2). Songs are produced by males (Watkins *et al.*, 2000; Croll *et al.*, 2002) and are primarily recorded from fall until spring with a peak in winter (Watkins *et al.*, 2000; Stafford *et al.*, 2007), which coincides with the prime breeding time between November and January (Kjeld *et al.*, 1992; Lockyer and Sigurjonsson, 1992). Although singing is believed to be part of breeding behavior, the exact function of songs remains unknown. Song structure is far less complex than in closely related baleen whale species. Sequences of pulses lasting approximately 3–20 min are separated by periods of silence—here called rests—often associated with surfacing episodes. Several sequences form a bout; bouts are typically isolated from other bouts by at least 2 h and last up to 32 h. Within a sequence, individual pulses are separated by stereotyped in-

TABLE I. Recording periods in the GSL and at two locations in the GoM: the SBNMS and MDR. Recordings were obtained in months and years indicated by an "X." Fin whale songs were detected in months and years indicated by a "Y."

	Recording schedule														
	2005			2006									2007		
Area	Oct.	Nov.	Dec.	Jan.	Feb.	Mar.	Jul.	Aug.	Sep.	Oct.	Nov.	Dec.	Jan.	Feb.	Sep.
GSL	XY	XY	XY	X	X										
SBNMS				XY	XY	XY			XY	XY	XY	XY	XY	X	
MDR							X	X	XY						XY

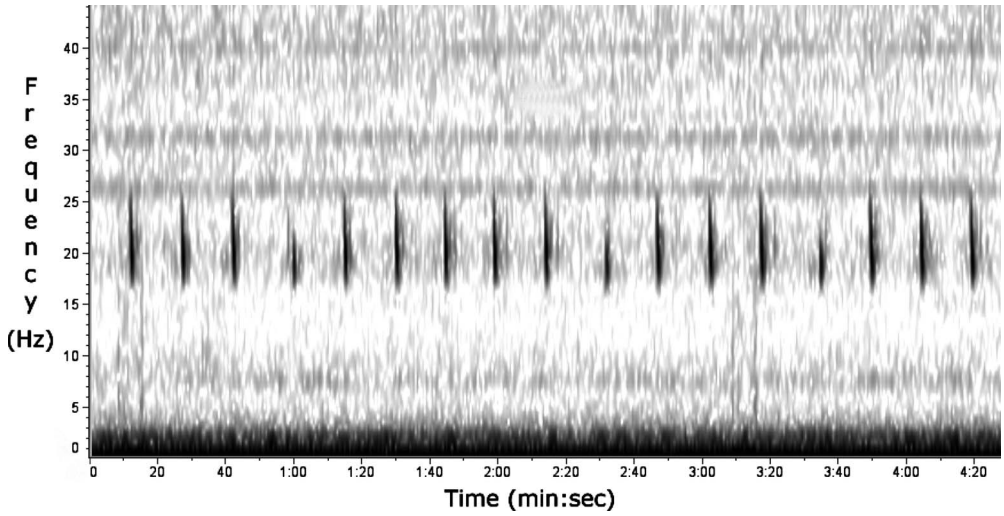


FIG. 2. Sequence of 20-Hz pulses. The spectrogram was generated using a Hanning window, 2048-point Fast Fourier Transform (FFT), 75% overlap, and a window length of 0.5.

tervals, called PIs, measured from the beginning of a pulse to the beginning of the next one (Watkins *et al.*, 1987). PIs show little intra- or interindividual variation within a geographical area. Differences in PI, and to a lesser extent median pulse frequency, have been found to explain up to 82% of the variation between songs from different areas and persist over periods of years with limited intra- and interannual variation (Hatch and Clark, 2004). Thus, PIs provide a good measure of acoustic divergence when comparing songs recorded in different areas.

C. Song selection

Passive acoustic data were visualized and analyzed using the eXtensiBle Acoustic Tool software (XBAT) (http://www.xbat.org). Spectrogram parameters were standardized to maintain constant time and frequency resolution for recordings sampled at different rates. An automated detector was built using the XBAT template detector tool to search for sequences of 20-Hz pulses in each of the three data sets. Given the large number of songs recorded in all areas, the following process was applied to select specific sequences for analysis. Only sequences that could be unambiguously attributed to a single singer and containing pulses with a high signal-to-noise ratio were considered. In order to lower the risk of nonindependence of data, each selected song had to be at least 24 h but preferably several days apart. Only songs containing a minimum of 100 usable pulses, which typically correspond to two to four sequences, were considered for analysis. From all the suitable songs, a random temporally stratified subset was selected from each area for analysis.

D. Song analysis

For each song, PIs were measured and collated to create a histogram using 0.5-s-wide frequency bins. PIs corresponding to, or occurring during, rests were not included. Histograms were classified as unimodal or bimodal based on whether they showed one or two peaks in PI (see Fig. 3). The mean PI was calculated for each song and each area. For songs characterized by a bimodal PI frequency distribution, a boundary PI separating both peaks was defined as the midpoint of the range of contiguous frequency bins containing less than 1% of all PIs. PIs shorter and longer than the boundary were averaged separately for a given singer and pooled with short and long PIs from other singers, respectively, in order to calculate the regional PI average.

PI frequency distributions were compared between regions using a Kolmogorov–Smirnov two sample test. We also investigated geographic and individual variation in song structure with classification trees using R (R Development Core Team 2004, http://www.R-project.org) and the RPART library (Therneau and Atkinson, 2004). To test for geographic variation, all PIs of each recording area were pooled; to test for individual variation, all PIs of each song were pooled. In both cases, data were split into increasingly homogeneous groups based on PIs, and the variation explained by the split was visualized graphically through the depth of the split; i.e., the longer the node, the more variation in the data explained by the corresponding split [for a detailed explanation of classification trees, see Risch *et al.* (2007)].

III. RESULTS

The results are based on the analysis of 19 songs from the GSL (2234 pulses), 10 songs from MDR (1591 pulses),

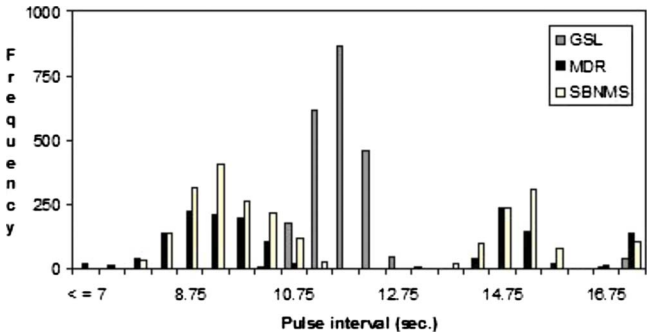


FIG. 3. (Color online) Overall PI frequency histogram for fin whale songs from the GSL (n=2234), MDR (n=1591), and SBNMS (n=2432). Frequency bins were 0.5 s in width. Numbers along the x-axis represent bins' midpoint.

TABLE II. Mean PI and standard deviation (SD) for song sequences recorded in the GSL and at MDR and SBNMS in the GoM.

Gulf of St. Lawrence					Gulf of Maine						
Date	Area	Mean	SD	n	Date	Area	Mean_1	SD_1	Mean_2	SD_2	n (n_1/n_2)
9/18/05	GSL	11.42	0.50	123	1/6/06	SBNMS	14.78	1.25	⋯	⋯	100
9/27/05	GSL	11.50	0.81	127	1/11/06	SBNMS	15.34	0.78	⋯	⋯	125
10/8/05	GSL	11.86	1.47	100	1/20/06	SBNMS	15.26	1.12	⋯	⋯	155
10/12/05	GSL	11.59	0.80	122	2/14/06	SBNMS	15.87	0.99	⋯	⋯	111
10/20/05	GSL	11.76	1.20	115	2/20/06	SBNMS	15.72	1.20	⋯	⋯	121
10/22/05	GSL	11.66	1.27	135	3/12/06	SBNMS	15.24	0.81	⋯	⋯	161
10/28/05	GSL	11.70	1.48	108	8/29/06	MDR	8.77	1.16	15.52	1.35	378 (175/203)
10/31/05	GSL	11.69	0.41	133	8/31/06	MDR	9.24	0.89	15.37	1.18	163 (80/83)
11/4/05	GSL	11.55	0.76	124	9/2/06	MDR	8.87	1.30	16.17	2.74	143 (82/61)
11/7/05	GSL	11.90	1.30	105	9/8/06	MDR	15.49	1.17	⋯	⋯	164
11/10/05	GSL	11.75	0.65	120	9/9/06	SBNMS	8.80	1.14	⋯	⋯	107
11/13/05	GSL	11.77	0.38	101	9/15/06	MDR	14.37	3.14	⋯	⋯	103
11/15/05	GSL	11.78	0.36	119	9/19/06	SBNMS	14.61	1.29	⋯	⋯	100
11/22/05	GSL	11.77	1.03	131	9/25/06	SBNMS	8.81	0.69	⋯	⋯	114
11/25/05	GSL	11.89	0.35	105	10/1/06	SBNMS	9.02	0.65	⋯	⋯	142
12/5/05	GSL	12.53	1.81	119	10/8/06	SBNMS	9.18	0.67	⋯	⋯	106
12/6/05	GSL	12.01	0.63	112	10/15/06	SBNMS	9.58	1.70	⋯	⋯	160
12/11/05	GSL	11.85	1.18	100	11/6/06	SBNMS	9.92	1.63	⋯	⋯	122
12/13/05	GSL	12.08	0.99	135	11/15/06	SBNMS	9.30	0.65	⋯	⋯	136
					11/23/06	SBNMS	9.62	1.42	⋯	⋯	118
					12/5/06	SBNMS	9.67	0.69	⋯	⋯	136
					12/13/06	SBNMS	9.93	1.15	⋯	⋯	117
					1/3/07	SBNMS	9.99	1.02	⋯	⋯	151
					1/24/07	SBNMS	9.88	1.17	⋯	⋯	152
					9/1/07	MDR	9.31	0.61	⋯	⋯	150
					9/7/07	MDR	9.57	1.98	⋯	⋯	128
					9/10/07	MDR	9.13	0.80	⋯	⋯	110
					9/14/07	MDR	9.40	0.97	⋯	⋯	150
					9/21/07	MDR	9.05	0.78	⋯	⋯	102

and 19 songs from SBNMS (2432 pulses). Songs in the GSL had a mean PI of 11.79 s (SE=0.02). In the GoM, the overall PI frequency distribution was bimodal. At MDR, PIs averaged 9.14 s (SE=0.04) and 15.36 s (SE=0.08). At SBNMS, PIs averaged 9.5 s (SE=0.03) and 15.28 s (SE=0.04) (Table II).

PIs in the GSL and GoM (Fig. 3) were significantly different (Kolmogorov–Smirnov two sample test statistics =0.589; $p<0.001$). This trend persisted when songs from prime breeding months only [November to January (Kjeld *et al.*, 1992; Lockyer and Sigurjonsson, 1992)] were considered: PIs from late fall 2005 in the GSL were significantly different from SBNMS PIs in winter 2006 (Kolmogorov–Smirnov two sample test statistics=0.973; $p<0.001$) and in fall 2006 to winter 2007 (Kolmogorov–Smirnov two sample test statistics=0.915; $p<0.001$).

The only noticeable overlap in frequency distribution occurred for PIs between 10.5 and 11 s and represented 8.0% of all GSL PIs and 5.1% and 1.3% of all PIs at SBNMS and MDR, respectively. Considering only PIs centered around the peak in short PI, the overlap represented 7.9% of SBNMS and 2.1% of MDR PIs. Although overall PI distributions at MDR and SBNMS were significantly different (Kolmogorov–Smirnov two sample test statistics=0.091; $p<0.001$), they followed the same pattern, and songs recorded in these areas exhibited a similar structure, as shown in Fig. 3.

The classification tree representing geographic variations (Fig. 4) revealed differences in the structure of songs recorded in the GSL and GoM. The overall classification error rate for that tree was 3.53%. The first split separated songs from SBNMS and MDR characterized by a short PI from the other songs. The second split separated GSL songs from SBNMS and MDR songs characterized by a long PI. For a given song type, SBNMS and MDR songs could not be differentiated. Consequently, both areas will be hereafter jointly referred to as GoM. The classification tree representing interindividual variations (Fig. 5) showed the same pattern, with an overall misclassification rate of 4.0%. Songs split into three groups ($\text{PI}>13.43$, $10.64\leqslant\text{PI}\leqslant13.43$, and $\text{PI}<10.64$) reflecting the three types of PI described above. Most of the variations in this tree were explained by the second split separating GSL and short PI GoM songs. Fin whales singing a particular song type could not be individually differentiated.

In the GoM, songs recorded during the winter (January to March) 2006 were different from those recorded in and after fall 2006, with an apparent transition period. Songs recorded in January through March 2006 at SBNMS were all characterized by a long PI. However, we found both the short and long song structure among songs recorded in August and September 2006 at SBNMS and MDR, including three songs incorporating short and long PIs grouped in different se-

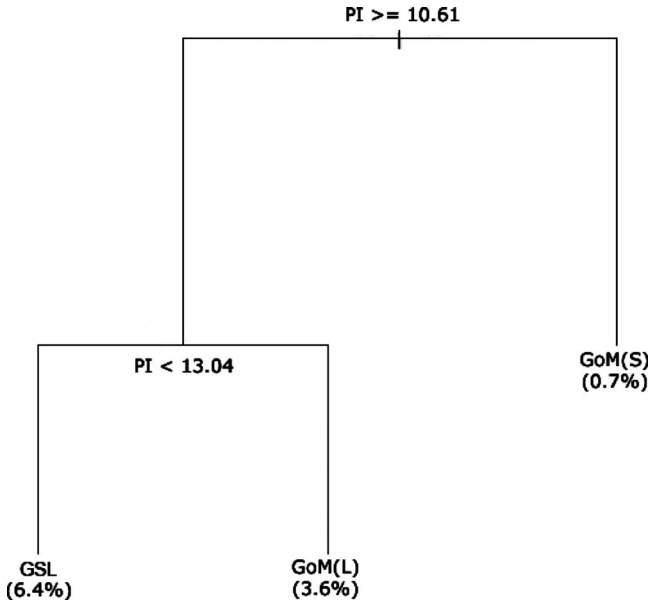


FIG. 4. Classification tree showing how songs recorded in the GSL and the GoM split based on PIs. GoM(L) and GoM(S) refer to long and short PI songs recorded in the GoM, respectively. The PI cutoff values are represented at each split. The numbers in parentheses indicate the misclassification rate for each node. The vertical depth of each split indicates the proportion of total variation in the data explained by it.

quences displayed consecutively. From October 2006 and until September 2007, all songs were characterized by a short PI at both locations (Fig. 6).

In the GSL, mean PI was positively correlated to the day of recording (Pearson $r=0.725$, $n=19$, $p<0.0005$). In the GoM, mean PIs recorded at SNNMS in winter 2006 were not correlated with the day of year (Pearson $r=0.511$, $n=6$, $p>0.05$) while fall 2006 to winter 2007 mean PIs were positively correlated with the day of year (Pearson $r=0.853$, $n=12$, $p<0.005$). MDR fall 2007 mean PIs were not correlated with recording day (Pearson $r=-0.515$, $n=5$, $p>0.05$).

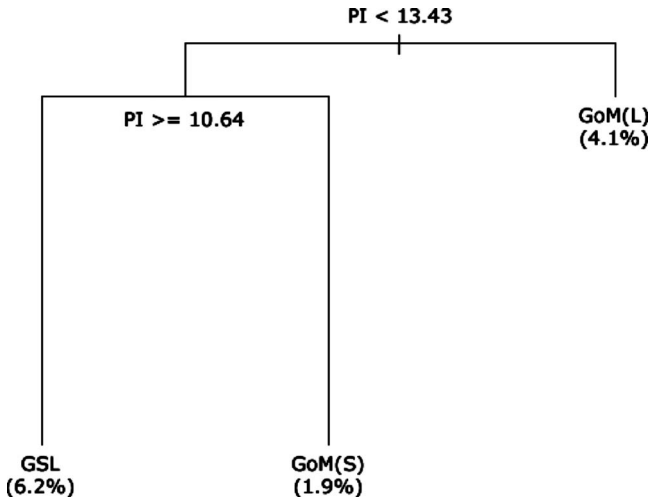


FIG. 5. Classification tree showing how singers from the GSL and GoM are split based on their song's PIs. GoM(L) and GoM(S) refer to long and short PI songs recorded in the GoM, respectively. The PI cutoff values are represented at each split. The numbers in parentheses indicate the misclassification rate for each node. See the legend in Fig. 4 for further details.

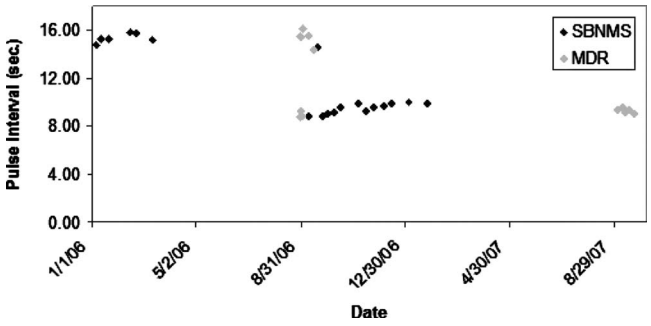


FIG. 6. Mean PI temporal distribution for songs recorded at MDR and SBNMS.

IV. DISCUSSION

A. Comparison of results with other stock structure assessment methods

Results of the PI analysis indicate that GSL songs differ from those recorded in the GoM. Fin whale stock structure in the Northwest Atlantic has previously been assessed using several recognized methods. First, several GSL fin whales were tagged during the 1965–1972 Eastern Canadian whaling period. Although no whaling occurred in the GSL, none were recaptured on the whaling grounds of NL or NS (Mitchell, 1974), which suggests nonhomogeneous mixing between these areas. In addition, population estimates for the GSL derived using line-transects during the same period remained relatively constant, while adjacent stocks under exploitation declined steadily. If GSL and NS or NL fin whales belong to the same stock and mix homogeneously and taking into account the small size of the GSL stock in relation to NS or NL, GSL whales should also have declined to some degree. Thus, GSL fin whales may be relatively isolated from NL or NS (Mitchell, 1974; Sergeant, 1977).

Second, levels of PCBs and persistent organochlorine pollutants in the blubber of fin whales caught in NS and NL during the 1965–1972 whaling period and sampled using biopsy darts in the GSL in the early 1990s (Gauthier *et al.*, 1997) suggested that GSL fin whales are not part of the NL stock and are relatively distinct from NS whales (Hobbs *et al.*, 2001).

Third, long-term photoidentification studies indicate that fin whales in the GoM and GSL exhibit strong site fidelity to their feeding grounds (Agler *et al.*, 1992; Robbins *et al.*, 2007; Mingan Island Cetacean Study, unpublished data; Edds and Macfarlane, 1987). In addition, a recent study noted that of 57 fin whales photoidentified off Halifax, NS in 1997, nine had also been seen in the GoM, four in the GSL, and one in both areas (Coakes *et al.*, 2005). Considering the number of individuals known in both areas (~840 in the GoM, ~430 in the GSL; Robbins *et al.*, 2007), the fact that only one individual was seen in both GSL and GoM indicates that very few movements occur between them, even though some mixing between GoM and GSL fin whales may occasionally happen within or outside these areas.

Residency and movement patterns (Agler *et al.*, 1992; Robbins *et al.*, 2007; Edds and MacFarlane, 1987) as well as differences in population trends (Mitchell, 1974) and contaminant loads (Hobbs *et al.*, 2001) suggest that GSL fin

whales are likely isolated from adjacent stocks and that fin whale feeding aggregations may qualify as demographically isolated units. This study showed that results provided by comparative song structure analyses are consistent with other stock structure assessment methods and may thus be helpful in identifying and tracking whales from different feeding aggregations. In addition, it showed that songs differ on a relatively small spatial scale. The stock structure currently adopted by the IWC in the Northwest Atlantic defines two broad stocks that are likely to encompass more than one feeding aggregation. Clapham *et al.* (2008) warned that current stock definition typically targets areas too large for management to be effective, and acoustically identifiable feeding aggregations may thus represent a more appropriate management unit in fin whales. Although fin whales from different feeding aggregations may be interbreeding, individuals' maternally directed fidelity (Clapham and Seipt, 1991) to their feeding ground also suggests that they should be managed at the level of feeding aggregations, as already recommended for humpback whales (Clapham *et al.*, 2008).

This study also confirmed recent findings indicating that estimates of genetic and acoustic divergence are not correlated (Hatch and Clark, 2004). Bérubé *et al.* (1998) did not detect any genetic differentiation between GSL and GoM fin whales. Evidence showing that fin whales in these areas may be at an early stage of divergence combined to the fact that samples used in these analyses were collected on two summer feeding grounds known to exchange a small number of individuals could explain this finding. As movements between areas are likely to occur broadly (Bérubé *et al.*, 2006), genetic stock structure assessments of feeding aggregations may fail to detect differences between areas with different conservation needs. Song structure analyses may provide a good alternative.

B. Temporal trends in song structure

PIs of songs from the Northwestern Atlantic were previously found to average around 9 and 14–15 s (see Hatch and Clark, 2004, Fig. 4). Both values, measured for songs recorded from January to March 1994–1995 and November to December 2002–2003, are in agreement with our recordings from 2006 and 2007 in the GoM. In addition, Hatch and Clark (2004) recorded both song types sequentially and not simultaneously, starting with songs characterized by the short PI. In this study, the reverse pattern was observed: the initial period was characterized by long PIs, followed by a short transition period during which both PIs were recorded while the final recordings contained only short PIs. The transition also appears to have taken place at the same time at MDR and SBNMS (Fig. 6), and one can speculate whether our recordings overlap with the transition period from one song type to the other. Regardless, it seems that song structure in the GoM varies in a nonrandom way over a relatively short time period. Therefore, at least on a decadal time scale, a fin whale song structure remains fairly constant and can be used to characterize geographic areas and groups of individuals living therein.

C. Causes of song variations

Hatch and Clark (2004) found that estimates of genetic divergence were not correlated with geographic variations in song structure. Consequently, males from different populations in acoustic or physical contact are more likely to sing different songs than allopatric, unrelated males. This finding could explain differences observed in this study. In winter, fin whales are forced out of the GSL by sea ice. After leaving the GSL, fin whales might potentially winter off NS waiting for the ice to retreat (Sergeant, 1977). There, they might come into contact with fin whales from adjacent stocks. Character displacement theory predicts that under conditions where reproductive signals could attract mates from different populations, selective forces should act on calls to become regionally distinctive (Gerhardt, 1994; Hatch and Clark, 2004). Recordings of songs in fall, winter, and spring in the GoM and other northern North Atlantic areas (Clark, 1995; Watkins *et al.*, 1987) as well as observations of singers displaying apparent courtship behavior in the GoM (Watkins *et al.*, 1987) suggest that breeding could take place, at least partly, on the feeding grounds. Therefore, it is reasonable to presume that the different songs recorded in the GoM and GSL may have originated from temporary range overlap of two or more stocks during the breeding period due to seasonal changes in distribution.

The hypothetical winter range overlap of two stocks in the GoM to explain the presence of two song types in this area by character displacement theory should correlate with simultaneous recording of their respective songs, but this was not the case in this study. The observed temporal segregation of song types cannot be explained by spatial segregation since both song types were recorded at both locations. GoM fin whales may in fact be characterized by two song types, as indicated by the three songs incorporating both PIs, whose structure alternates on a multiyear basis. Despite this shift, these songs appear to characterize the GoM feeding aggregation consistently and could also be used to track fin whales from this area.

D. Implications for acoustic stock structure assessment

The main purpose of this paper was to assess the use of fin whale songs for characterizing management stocks. Since some movements occur between feeding grounds (Coakes *et al.*, 2005), it is important to consider and differentiate the impacts that permanent dispersal and temporary movements may have on their acoustic signature. Dispersal, though hard to detect and quantify, is a critical factor in evaluating whether individuals from different areas are demographically isolated and should be managed together or separately. It is thus essential to determine whether male fin whale interstock movements can be detected acoustically.

Seasonal movements may result in temporary range overlap that could confound genetic stock structure assessments but should theoretically have no impact on the long-term composition of a stock. In fin whales, range overlap is likely to be caused by environmental factors leading to shifts in prey type or abundance (Payne *et al.*, 1990) and should be

occasional and most prominent during the feeding season. Temporarily sympatric fin whale stocks that intermingle to exploit productive areas but not to mate are likely to retain the original structure of their song. Considering that songs can be recorded as early as September when fin whales usually still frequent their feeding grounds, acoustic sampling of feeding aggregations may provide a better knowledge of the distribution and range overlap of adjacent stocks that have been characterized acoustically. It may also prove to be the best way to investigate whether individuals from different feeding stocks share the same area in winter and thus potentially interbreed.

Studies have pointed out the existence of a possibly important gene flow between fin whale populations (Hatch and Clark, 2004; Palsbøll *et al.*, 2004; Bérubé *et al.*, 2006), and there is genetic evidence suggesting that some males are successful postdispersal breeders (Hatch and Clark, 2004). Detecting these breeding-related, potentially permanent, movements acoustically may not be possible if males are capable of modifying their songs when exposed to new ones. As discussed above, these movements could explain the lack of concordance between geographic patterns of acoustic and genetic variation (Hatch and Clark, 2004). Studies on birds have revealed the same phenomenon in a number of species (Hafner and Petersen, 1985; Wright and Wilkinson, 2001), and postdispersal vocal learning in males was found to be critical in preventing populations from diverging genetically while maintaining local dialects (Ellers and Slabbekoorn, 2003; Wright *et al.*, 2005).

Whether fin whales are capable of vocal learning is unknown. However, if male fin whales can learn new songs after dispersing, they may adjust to the local song of their new environment. The selective pressure for song matching will depend on the role of singing in the mating system of fin whales and on whether singing the local song improves a male's reproductive fitness, as seen in birds (Payne, 1982). Regardless, vocal learning could prevent the detection of dispersal, but it would also allow immigrants to be correctly assigned to their new stock.

A potential drawback of using songs as a stock assessment tool is that singing is a male behavior, and song-based surveys would therefore not include females. However, if singing is indeed a mating display and given the geographic discreteness of various song types, one could expect males to sing in the presence of females, presumably of their own stock. Thus, acoustic sampling could potentially delineate stocks encompassing the range of both males and females. More research is needed to investigate whether fin whales are capable of vocal learning, the role of singing in the species' mating system, i.e., whether it functions as a display to attract females, and if females only respond to the song produced by males from the same stock before song-based stock structure assessments become fully functional.

E. Conclusion

This study showed that for the areas and periods sampled, geographic song variations are consistent with the stock structure model derived from differences in contaminants loads, demographic parameters, and residency patterns. Despite limited movements between areas, songs seem to be a stable regional indicator and could therefore offer an appropriate means to define management units, in particular at small spatial scales where genetic assessments may be affected by spatial overlap and gene flow. Because of the existing knowledge regarding the distribution of feeding aggregations, the Northwest Atlantic metapopulation presents an ideal opportunity to further test the applicability of acoustic stock structure assessment in fin whales. Studies to better quantify the rate of movements between the GSL, GoM, NS, and NL concurrently with thorough acoustic and genetic sampling could help resolve the uncertainty regarding the potential impact of movement patterns and vocal learning on geographic variations in song structure. It would also improve the current knowledge regarding the relationship between acoustic and genetic divergence.

ACKNOWLEDGMENTS

The authors would like to thank Leila Hatch as well as two anonymous reviewers for helpful comments on an earlier version of this paper. They are also thankful to the staff and volunteers at Allied Whale and Cornell Bioacoustics Research Laboratory and more particularly to Chris Tremblay, Dan DenDanto, and Toby Stephenson for their assistance in the deployment and retrieval of pop-up buoys at MDR. Funding for MDR data collection was provided by the Marisla Foundation. For the SBNMS data collection, David Wiley, Chris Clark, Leila Hatch, and Michael Thompson provided helpful support. Funding was provided by the Northeast Fisheries Science Center, the Northeast Regional Fisheries Office, and the SBNMS. Funding for the GSL data collection was provided by Forschungskommission der Universität Zürich, Zürcher Tierschutz, Schweizerische Akademie der Naturwissenschaften, and Basler Stiftung für biologische Forschung.

Agler, B. A., Robertson, K. A., DenDanto, D., Katona, S. K., Allen, J. M., Frohock, S. E., Seipt, I. E., and Bowman, R. S. (**1992**). "The use of photographic identification for studying fin whales (*Balaenoptera physalus*) in the Gulf of Maine," Rep. Int. Whal. Comm. **42**, 711–722.

Bérubé, M., Aguilar, A., Dendanto, D., Larsen, F., Notarbartolo Di Sciara, G., Sears, R., Sigurjonsson, J., Urban, J., and Palsbøll, P. J. (**1998**). "Population genetic structure of North Atlantic, Mediterranean Sea and Sea of Cortez fin whales, *Balaenoptera physalus*, (Linnaeus 1758): Analysis of mitochondrial and nuclear loci," Mol. Ecol. **7**, 585–599.

Bérubé, M., Danielsdóttir, A. K., Aguilar, A., Árnason, A., Bloch, D., Dendanto, D., Larsen, F., Lien, J., Notarbartolo-di-Sciara, G., Sears, R., Sigurjonsson, L., Urban, J., Witting, L., Øien, N., Vikingsson, G. A., and Palsbøll, P. J. (**2006**). "High rates of gene flow among geographic locations in North Atlantic fin whales," Paper presented to the IWC Scientific Committee, St. Kitts and Nevis, June, Paper No. SC/58/PF16, pp. 1–17.

Breiwick, J. M. (**1993**). "Population dynamics and analyses of the fisheries for fin whales (*Balaenoptera physalus*) in the northwest Atlantic Ocean," Ph.D. dissertation, University of Washington, Seattle, WA.

Clapham, P. J., Aguilar, A., and Hatch, L. T. (**2008**). "Determining spatial and temporal scales for management: Lessons from whaling," Marine Mammal Sci. **24**, 183–201.

Clapham, P. J., and Palsbøll, P. (**1999**). "Review of studies on stock identity of the humpback whale in the North Atlantic," Paper presented to the IWC Scientific Committee, Grenada, May, Paper No. SC/51/RMP22, pp. 1–7.

Clapham, P. J., and Seipt, I. E. (**1991**). "Resightings of independent fin whales, Balaenoptera physalus, on maternal summer ranges," J. Mammal. **72**, 788–790.

Clark, C. W. (**1995**). "Application of US Navy underwater hydrophone arrays for scientific research on whales," Rep. Int. Whal. Comm. **45**, 210–212.
Clark, C. W., Borsani, J. F., and Notarbartolo-di-Sciara, G. (**2002**). "Vocal activity of fin whales, *Balaenoptera physalus*, in the Ligurian Sea," Marine Mammal Sci. **18**, 281–285.
Coakes, A., Gowans, S., Simardi, P., Giard, J., Vashro, C., and Sears, R. (**2005**). "Photographic identification of fin whales (*Balaenoptera physalus*) off the Atlantic Coast of Nova Scotia, Canada," Marine Mammal Sci. **21**, 323–326.
Croll, D. A., Clark, C. W., Acevedo, A., Tershy, B., Flores, S., Gedamke, J., and Urban, J. (**2002**). "Only male fin whales sing loud songs," Nature (London) **417**, 809.
Dizon, A. E., Lockyer, C., Perrin, W. F., Demaster, D. P., and Sisson, J. (**1992**). "Rethinking the stock concept: A phylogeographic approach," Conserv. Biol. **6**, 24–36.
Donovan, G. P. (**1991**). "A review of IWC stock boundaries," Rep. Int. Whal. Comm. **13**, 39–68.
Edds, P. L. (**1988**). "Characteristics of finback *Balaenotera physalus* vocalizations in the St. Lawrence Estuary," Bioacoustics **1**, 131–149.
Edds, P. L., and Macfarlane, J. A. F. (**1987**). "Occurrence and general behavior of balaenopterid cetaceans summering in the St. Lawrence Estuary, Canada," Can. J. Zool. **65**, 1363–1376.
Ellers, J., and Slabbekoorn, H. (**2003**). "Song divergence and male dispersal among bird populations: A spatially explicit model testing the role of vocal learning," Anim. Behav. **65**, 671–681.
Ford, J. K. B., and Fisher, H. D. (**1982**). "Killer whale (*Orcinus orca*) dialects as an indicator of stocks in British Columbia," Rep. Int. Whal. Comm. **32**, 671–679.
Gauthier, J. M., Metcalfe, C. D., and Sears, R. (**1997**). "Chlorinated organic contaminants in blubber biopsies from northwestern Atlantic balaenopterid whales summering in the Gulf of St. Lawrence," Mar. Environ. Res. **44**, 201–223.
Gerhardt, H. C. (**1994**). "Reproductive character displacement of female mate choice in the grey treefrog, *Hyla chrysoscelis*," Anim. Behav. **47**, 959–969.
Gorbics, C. S., and Bodkin, J. L. (**2001**). "Stock structure of sea otters (*Enhydra lutris kenyoni*) in Alaska," Marine Mammal Sci. **17**, 632–647.
Hafner, D. J., and Petersen, K. E. (**1985**). "Song dialects and gene flow in the white-crowned sparrow, *Zonotrichia leucophrys nuttalli*," Evolution (Lawrence, Kans.) **39**, 687–694.
Harvey, P. H., and Pagel, M. D. (**1991**). *The Comparative Method in Evolutionary Biology* (Oxford University Press, Oxford, UK).
Hatch, L. T., and Clark, C. W. (**2004**). "Acoustic differentiation between fin whales in both the North Atlantic and North Pacific Oceans, and integration with genetic estimates of divergence," Paper presented to the IWC Scientific Committee, Sorrento, Italy, July, Paper No. SC/56/SD6, pp. 1–37.
Hobbs, K. E., Muir, D. C. G., and Mitchell, E. (**2001**). "Temporal and biogeographic comparisons of PCBs and persistent organochlorine pollutants in the blubber of fin whales from eastern Canada in 1971–1991," Environ. Pollut. **114**, 243–254.
Kellogg, R. (**1929**). "What is known of the migrations of some of the whalebone whales," Smithsonian Institution Annual Report **1929**, 467–494.
Kjeld, J. M., Sigurjonsson, J., and Arnason, A. (**1992**). "Sex hormone concentrations in blood serum from the North Atlantic fin whale (Balaenoptera physalus)," J. Endocrinol. **134**, 405–413.
Lockyer, C., and Sigurjonsson, J. (**1992**). "The Icelandic fin whale (Balaenoptera physalus): Biological parameters and their trends over time," Paper presented to the IWC Scientific Committee Special Meeting on North Atlantic Fin Whales, February 1991, Paper No. SC/F91/F8, pp. 1–79.
McDonald, M. A., Mesnick, S. L., and Hildebrand, J. A. (**2006**). "Biogeographic characterization of blue whale song worldwide: Using song to identify populations," J. Cetacean Res. Manage. **8**, 55–65.
Mellinger, D. K., and Barlow, J. (**2003**). "Future directions for acoustic marine mammal surveys: Stock assessment and habitat use," NOAA OAR, NOAA/PMEL Contribution 2557, pp. 1–37.
Mitchell, E. D. (**1974**). "Present status of northwest Atlantic fin and other whale stocks," in *The Whale Problem: A Status Report*, edited by W. E. Schevill (Harvard University Press, Cambridge, MA), pp. 108–169.
Mizroch, S. A., Rice, D. W., and Breiwick, J. M. (**1984**). "The Fin Whale, *Balaenoptera physalus*," Mar. Fish. Rev. **46**, 20–24.
Palsbøll, P. J., Bérubé, M., Aguilar, A., Notarbartolo-di-Sciara, G., and Nielsen, R. (**2004**). "Discerning between recurrent gene flow and recent divergence under a finite-site mutation model applied to North Atlantic and Mediterranean fin whales (*Balaenoptera physalus*) populations," Evolution (Lawrence, Kans.) **58**, 670–675.
Payne, P. M., Wiley, D. N., Young, S. B., Pittman, S., Clapham, P. J., and Jossi, J. W. (**1990**). "Recent fluctuations in the abundance of baleen whales in the southern Gulf of Maine in relation to changes in selected prey," Fish. Bull. **88**, 687–696.
Payne, R. B. (**1982**). "Ecological consequences of song matching: Breeding success and intraspecific song mimicry in indigo buntings," Ecology **63**, 401–411.
Payne, R. S., and Guinee, L. N. (**1983**). "Humpback whale (*Megaptera novaeangliae*) songs as an indicator of 'stocks'," in *Communication and Behavior of Whales*, edited by R. S. Payne (Westview, Boulder, CO), pp. 333–358.
Payne, R. S., and Webb, D. (**1971**). "Orientation by means of long range acoustic signaling in baleen whales," Ann. N.Y. Acad. Sci. **188**, 110–141.
Rendell, L. E., and Whitehead, H. (**2003**). "Vocal clans in sperm whales (*Physeter macrocephalus*)," Proc. R. Soc. London, Ser. B **270**, 225–231.
Risch, D., Clark, C. W., Corkeron, P. J., Elepfandt, A., Kovacs, K. M., Lydersen, C., Stirling, I., and Van Parijs, S. M. (**2007**). "Vocalizations of male bearded seals, *Erignathus barbatus*: Classification and geographical variation," Anim. Behav. **73**, 747–762.
Robbins, J., DenDanto, D., Giard, J., Panigada, S., Sears, R., and Zanardelli, M. (**2007**). "Photo-ID studies on North Atlantic fin whales," Paper presented to the IWC Scientific Committee, Anchorage, AK, June, Paper No. SC/59/PFI1, pp. 1–4.
Sergeant, D. E. (**1977**). "Stocks of fin whales *Balaenoptera physalus* L. in the North Atlantic Ocean," Rep. Int. Whal. Comm. **27**, 460–473.
Smith, T. D., Allen, J., Clapham, P. J., Hammond, P. S., Katona, S., Larsen, F., Lien, J., Mattila, D., Palsbøll, P. J., Sigurjónsson, J., Stevick, P. T., and Øien, N. (**1999**). "An ocean-basin-wide mark-recapture study of the North Atlantic humpback whale (*Megaptera novaeangliae*)," Marine Mammal Sci. **15**, 1–32.
Stafford, K. M., Mellinger, D. K., Moore, S. E., and Fox, C. G. (**2007**). "Seasonal variability and detection range modeling of baleen whale calls in the Gulf of Alaska, 1999–2002," J. Acoust. Soc. Am. **122**, 3378–3390.
Taylor, B. L. (**1999**). "Refining the definition of stocks: A discussion paper," Paper presented to the IWC Scientific Committee, Grenada, May, Paper No. IWC/51/RMP23, pp. 1–8.
Taylor, B. L. (**2005**). "Identifying units to conserve," in *Marine Mammal Research: Conservation Beyond Crisis*, edited by J. E. Reynolds III, W. F. Perrin, R. R. Reeves, S. Montgomery, and T. J. Ragen (The John Hopkins University Press, Baltimore, MD), pp. 146–164.
Therneau, T. M., and Atkinson, B. (**2004**). http://www.mayoresearch.mayo.edu/mayo/research/biostat/splusfunctions.cfm
Thompson, P. O., Findley, L. T., and Vidal, O. (**1992**). "20-Hz pulses and other vocalizations of fin whales, *Balaenoptera physalus*, in the Gulf of Mexico," J. Acoust. Soc. Am. **92**, 3051–3057.
Watkins, W. A. (**1981**). "Activities and underwater sounds of fin whales," Sci. Rep. Whales Res. Inst. **33**, 83–117.
Watkins, W. A., Daher, M. A., Reppucci, G. M., George, J. E., Martin, D. L., DiMarzio, N. A., and Gannon, D. P. (**2000**). "Seasonality and distribution of whale calls in the North Pacific," Oceanogr. **13**, 62–67.
Watkins, W. A., Tyack, P., Moore, K. E., and Bird, J. E. (**1987**). "The 20-Hz signals of finback whales (*Balaenoptera physalus*)," J. Acoust. Soc. Am. **82**, 1901–1912.
Winn, H. E., Thompson, T. J., Cummings, W. C., Hain, J., Hudnall, H., Hays, H., and Steiner, W. W. (**1981**). "Song of the humpback whale—population comparisons," Behav. Ecol. Sociobiol. **8**, 41–46.
Wright, T. F., Rodriguez, A. M., and Fleischer, R. C. (**2005**). "Vocal dialects, sex-biased dispersal, and microsatellite population structure in the parrot *Amazona auropalliata*," Mol. Ecol. **14**, 1197–1205.
Wright, T. F., and Wilkinson, G. S. (**2001**). "Population genetic structure and vocal dialects in an amazon parrot," Proc. R. Soc. London, Ser. B **268**, 609–616.

Feeding at a high pitch: Source parameters of narrow band, high-frequency clicks from echolocating off-shore hourglass dolphins and coastal Hector's dolphins

Line A. Kyhn[a)]
Department of Arctic Environment, National Environmental Research Institute, University of Aarhus, Frederiksborgvej 399, DK-4000 Roskilde, Denmark and Department of Biological Sciences, Aarhus University, C. F. Møllers Allé Building 1131, DK-8000 Aarhus, Denmark

J. Tougaard
Department of Arctic Environment, National Environmental Research Institute, University of Aarhus, Frederiksborgvej 399, DK-4000 Roskilde, Denmark

F. Jensen
Department of Biological Sciences, Aarhus University, C. F. Møllers Allé Building 1131, DK-8000 Aarhus, Denmark

M. Wahlberg
Fjord & Baelt, Margrethes Plads 1, DK-5600 Kerteminde, Denmark

G. Stone and A. Yoshinaga
England Aquarium Central Wharf, Boston, Massachusetts 02110

K. Beedholm
Department of Biological Sciences, Aarhus University, C. F. Møllers Allé Building 1131, DK-8000 Aarhus, Denmark

P. T. Madsen
Department of Biological Sciences, Aarhus University, C. F. Møllers Allé Building 1131, DK-8000 Aarhus, Denmark and Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 05432

(Received 26 September 2008; revised 22 December 2008; accepted 29 December 2008)

Toothed whales depend on echolocation for orientation and prey localization, and source parameters of echolocation clicks from free-ranging animals therefore convey valuable information about the acoustic physiology and behavioral ecology of the recorded species. Recordings of wild hourglass (*Lagenorhynchus cruciger*) and Hector's dolphins (*Cephalorhynchus hectori*) were made in the Drake Passage (between Tierra del Fuego and the Antarctic Peninsular) and Banks Peninsular (Akaroa Harbour, New Zealand) with a four element hydrophone array. Analysis of source parameters shows that both species produce narrow band high-frequency (NBHF) echolocation clicks. Coastal Hector's dolphins produce clicks with a mean peak frequency of 129 kHz, 3 dB bandwidth of 20 kHz, 57 μs, 10 dB duration, and mean apparent source level (ASL) of 177 dB re 1 $\mu Pa_{p.-p.}$. The oceanic hourglass dolphins produce clicks with mean peak frequency of 126 kHz, 3 dB bandwidth of 8 kHz, 116 μs, 10 dB duration, and a mean estimated ASL of 197 dB re 1 $\mu Pa_{p.-p.}$. Thus, hourglass dolphins apparently produce clicks of higher source level, which should allow them to detect prey at more than twice the distance compared to Hector's dolphins. The observed source parameter differences within these two NBHF species may be an adaptation to a coastal cluttered environment versus a deep water, pelagic habitat.
 [DOI: 10.1121/1.3075600]

PACS number(s): 43.80.Ka, 43.80.Ev [WWA] Pages: 1783–1791

I. INTRODUCTION

Toothed whales echolocate to navigate and find prey by processing echoes generated from emission of ultrasonic short clicks of high directionality and source level (Au, 1993). The performance of a toothed whale biosonar system depends on the source parameters of the transmitted echolocation clicks, and analysis of click properties can thus convey valuable information about the acoustic physiology and behavioral ecology of recorded toothed whale species (e.g., Au, 1993; Madsen and Wahlberg, 2007). The variation in habitat and prey type from ice filled fjords, muddy rivers, and deep open oceans provides diverse acoustic environments that may have contributed to the evolutionary shaping of different click types, but little is known about how click

a)Electronic mail: lky@dmu.dk

source parameters may be linked to the physical environment of the different echolocating toothed whale species (Wood and Evans, 1980).

A biosonar is ultimately limited in detection range by either noise or clutter (Urich, 1983). By clutter is meant echoes from irrelevant objects ensonified by the sonar that masks target echo reception when there is temporal overlap in the arrival of target and clutter echoes (Au and Benoit-Bird, 2008). Increasing the source level of the sonar potentially leads to significant ensonification of more unwanted objects in the water meaning that the received echo-to-clutter ratio remains constant or even deteriorates with increasing source level. Instead an improved echo-to-clutter ratio can be achieved by increasing the directionality of the sonar signal as objects outside the sound beam do not add significantly to the clutter, but this gain is at the expense of search width of the sonar beam.

A noise-limited biosonar, on the other hand, is limited by either the ambient noise or the self-noise of the auditory system. For most young and healthy toothed whales (Kastelein *et al.*, 1999; Johnson, 1967) the hearing threshold seems to be close to or below normal ambient noise level in the relevant frequency band, if the current interpretations of how toothed whales integrate noise are correct. Accordingly, an increase in source level increases the echo-to-noise ratio and hence the performance of the biosonar. The echo-to-noise ratio can also be raised by increasing the receiving directionality, which serves to reduce reception of anisotropic noise. Finally, echolocation signals may be shifted to higher frequencies where the ambient noise levels are lower (Møhl and Andersen, 1973; Urich, 1983; Au, 1993) but at the cost of increased sound absorption (Urich, 1983). Thus, different toothed whale species likely have optimized their echolocation capabilities to the specific habitat they have evolved in as seen for bats (Neuweiler and Fenton, 1988), as a trade-off between click source parameters matched to prey properties, sound absorption, ambient noise levels, clutter, and the functional constraints imposed by the morphology and size of their sound generators (Madsen and Wahlberg, 2007; Madsen *et al.*, 2005).

For example, the large sperm whale operates a powerful long range biosonar system to locate mesopelagic prey patches at long ranges by using clicks with high source levels and centroid frequencies between 15 and 20 kHz where absorption is low (Madsen *et al.*, 2002; Møhl *et al.*, 2000, 2003). At the other extreme, several smaller toothed whales have been shown to produce narrow band high-frequency (NBHF) clicks around 130 kHz (Møhl and Andersen, 1973; Dawson, 1988; Madsen *et al.*, 2005; Li *et al.*, 2007). These signals suffer from range-dependent absorption about 40 times higher than sperm whale clicks, so the sonar can only operate at short ranges. The species producing NBHF clicks are found in three different odontocete families and they all produce echolocation clicks at peak frequencies of more than 120 kHz, 3 dB bandwidth of 6–26 kHz and Q-values between 8 and 20 [e.g., *Phocoena phocoena*, *Neophocaena phocaenoides*, *Kogia breviceps*, and *Cephalorhynchus hectori* (Møhl and Andersen, 1973; Au, 1993; Dawson, 1988; Madsen *et al.*, 2005; Li *et al.*, 2007)].

The NBHF clicks are apparently very similar, yet it is not clear what factors may have driven the seemingly convergent evolution of NBHF clicks in species that live in very different habitats. Morisaka and Connor (2007) suggested that the NBHF signal evolved for acoustic camouflage as an anti-predator strategy against killer whales (*Orcinus orca*) whose hearing sensitivity decreases sharply at frequencies above 60 kHz and is practically zero above 100 kHz (Szymanski *et al.*, 1999). For this anti-predator strategy to be effective all energies of the NBHF signal must be emitted at frequencies over 100 kHz, above the upper hearing limit of the killer whale. Since absorption at the same time increases considerable with frequency above 100 kHz, the NBHF species have a small frequency band at their disposal to adapt to their environment. Yet, it seems that the very different habitats from deep water of *Kogia* to the coastal environment of porpoises would face these animals with different echolocation tasks in terms of prey, predation, noise, and clutter levels. Examples of such NBHF species living in different habitats are Hector's dolphin (*Cephalorhynchus hectori*) and the hourglass dolphin (*Lagenorhynchus cruciger*).

Hector's dolphins are coastal and have a body length of about 1.5 m and weigh around 50 kg (Reeves *et al.*, 2002). They are only found in New Zealand and are most often found within 1 km from land (Slooten *et al.*, 1993; Bräger *et al.*, 2002). They feed opportunistically on smaller fish and squid caught at the bottom and at the surface (Slooten and Dawson, 1988). The hourglass dolphins are oceanic and slightly larger than Hector's dolphins, measuring ~1.4–1.9 m and weighing 74–88 kg (Godall *et al.*, 1997). Molecular phylogenies (May-Collado and Agnarsson, 2006) suggest close taxonomic affinity to the *Cephalorhynchus* genus. With its oceanic circumpolar sub-Antarctic distribution knowledge about this species is primarily circumstantial (Godall *et al.*, 1997). The few collected stomachs of hourglass dolphins contained remains of small fish (*Mycophidae*) and squid (*Onychoteuthidae* and *Enoloteuthidae* families) (Godall *et al.*, 1997). So, while NBHF species seemingly produce almost identical clicks, it is possible from their differences in habitat and morphology that they do display differences in source parameters within the NBHF click class, in particular, with respect to source level.

Here we show that echolocation clicks of Hector's and hourglass dolphins have different source levels, duration, and bandwidth, possibly due to the different acoustic conditions posed by their respective habitats.

II. MATERIALS AND METHODS

A. Recording chain and field sites

Recordings were made with a linear, vertical array of four Reson TC 4034 spherical hydrophones (Reson A/S, Slangerup, Denmark) with 20 m cable and a measured sensitivity of −222 dB re 1 V/μPa between 100 and 150 kHz. Hydrophones were mounted horizontally in the same direction along a vertical Perspex rod with 1 m hydrophone spacing. The entire array was suspended either 1 m (Hector's) or 2 m (hourglass) below a buoy in the surface and with a 0.5 kg lead weight attached to the other end of the array (*sensu*

Madsen *et al.*, 2004a). Signals were bandpass filtered [100 Hz (one pole) to 200 kHz (four poles)], amplified in a custom-built four-channel amplifier, and digitized [500 kHz, 12 bit NuDAQ pci9812, AdLink, Los Angeles, CA]. The measured frequency response of the entire recording chain was flat (±2 dB) from 200 Hz to 180 kHz and allowed for continuous streaming of data to disk. Clip level of the recording chain was 189 dB re μPa (peak) with 50 dB gain for hourglass dolphins, and either 169 or 189 dB re μPa (peak) for Hector's dolphins (70 or 50 dB gain) set by the max input voltage of ±5 V peak in the analog-to-digital converter.

Recordings were obtained at two field sites. Hector's dolphins were recorded in the coastal habitat around Akaroa Harbour, New Zealand (43°52′9″S; 172°56′16″E) on January 7 and 9, 2007. When dolphins approached the large rigid-hulled inflatable boat (RHIB) to bow ride, the engine was turned off and the recording array lowered in the water. No other marine mammals were in sight or detected acoustically. Recordings were made under very calm weather conditions (low winds, sea state 0–1 Beaufort, sea temperature of 14 °C, and salinity of 34.5‰).

Hourglass dolphins were recorded on January 30, 2007, in the Drake Strait (58°17′50″S; 61°29′39″W) at open sea between Tierra del Fuego and the Antarctic Peninsular. A group of about 20 hourglass dolphins was sighted from the naval vessel HDMS "Vædderen" and identified by their characteristic hourglass-shaped white markings on the flanks and prominent dorsal fin. No other marine mammals were in sight or detected acoustically. The dolphins were approached in a RHIB and when animals were encountered within 100 m of the RHIB the recording array was deployed. Recordings were made under calm weather conditions (low winds, sea state 2 Beaufort, moderate swell, sea temperature of 5.1 °C, and salinity of 33.9‰).

B. Click analysis

Dolphin clicks are very directional and it is thus essential to ensure that clicks used for analysis are recorded as close to on-axis as possible (Au, 1993; Madsen and Wahlberg, 2007). Off-axis clicks are distorted and with much lower apparent source level (ASL) (Au *et al.*, 1986), and the click parameters may be different than on-axis (Au, 1993). For click analysis we applied a set of criteria for on-axis properties following Villadsgaard *et al.* (2007) to ensure that only clicks recorded close to on-axis were used. However, since we did not know whether focal animals were actually pointing their acoustic axis at the array, we used the definition of Møhl *et al.* (2000) of ASL, i.e., the back calculated sound intensity at a distance of 1 m from a directional source recorded in an unknown aspect, for the clicks complying with the following on-axis criteria: (i) recorded on all four channels; (ii) part of a scan, i.e., a series of clicks closely spaced in time normally first increasing then decreasing in amplitude (*sensu* Møhl *et al.*, 2003); (iii) of maximum amplitude on one of the two middle hydrophone channels; and (iv) the direct path of the click had to be stronger than any trailing bottom or surface reflections. Further, we determined that the localization error (see below) could not give rise to a transmission error of more than 2 dB in the source level calculations. Additionally, we visually inspected all on-axis clicks and removed clicks that contained double or triple pulses, since these likely arise from surface reflections and are not source generated (Li *et al.*, 2005).

The source properties were quantified using a series of parameters *sensu* Au (1993) and Madsen and Wahlberg (2007) for each click accepted as on-axis: Duration (defined as between −10 dB points on the envelope, calculated as the absolute value of the analytical waveform. From here on stated as 10 dB duration); peak frequency, centroid frequency (i.e., the frequency dividing the spectrum in two halves of equal energy on a linear scale), 10 dB bandwidth, (i.e., bandwidth at −10 dB points below the spectrum peak), 3 dB bandwidth (i.e., bandwidth at −3 dB points below the spectrum peak), rms bandwidth (i.e., spectral standard deviation around the centroid frequency on a linear scale), and Q-value (centroid frequency divided by the rms bandwidth).

Interclick intervals (ICIs) were found as the interval between the on-axis click and the click preceding the on-axis click in the same click scan. ICI is given in milliseconds. A few clicks were too closely spaced with echoes and clicks from several scans to objectively derive the ICI and these clicks were thus not included in the ICI analysis

The recordings were browsed using ADOBE AUDITION 1.5 (Syntrillium, Adobe, Mountain View, CA) and all analysis and signal processing were performed with custom written scripts in MATLAB (Mathworks).

C. Calibration of localization routines

The array performance was evaluated in Aarhus Harbour, Denmark, by playing out calibrated tone pips with source parameters similar to NBHF clicks at known ranges from the array. The four-hydrophone-array was suspended from a buoy with the top hydrophone 1 m below the surface. The sound source (at 3 m depth) was moved gradually away from the array at measured ranges in 10 m steps. Tone pips (130 kHz sinus pulses of 15 cycles and a 100 μs duration) were transmitted with an omni-directional hydrophone (B&K8105) connected to a sound generator (Agilent, model 33220A). The same recording chain and settings as used for the dolphin recordings were used to record signals transmitted in a range interval from 10 to 80 m from the array. Speed of sound was estimated from the Leroy equation (Urich, 1983) from measured temperature and salinity. A precise localization range was defined as a range within which the rms error (Villadsgaard *et al.*, 2007) with respect to the actual range was within a range jitter corresponding to a variation in transmission loss (TL) (spherical spreading) of <2 dB.

D. Estimation of source level

Synchronized recordings of the same click on four channels allow localization of the clicking dolphin with three hyperbolas calculated from time of arrival differences of the click pair wise among the four hydrophones. Localizations were performed using MATLAB implementing the localization routines of Wahlberg *et al.* (2001) and Madsen and Wahlberg (2007). It proved essential to use a robust measure of the

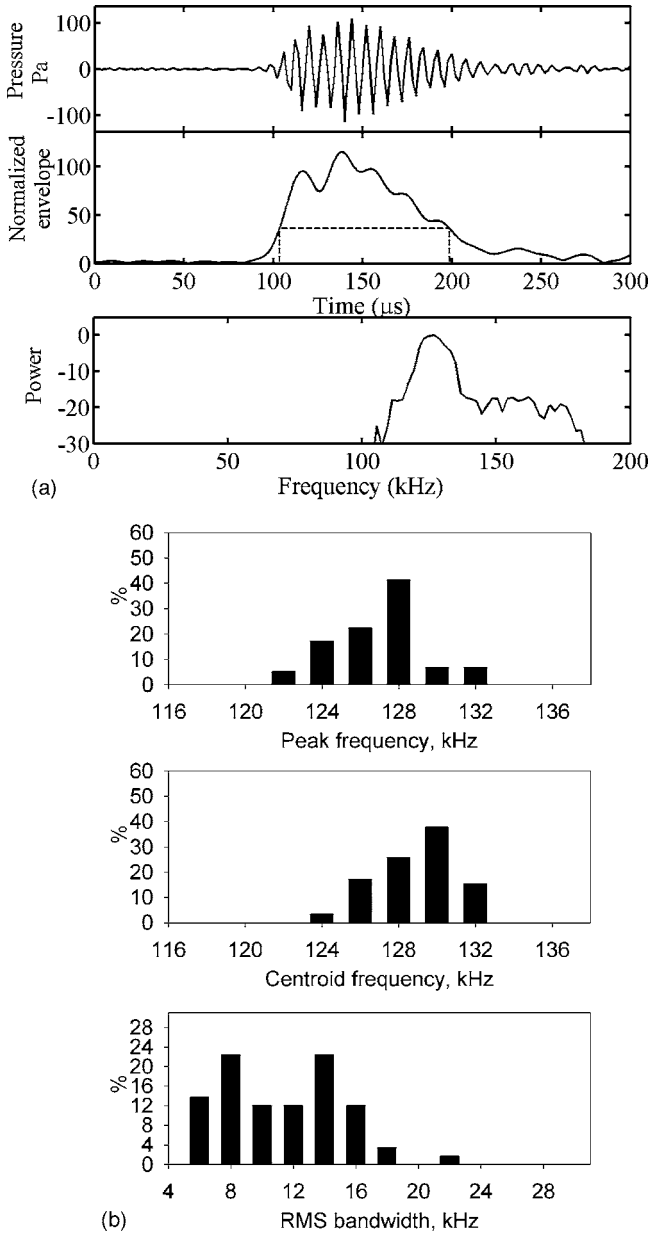


FIG. 1. (a) Time domain, envelope, and power spectrum of a representative hourglass dolphin signal. The dashed square in the envelope denotes the 10 dB duration. [Fast Fourier transform (FFT) size of 256, spectrum interpolated with a factor 100, sampling rate of 500 kHz, and rectangular window.] (b) Histograms of 58 on-axis hourglass dolphins' clicks with peak frequency, centroid frequency, and rms bandwidth. Binwidth is 2 kHz.

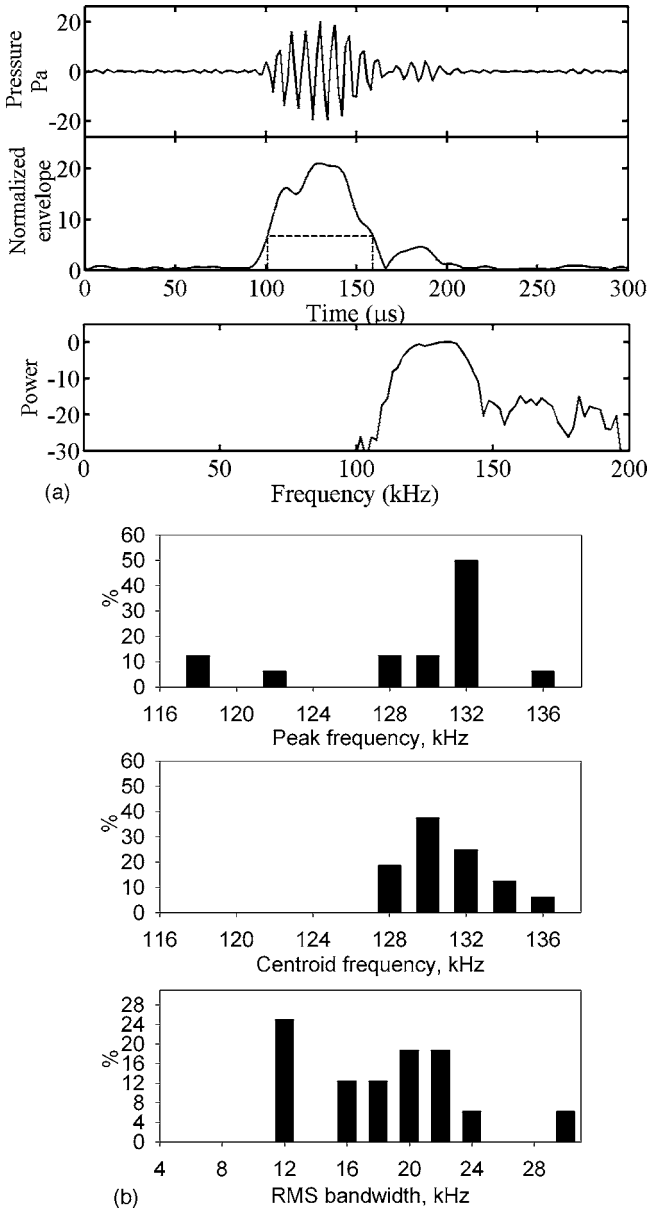


FIG. 2. (a) Time domain, envelope, and power spectrum of a representative Hector's dolphin signal. The dashed square in the envelope denotes the 10 dB duration. (FFT size of 256, spectrum interpolated with a factor 100, sampling rate of 500 kHz, and rectangular window.) (b) Histograms of 16 on-axis Hector's dolphins' clicks with peak frequency, centroid frequency, and rms bandwidth. Binwidth is 2 kHz.

time of arrival differences for the same signal recorded on the four channels. We determined the timing of a click from the first sample exceeding −10 dB of the peak of the click envelope [Figs. 1(a) and 2(a)].

Once the range to the animal has been estimated, TL can be calculated and added to the received level (RL) of a click. Villadsgaard *et al.* (2007) found that propagation loss of 130 kHz porpoise clicks in a shallow water habitat was well approximated by spherical spreading plus the frequency dependent absorption. Given the short distances and good mixing of the water column in the two habitats in the present study, we also assumed that TL could be approximated by spherical spreading plus absorption. ASL of echolocation clicks was thus calculated using the equation

$$\mathrm{ASL} = \mathrm{RL} + \mathrm{TL}(=20\log r + \alpha r) \quad \text{(Urich, 1983)},$$

where α is the absorption coefficient in dB/m and r is range in m. For the present field sites of 14 and 5.1 °C for Akaroa and the Drake Passage, α was calculated following expressions from Fisher and Simmons (1977) using the centroid frequency of the clicks (α is 0.037 for Hector's dolphin and 0.029 or hourglass dolphin). ASLs are given as peak-peak pressure, rms pressure, and energy flux density (EFD) and were computed as follows. $\mathrm{RL}_{\mathrm{p.-p.}}$ (dB re 1 $\mu\mathrm{Pa}_{\mathrm{p.-p.}}$) was measured directly from the maximum and minimum peak pressures of the waveform. $\mathrm{RL}_{\mathrm{rms}}$ (dB//1 μPa rms) is the rms pressure calculated over the 10 dB duration of the signal. $\mathrm{RL}_{\mathrm{EFD}}$ (dB//1 $\mu\mathrm{Pa}^2$ s) is the signal energy integrated over the 10 dB duration (Madsen, 2005).

TABLE I. Mean (±standard deviation) and range of echolocation click source parameters of hourglass dolphins (*Lagenorhynchus cruciger*) and Hector's dolphins (*Cephalorhynchus hectori*).

	Hector's dolphin *Cephalorhynchus hectori*		Hourglass dolphin *Lagenorhynchus cruciger*	
Parameters	Mean values (stdev)	Range	Mean values (stdev)	Range
10 dB duration (μs)	57(±6)	41–65	115(±24)	79–176
$RL_{p.-p.}$, dB re 1 $\mu Pa_{p.-p.}$	156(±6)	145–166	162(±4)	155–168
$ASL_{p.-p.}$, dB re 1 $\mu Pa_{p.-p.}$ [a]	177(±6)	161–187	197(±4)[a]	190–203[a]
$RL_{-10\ dB}$, dB re 1 μPa rms	145(±6)	133–154	151(±4)	144–158
$ASL_{-10\ dB}$, dB re 1 μPa rms[a]	166(±6)	152–175	186(±4)[a]	179–193[a]
$EFD_{-10\ dB}$, dB re 1 μPa^2 s [a]	121(±4)	110–126	146(±3)[a]	140–152[a]
Peak frequency (kHz)	129(±5)	117–135	126(±2)	122–131
Centroid frequency (kHz)	128(±3)	125–132	128(±2)	124–132
3 dB bandwidth (kHz)	20(±3)	12–26	8(±2)	5–11
10 dB bandwidth (kHz)	30(±10)	24–66	13(±2)	9–18
rms bandwidth (kHz)	18(±5)	11–29	11(±4)	5–22
$Q_{-3\ dB}$	6(±1)	5–11	17(±4)	12–25
Q_{rms}	8(±2)	4–12	13(±5)	6–26
Mean range from array (m)	11(±4)	4–19	50[b]	50[b]
n	16		58	

[a]ASL for hourglass dolphins is calculated from an estimated range.
[b]Estimated minimum range of 50 m, corresponding to a transmission loss of 35 dB.

III. RESULTS

Several hundred Hector's dolphin clicks were recorded over 2 days from 12 groups of between two and eight animals approaching to within 1 m from the hydrophone array. Due to the fission-fusion structure of Hector's dolphins groups (Slooten and Dawson, 1988; Slooten *et al.*, 1993) and their small home range (Bräger *et al.*, 2002) some animals have likely been recorded more than once. The dolphins approached the boat to bow ride.

Around 200 hourglass dolphin clicks were recorded on one occasion from a group of around 20 animals. The clustering of the dolphins and lack of directional movement suggested that they were engaged in feeding activities. Fifty-eight hourglass dolphins' clicks were judged to have been recorded on-axis; however, all were recorded at too great a range (>40 m) from the hydrophone array to allow for accurate localization. We thus based ASL calculations on estimated range. In order to ensure a conservative ASL estimate we used a range of 50 m to calculate TL (see equation above) between the clicking dolphin and the hydrophone array. A TL of 35 dB was therefore added to each RL for this species. No dolphins were recorded closer to the array than 40 m, and we therefore assume that the derived source levels are minimal estimates with the possibility of the estimated source levels to be higher, but very unlikely to be lower.

Source signal parameters of both species are summarized in Table I. Hourglass dolphins and Hector's dolphins both produced clicks with a centroid frequency of 128 kHz with little intraspecific variation and identical maximum values of 132 kHz. Bandwidths were much narrower for hourglass dolphins meaning that the click energy was contained in a smaller frequency band giving a higher *Q*-value. The RLs of hourglass dolphins ranged from 155 to 168 dB re 1 $\mu Pa_{p.-p.}$ and Hector's dolphins RL ranged from 145 to 166 dB re 1 $\mu Pa_{p.-p.}$. The peak-to-peak source level of hourglass dolphin was thus estimated to be about 20 dB higher than Hector's dolphin clicks, since they were at least 50 m away whereas Hector's dolphins were localized to be <20 m from the array. Hourglass dolphin clicks were on average more than twice as long as Hector's dolphin clicks and appeared to have an EFD at least 25 dB higher due to the higher source level and longer duration. Representative clicks and histograms of click parameters are shown in Figs. 1(a), 1(b), 2(a), and 2(b). We found no correlation for click duration or bandwidth with ICI, for either species (Fig. 5).

For the array calibration 2451 clicks were included in the analysis of the localization routines. 130 kHz clicks could be localized precisely (with a rms error up to 2 dB) out to 40 m from the hydrophone array (Fig. 3). At greater distances the routines underestimated the actual distances to the sound source, giving rise to errors larger than 3 dB in TL.

IV. DISCUSSION

NBHF clicks of porpoises and the *Cephalorhynchus* genus have been described as stereotypical (Au, 1993; Madsen *et al.*, 2005). The source parameters of Hector's and hourglass dolphins recorded in this study, however, displayed some apparent differences. Both species produced NBHF clicks with a centroid frequency of 128 kHz, but clicks of hourglass dolphins were about twice as long, with a narrower bandwidth and therefore higher *Q*-value, than those of Hector's dolphins. Hourglass dolphins consequently concentrated their energy in a narrower frequency band while the peak frequency and centroid frequency were similar for the two species.

Hourglass dolphins had higher estimated source levels than Hector's dolphin, and the fact that even the highest of the Hector's source levels found in this study did not exceed the lowest of the estimated hourglass dolphin source levels suggests that there is a genuine source level difference be-

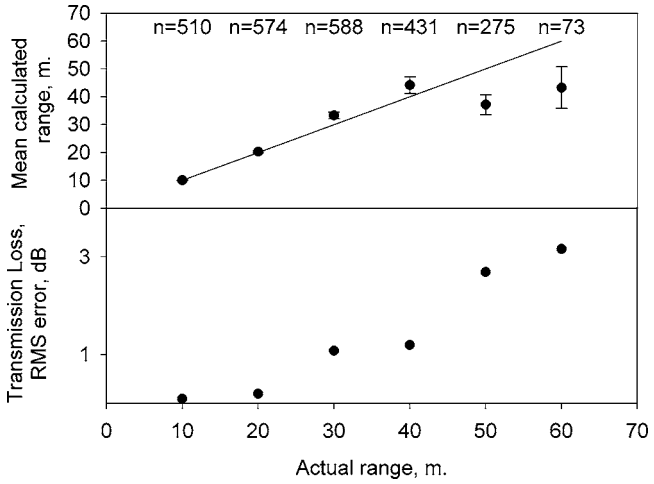


FIG. 3. Calibration of localization precision using a 3 m aperture four element hydrophone array. (a) Mean calculated ranges with standard deviation plotted against actual range. The line denotes the expected localization range from each of the actual ranges. (b) Effect of localization error on TL expressed in rms error, dB. With a 3 m aperture array NBHF species may be localized out to 40 m with ±2 dB precision on source level calculations.

tween the two species. Despite the apparent on-axis approach used in this study we can, however, only ensure that the clicks used for analysis were those recorded closest to on-axis of the clicks in a given scan. Since the ASL decreases with increasing angle to the acoustic axis (Au, 1993), source levels may be underestimated in our analysis compared to true on-axis clicks. Furthermore, the ASL of hourglass dolphins is a rough (but most likely conservative) estimate since the dolphins were too far away to be localized. The estimated range of 50 m for the calculation of TL was based on a visual estimation made at the time of recording and on the fact that 130 kHz clicks can be located accurately out to 40 m from the array. We could not see the animals under water, but the fact that we could not locate any of the recorded clicks leads us to suspect that all dolphins were at greater ranges than 40 m from the array during recording. Further, the hourglass dolphins had a mean RL of 162 dB re 1 $\mu Pa_{p.-p.}$ and was ~50 m away, whereas Hector's dolphins were <20 m away and had a mean RL of 156 dB re 1 $\mu Pa_{p.-p.}$. A mean ASL of 197±4 dB re 1 $\mu Pa_{p.-p.}$ at 1 m and unknown aspect is thus a conservative best estimate for the hourglass dolphin (a range increase from 50 to 100 m would increase mean SL to 204 dB re 1 $\mu Pa_{p.-p.}$), whereas the estimate of ASL for Hector's dolphins was 177±6 dB re 1 $\mu Pa_{p.-p.}$ at 1 m (Table I).

Hector's dolphins came to bow ride, and as they were very close to the boat and the hydrophone array at the time of recording their source levels may likely have been lower than during natural foraging, which the hourglass dolphins were engaged in. However, Dawson and Thorpe (1990) also found low ASLs of ~150 dB re 1 $\mu Pa_{p.-p.}$ for Hector's dolphins foraging at the surface within 5 m from their boat. They also recorded clicks of up to 163 dB re 1 $\mu Pa_{p.-p.}$, but could not discern the vocalizing animal and thus estimate distance to the hydrophone.

If the source parameters measured in this study are representative for the two species, the ASLs of hourglass dolphins are an order of magnitude higher than those of Hector's dolphins. Hourglass dolphins have been found up to around 40 cm longer and 40 kg heavier than Hector's dolphin and it is possible that the higher ASL of hourglass dolphins can be ascribed to this size difference. However, Villadsgaard *et al.* (2007) reported that ASLs clicks of wild harbor porpoises, similar in size to Hector's dolphins, vary considerable between different recording sites (178–205 dB re 1 $\mu Pa_{p.-p.}$), possibly depending on background noise level and behavioral states of the animals. The variation in ASL of harbor porpoises thus spans both Hector's and hourglass dolphins, also spanning the size differences between the two. Hector's and hourglass dolphins likely also have a large dynamic range within which they produce clicks since we cannot ascertain that the full source level repertoire was sampled during these recordings.

Source level influences heavily on the range at which a dolphin can detect prey and the source level differences found here may thus provide a hint to the ranges at which the dolphins have adapted their sonars to search for prey. Detection range can only be estimated knowing the echo level threshold (DT) of a dolphin for a prey object with known target strength (TS). Kastelein *et al.* (1999) measured the psychophysical target detection threshold (expressed as echo energy flux density, EE) for a captive harbor porpoise echolocating at two water-filled stain-less steel spheres to be ~27 dB re 1 μPa^2 s (between 22.4 and 27.4 dB re 1 μPa^2 s); however, these calculations were based on too low source level (Au *et al.*, 2007) and the present best estimate of harbor porpoise detection threshold is 44–45 dB re 1 μPa^2 s (Au, personal communication). To calculate detection range Kastelein *et al.* (1999) used the active sonar equation ($EE=SE-2\times TL+TS_E=DT$) solved for EE at maximum range of detection, where SE is source EFD, TL is transmission loss, and TS_E is target strength energy. The non-noise-limited form of the sonar equation was used due to the low background noise usually found around 130 kHz (Kastelein *et al.*, 1999; Au *et al.*, 2007). Au *et al.* (2007) found that a 30 cm cod has a broadside target strength of −25 dB emitting artificial NBHF clicks with a centroid frequency of 130 kHz. Assuming that Hector's dolphin and hourglass dolphin have detection thresholds comparable to that of the harbor porpoise of 45 dB re 1 μPa^2 s and using the porpoise as a model we can thus estimate detection ranges of these two species. Under the assumption that both species were recorded with representative maximum source levels, Hector's dolphin is predicted to be able to detect a 25 cm cod broadside up to 10–24 m away, while hourglass dolphins are predicted to detect the same fish out to about 52–96 m, using minimum and maximum EFD source levels of 110 and 126 dB re 1 μPa^2 s and 140 and 152 dB re 1 μPa^2 s and respective absorption values (α, see above) (Fig. 4). Using the same assumptions wild harbor porpoises are predicted to detect the same cod 20–84 m away using minimum and maximum EFDs of 123 and 150 dB re 1 μPa^2 s and α of 0.04 (Villadsgaard *et al.*, 2007). The 25 dB lower EFD of Hector's dolphin clicks thus more than halves the detection range compared to the hourglass dolphin.

Hector's dolphins live within the coastal zone. Shallow coastal areas generally have a higher productivity and thus

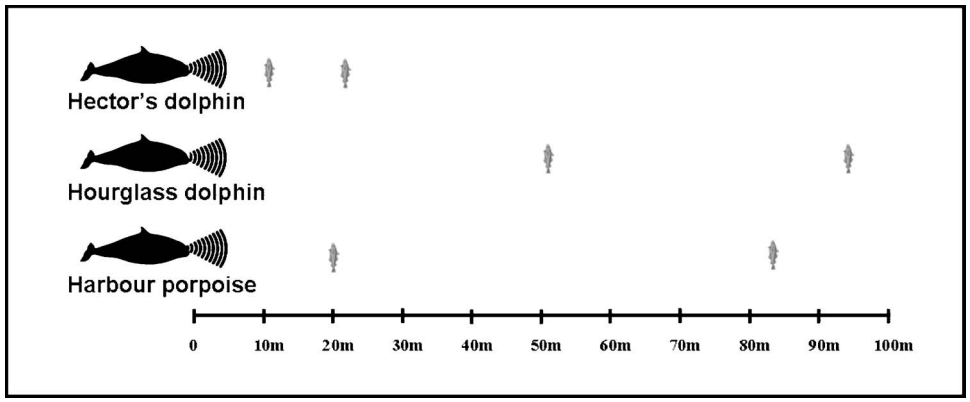


FIG. 4. Estimated detection ranges of hourglass dolphin and Hector's dolphin calculated for minimum and maximum EFD source levels using the harbor porpoise as model (Kastelein *et al.*, 1999, see text for further explanation) and target strength of a 25 cm cod measured broad side (Au *et al.*, 2007). Minimum and maximum detection ranges of harbor porpoise are shown for comparison, build on data from Villadsgaard *et al.* (2007) and Au (personal communication).

greater prey availability, but clutter and reverberation levels are also higher. Since an increase in source level also increases clutter this may explain why Hector's dolphins use a comparatively low source level. Increased transmission and receiving directivity will increase detection range in a cluttered environment since a more directional beam or sound reception will result in fewer unwanted echoes. Transmission directivity depends on size of the sound transducer relative to the emitted wavelength. If the sound production apparatus scales with head diameter of the animal (Au *et al.*, 1999), transmission directivity likely compares among the similar sized Hector's and hourglass dolphins presenting no special adaptation of Hector's dolphin to a cluttered environment. However, if prey density is also higher in the coastal environment, Hector's dolphin may not need high source levels to locate prey there. Future studies may uncover whether Hector's dolphins are capable of producing higher source levels at other behavioral states as observed for harbor porpoises (Villadsgaard *et al.*, 2007).

The higher source levels of hourglass dolphins oppositely suggest that they forage in an environment with lower prey density or with longer distances between prey patches. Since their click duration is also about twice that of Hector's dolphins, the energy content is increased two-ways compared to Hector's dolphin: higher source level (20 dB) and longer duration (doubling of energy content, 3 dB). If there is a size restricted maximum output for NBHF species, as suggested by Au (1993), the longer click duration suggests that hourglass dolphins may be noise limited and that they maximize energy content by making longer clicks to facilitate a longer detection range. The source parameters of hourglass' clicks may be the result of selection for increased target range in a noise-limited sonar situation.

For odontocete clicks, duration and bandwidth are inversely related (Wiersma, 1988; Au, 1993; Beedholm, 2008) so that a change in one parameter will change the other as well. It is therefore not surprising to find that bandwidth of hourglass dolphin clicks is about half that of Hector's clicks. However, it is interesting to note that despite differences in bandwidth both species have all click energy above 100 kHz [Figs. 1(a) and 2(a)] and the differences in source parameters observed between these two NBHF species are thus not in disagreement with the NBHF anti-predator hypothesis of Morisaka and Connor (2007).

The duration of Hector's dolphins' clicks of ~60 μs found in this study is lower than in previous published studies of 80–800 μs (Dawson, 1988; Dawson and Thorpe, 1990; Dawson, 1991; Thorpe *et al.*, 1991; Thorpe and Dawson, 1991). These differences in duration likely arise from the fact that previous studies included so-called double, triple, and quadruple pulsed clicks in their analysis, which we omitted here as they are most likely the result of surface and bottom reflections (Li *et al.*, 2005) and possibly off-axis distortions. In addition, previous studies measured duration by hand, whereas we used a definition of −10 dB from maximum amplitude of the signal envelope. The durations of Hector's clicks measured here correspond to that of another NBHF species the finless porpoise (*Neophocaena phocaenoides*) of ~30–60 μs (unknown recording aspect and definition of duration) (Akamatsu *et al.*, 1998) and are among the shortest of NBHF clicks. Villadsgaard *et al.* (2007), however, reported a range in 10 dB duration of harbor porpoise clicks from 44 to 113 μs thus spanning the range of both Hector's and hourglass dolphin clicks. It is possible that some click source parameters change with behavior and thus that the variation in duration between Hector's and hourglass' clicks could be caused by differences in behavior at the time of recording. Since ICIs change during, e.g., approach and capture phases of prey in other toothed whales (Johnson *et al.*, 2007), we used ICI as a proxy for behavior, and to test whether Hector's dolphins and hourglass dolphins changed their click duration and/or bandwidth with ICI we thus plotted each ICI preceding an on-axis click against its 10 dB duration and rms bandwidth for both species in Fig. 5. ICIs were generally longer for hourglass dolphins, but where the two species overlapped in ICI, the click duration and bandwidth were distinctly different (Fig. 5). Although it is possible that more recordings of the species measured here engaged in different behaviors may show a greater dynamic range in click duration resulting from changes in the context of sonar use, and not only from differences in habitat, this study suggests a genuine difference in click duration between Hector's dolphin and hourglass dolphin.

So far ASLs have only been measured for three free-ranging NBHF species: the harbor porpoise (Villadsgaard *et al.*, 2007) and the two present species (this study and Dawson and Thorpe, 1990). The highest levels were found for hourglass dolphins with a range 190–203 dB re 1 $\mu Pa_{p.-p.}$

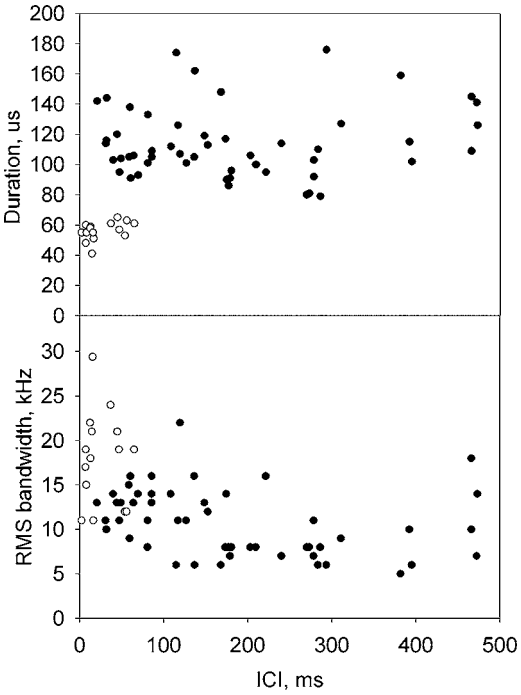


FIG. 5. rms bandwidth and 10 dB duration as a function of ICI. ICI is the one preceding the on-axis click for which bandwidth and duration was measured. Closed circles are hourglass dolphins and open circles are Hector's dolphins.

(Table I) and harbor porpoises with a range from 178 to 205 dB re 1 $\mu Pa_{p.-p.}$ (Villadsgaard *et al.*, 2007) and Hector's dolphins with a range 161–187 dB re 1 $\mu Pa_{p.-p.}$ (Table I). These levels are 10–30 times lower than in general for broadband dolphin clicks that have source levels of up to ~220 dB re 1 $\mu Pa_{p.-p.}$ (Au, 1993; Rasmussen *et al.*, 2002; Madsen *et al.*, 2004b), which again is reflected in the much longer detection range. It is precarious to generalize from only two species, especially with a very limited sample size in terms of behavior; but at least for *Cephalorhynchus* dolphins and the hourglass dolphin, offshore NBHF dolphins seem to produce clicks with higher source level than coastal dolphins do. To gain a better insight into the evolution and plasticity of the NBHF clicks this hypothesis should be tested by recording clicks from other *Cephalorhynchus* and porpoise species engaged in different behaviors. On the basis of the present findings we hypothesize that the oceanic Dall's porpoise will have a greater source level than coastal-offshore species such as the harbor porpoise and that riverine and very coastal species will have the lowest source levels.

ACKNOWLEDGMENTS

This study was conducted as part of the Danish Galathea3 oceanographic expedition, which took place in 2006–2007 and was organized and sponsored by the Danish Expedition Foundation. Special thanks goes to the crew on HDMS "Vædderen," in particular, Chief of Operations D. Vad and the RHIB crew M. Erdman and S. S. Nielsen for excellent seamanship and devotion to the project. This study was funded by Villum Kann Rasmussen's Foundation and Knud Højgaard's Foundation. Additional support from RESON A/S and National Instruments, Denmark is also acknowledged. This is publication No. P16 from the Galathea3 Expedition.

Au, W. W. L., Moore, P. W. B., and Pawlowski, D. (**1986**). "Transmission beam pattern of Atlantic bottlenose dolphin," J. Acoust. Soc. Am. **80**, 688–691.
Au, W. W. L. (**1993**). *The Sonar of Dolphins* (Springer-Verlag, New York).
Au, W. W. L., Kastelein, R. A., Rippe, T., and Schooneman, N. M. (**1999**). "Transmission beam pattern and echolocation signals of a harbor porpoise (*Phocoena phocoena*)," J. Acoust. Soc. Am. **106**, 3699–3745.
Au, W. W. L., Kastelein, R. A., and Benoit-Bird, K. J. (**2007**). "Modeling the detection range of fish by echolocating bottlenose dolphins and harbor porpoises," J. Acoust. Soc. Am. **121**, 3954–3962.
Au, W. W. L., and Benoit-Bird, K. J. (**2008**). "Broadband backscatter from individual Hawaiian mesopelagic boundary community animals with implications for spinner dolphin foraging," J. Acoust. Soc. Am. **123**, 2884–2894.
Akamatsu, T., Wang, D., Nakamura, K., and Wang, K. (**1998**). "Echolocation range of captive and free-ranging baiji (*Lipotes vexillifer*), finless porpoise (*Neophocaena phocaenoides*), and bottlenose dolphin (*Tursiops truncatus*)," J. Acoust. Soc. Am. **104**, 2511–2516.
Beedholm, K. (**2008**). "Harbor porpoise clicks do not have conditionally minimum time bandwidth product," J. Acoust. Soc. Am. **124**, EL15–EL20.
Bräger, S., Dawson, S. M., Slooten, E., Smith, S., Stone, G. S., and Yoshinaga, A. (**2002**). "Site fidelity and along-shore range in Hector's dolphin, an endangered marine dolphin from New Zealand," Biol. Conserv. **108**, 281–287.
Dawson, S. M. (**1988**). "The high frequency sounds of free-ranging Hector's dolphins, *Cephalorhynchus hectori*," Rep. Int. Whal. Comm. **9**, 339–344.
Dawson, S. M., and Thorpe, C. W. (**1990**). "A quantitative analysis of the sounds of Hector's dolphin," Ethology **86**, 131–145.
Dawson, S. M. (**1991**). "Clicks and communication: The behavioural and social contexts of Hector's dolphins communication," Ethology **88**, 265–276.
Fisher, F. H., and Simmons, V. P. (**1977**). "Sound absorption in sea water," J. Acoust. Soc. Am. **62**, 558–564.
Godall, R. N. P., Baker, A. N., Best, P. B., Meyer, M., and Miyazaki, N. (**1997**). "On the biology of the hourglass dolphin *Lagenorhynchus cruciger*," Rep. Int. Whal. Comm. **47**, 985–999.
Johnson, C. S. (**1967**). in *Marine Bioacoustics II*, edited by W. N. Tavolga (Pergamon, Oxford), pp. 247–260.
Johnson, M., Hickmott, L. S., Aguilar Soto, N., and Madsen, P. T. (**2007**). "Echolocation behaviour adapted to prey in foraging Blainville's beaked whale (*Mesoplodon densirostris*)," Proc. R. Soc. London, Ser. B **275**, 133–139.
Kastelein, R. A., Au, W. W. L., Rippe, T., and Schooneman, N. M. (**1999**). "Target detection by an echolocating harbor porpoise (*Phocoena phocoena)*," J. Acoust. Soc. Am. **105**, 2493–2498.
Li, S., Wang, K., and Wang, D. (**2005**). "Origin of the double-and multipulse structure of echolocation signals in Yangtze finless porpoise (*Neophocaena phocaenoides asiaeorientialis*)," J. Acoust. Soc. Am. **118**, 3934–3940.
Li, S., Wang, D., Wang, K., Akamatsu, T., Ma, Z., and Han, J. (**2007**). "Echolocation click sounds from wild inshore finless porpoise (*Neophocaena phocaenoides sunameri*) with comparisons to the sonar of riverine *N. p. asiaeorientalis*," J. Acoust. Soc. Am. **121**, 3938–3946.
Madsen, P. T. (**2005**). "Marine mammals and noise: Problems with root mean square sound pressure levels for transients," J. Acoust. Soc. Am. **117**, 3952–3957.
Madsen, P.T., and Wahlberg, M. (**2007**). "Recording and quantification of ultrasonic echolocation clicks from free-ranging toothed whales," Deep-Sea Res., Part I **54**, 1421–1444.
Madsen, P. T., Kerr, I., and Payne, R. (**2004a**). "Source parameter estimates of echolocation clicks from wild pygmy killer whales (*Feresa attenuata*) (L)," J. Acoust. Soc. Am. **116**, 1909–1912.
Madsen, P. T., Kerr, I., and Payne, R. (**2004b**). "Echolocation clicks of two free-ranging, oceanic delphinids with different food preferences: False killer whales *Pseudorca crassidens* and Risso's dolphins *Grampus griseus*," J. Exp. Biol. **207**, 1811–1823.
Madsen, P. T., Wahlberg, M., and Møhl, B. (**2002**). "Male sperm whale (*Physeter macrocephalus*) acoustics in a high-latitude habitat: Implications for echolocation and communication," Behav. Ecol. Sociobiol. **53**, 31–41.

Madsen, P. T., Carder, D. A., Beedholm, K., and Ridgway, S. H. (**2005**). "Porpoise clicks from a sperm whales nose—Convergent evolution of 130 kHz pulses in toothed whale sonars?," Bioacoustics **15**, 195–206.
May-Collado, L., and Agnarsson, I. (**2006**). "Cytochrome *b* and Bayesian inference of whale phylogeny," Mol. Phylogenet. Evol. **38**, 344–354.
Morisaka, T., and Connor, R. C. (**2007**). "Predation by killer whales (*Orcinus orca*) and the evolution of whistle loss and narrow-band high frequency clicks in odontocetes," J. Evol. Biol. **20**, 1439–1458.
Møhl, B., and Andersen, S. (**1973**). "Echolocation: High-frequency component in the click of the harbour porpoise (*Phocoena ph. L.*)," J. Acoust. Soc. Am. **54**, 1368–1372.
Møhl, B., Wahlberg, M., Madsen, P. T., Miller, L. A., and Surlykke, A. (**2000**). "Sperm whale clicks: Directionality and source level revisited," J. Acoust. Soc. Am. **107**, 638–648.
Møhl, B., Wahlberg, M., Madsen, P. T., Heerfordt, A., and Lund, A. (**2003**). "The monopulsed nature of sperm whale clicks," J. Acoust. Soc. Am. **114**, 1143–1154.
Neuweiler, G., and Fenton, M. B. (**1988**). "Behaviour and foraging ecology of echolocating bats," in *Animal Sonar: Processes and Performance*, edited by P. E. Nachtigall and P. W. B. Moore (Plenum, New York), pp. 535–549.
Reeves, R. R., Stewart, B. S., Clapham, P. J., and Powell, J. A. (**2002**). *Guide to Marine Mammals of the World* (Alfred E. Knopf, Inc., Chanticleer Press, New York).
Slooten, E., and Dawson, S. M. (**1988**). "Studies on hector's dolphin, *Cephalorhynchus hectori*: A progress report," Rep. Int. Whal. Comm. **9**, 315–324.
Slooten, E., Dawson, S. M., and Whitehead, H. (**1993**). "Associations among photographically identified Hector's dolphins," Can. J. Zool. **71**, 2311–2318.
Szymanski, M. D., Bain, D. E., and Kiehl, K. (**1999**). "Killer whale (*Orcinus orca*) hearing: Auditory brainstem response and behavioral audiograms," J. Acoust. Soc. Am. **106**, 1134–1141.
Rasmussen, M. H., Miller, L. A., and Au, W. W. L. (**2002**). "Source levels of clicks from free-ranging white beaked dolphins (*Lagenorhynchus albirostris* Gray 1846) recorded in Icelandic waters," J. Acoust. Soc. Am. **111**, 1122–1125.
Thorpe, C. W., and Dawson, S. M. (**1991**). "Automatic measurement of descriptive features of Hector's dolphin vocalizations," J. Acoust. Soc. Am. **89**, 435–443.
Thorpe, C. W., Bates, R. H. T., and Dawson, S. M. (**1991**). "Intrinsic echolocation capability of Hector's dolphin, *Cephalorhynchus hectori*," J. Acoust. Soc. Am. **90**, 2931–2934.
Urich, R. J. (**1983**). *Principles of Underwater Sound*, 3rd ed. (McGraw-Hill, New York).
Villadsgaard, A., Wahlberg, M., and Tougaard, J. (**2007**). "Echolocation signals of wild harbour porpoises, *Phocoena phocoena*," J. Exp. Biol. **210**, 56–64.
Wahlberg, M., Møhl, B., and Madsen, P. T. (**2001**). "Estimating source position accuracy of a larger-aperture hydrophone array for bioacoustics," J. Acoust. Soc. Am. **109**, 397–406.
Wiersma, H. (**1988**). in *Animal Sonar*, edited by P. E. Nachtigall and P. W. B. Moore (Plenum, New York), pp. 129–145.
Wood, F. G., Jr., and Evans, W. E. (**1980**). in *Animal Sonar Systems*, edited by R. G. Bushnell and J. F. Fish (Plenum, New York), pp. 381–426.

Asymmetries in the individual distinctiveness and maternal recognition of infant contact calls and distress screams in baboons

Drew Rendall[a)]
Department of Psychology, University of Lethbridge, Lethbridge, Alberta T1K 3M4, Canada

Hugh Notman[b)]
Department of Anthropology, Athabasca University, 1 University Drive, Athabasca, Alberta T9S 3A3, Canada

Michael J. Owren[c)]
Department of Psychology, Georgia State University, P.O. Box 5010, Atlanta, Georgia 30302-5010

(Received 2 September 2008; revised 11 December 2008; accepted 12 December 2008)

A key component of nonhuman primate vocal communication is the production and recognition of clear cues to social identity that function in the management of these species' individualistic social relationships. However, it remains unclear how ubiquitous such identity cues are across call types and age-sex classes and what the underlying vocal production mechanisms responsible might be. This study focused on two structurally distinct call types produced by infant baboons in contexts that place a similar functional premium on communicating clear cues to caller identity: (1) contact calls produced when physically separated from, and attempting to relocate, mothers and (2) distress screams produced when aggressively attacked by other group members. Acoustic analyses and field experiments were conducted to examine individual differentiation in single vocalizations of each type and to test mothers' ability to recognize infant calls. Both call types showed statistically significant individual differentiation, but the magnitude of the differentiation was substantially higher in contact calls. Mothers readily discriminated own-offspring contact calls from those of familiar but unrelated infants, but did not do so when it came to distress screams. Several possible explanations for these asymmetries in call differentiation and recognition are considered.
 [DOI: 10.1121/1.3068453]

PACS number(s): 43.80.Ka, 43.80.Lb [WWA] Pages: 1792–1805

I. INTRODUCTION

A central and continuing theme in research on nonhuman primates concerns the importance of kinship and individual identity in structuring species' social behavior (reviewed in Chapais and Berman, 2004). Many species are noted for strongly differentiated social relationships which suggest a capacity for discriminating among, and perhaps explicitly recognizing, individual conspecifics and biological kin (reviewed in Rendall, 2004; Widdig, 2007). They are also noted for the variety of vocalizations they produce to mediate social behavior in different contexts, in some of which their individualized social relationships are literally called into play. That is, the structure of some call types has been shown to be differentiated according to the kinship and individual identity of callers (e.g., Marler and Hobbett, 1975; Macedonia, 1986; Gouzoules and Gouzoules, 1990; Hauser, 1991; Rendall *et al.*, 1998; Fischer *et al.*, 2001), and, in some cases, field and laboratory experiments have also confirmed the perceptual salience and behavioral significance of these identity cues to listening animals (Hansen, 1976; Waser, 1977; Kaplan *et al.*, 1978; Cheney and Seyfarth, 1980; Snowdon and Cleveland, 1980; Symmes and Biben, 1985; Pereira, 1986; Rendall *et al.*, 1996; Hammerschmidt and Fischer, 1998; Jovanic *et al.*, 2000; Nunn, 2000; Fischer, 2004). Thus, the animals appear to recognize and respond to the vocalizations of different callers in ways that are functionally commensurate with their kinship or individual relationship to them. Taken together, these patterns suggest that the organization of primate vocal repertoires might be determined in part by the social networks they function to support and, in turn, that flexible management of social networks is to some extent affected by the structural differentiation of vocalizations along lines of social identity.

It is not yet clear, however, whether all calls in a species' repertoire are sufficiently distinctive to allow ready recognition of callers, or whether this is true only for certain call types for which there has been a selective premium on identity signaling and recognition. It is also often unclear exactly what vocal production mechanisms are responsible for generating distinctive voice cues and whether they might create identity cues that are common across call types (Ghazanfar and Santos, 2004; Ghazanfar *et al.*, 2007). For example, although the calls of different individuals could differ in a range of acoustic features, Rendall *et al.* (1998) proposed that patterns of vocal tract resonance might be a

a)Electronic mail: d.rendall@uleth.ca
b)Electronic mail: hughn@athabascau.ca
c)Electronic mail: owren@gsu.edu

particularly common and productive source of identity cues in primates. Such cues might arise inevitably, and also largely incidentally, as a product of idiosyncrasies in individual vocal tract development. They might also impart a characteristically distinctive voice pattern across most or all of the calls produced by a given individual because, regardless of their source characteristics, all calls must ultimately be radiated through, and thus be structurally shaped by, the supralaryngeal airways. At the same time, primates have long been suspected to have only limited ability to modify the resonances of these airways through dynamic vocal-tract articulations during vocal production (Lieberman *et al.*, 1969). Together, these effects might yield characteristically distinctive voice resonance patterns in each individual (Rendall *et al.*, 1998; Ghazanfar *et al.*, 2007). However, any one of these assumptions might be faulty, or at least too simplistic: vocal tract development might not always be sufficiently different between individuals to yield perceptually distinctive voice resonance patterns; some laryngeal sources (e.g., those with very high fundamental frequency, F_0) might not highlight the supralaryngeal resonances very well, and primates might have a greater capacity for vocal tract modulation than previously suspected (Hauser, 1992; Riede and Zuberbühler, 2003).

It is thus important that we continue to test the production and recognition of identity cues in different vocalizations to more fully understand the ubiquity and underlying vocal production mechanics of identity signaling in primates and thus also how vocal communication patterns in these species integrate with the functional demands of their individualistic social lives. In this paper, we report acoustic analyses examining individual distinctiveness in two distinct call types produced by infant baboons and then follow-up field experiments testing maternal recognition of infants' calls.

We focus specifically on baboons because they are among the best-studied primate species, yielding a comparatively solid understanding of their social life (Altmann, 1980; Smuts, 1985; Strum, 1987; Cheney and Seyfarth, 2007). There has also been considerable recent research on their vocal communication, including studies related to identity cueing and vocal recognition of group members which point to individually distinctive and recognizable calls in at least some kinds of vocalizations produced by adults (e.g., Cheney *et al.*, 1996; Owren *et al.*, 1997; Cheney and Seyfarth, 1999; Rendall *et al.*, 1999, 2004; Semple, 2001; Fischer *et al.*, 2001, 2002; Engh *et al.*, 2006). We focus the current work on infants then because, by comparison to adults, there has been relatively little systematic research on infant vocal repertoires, either in baboons or in other primate taxa. At the same time, infants are, by definition, the most vulnerable members of primate groups, subject to high risks of predation and, in some species including baboons, also infanticide from adult males (Palombit *et al.*, 1997, 2000). As a result, infants are the individuals that most often need social support, and most often solicit it. Finally, we focused also on the responses of mothers to their infants' calls because they are the most natural source of social support in primates. Thus, of all possible social relationships, that between infants and their mothers

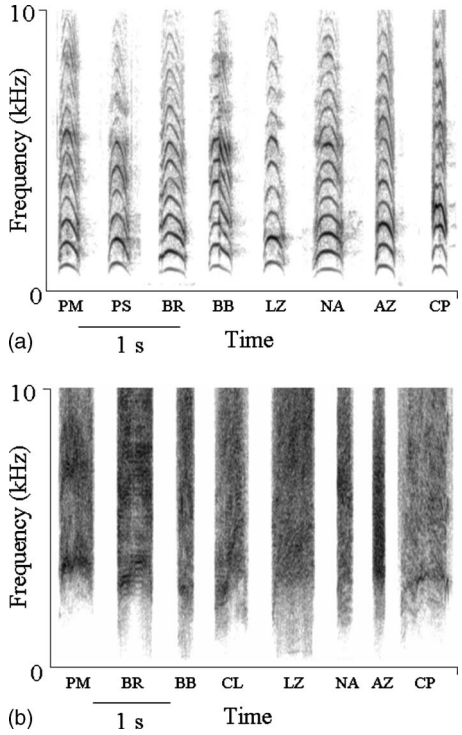


FIG. 1. Examples of contact calls (a) and distress screams (b) produced by infant baboons used in this study. Spectrograms were generated using a 1000-point FFT and a Gaussian window with 95% frame overlap.

was deemed the most likely to manifest unambiguous cues to identity in vocalizations and acute sensitivity to such cues.

The two contexts of vocalization chosen for study are common ones for baboons and most other social primates. The first involves situations in which individuals get physically separated from one another or the group at large. Such separation happens daily when the group is moving and foraging and group members invariably become widely dispersed, sometimes in dense habitats that preclude visual contact. For infants, it can also occur during group rest periods because infants often remain active playing with age mates. Sometimes they wander a considerable distance from their mothers and end up out of sight. In either situation, infants can become visibly agitated when they register the separation. They begin to scan their surroundings rapidly and repeatedly, sometimes climbing to an elevated position in a tree. Ultimately, they produce loud harmonically structured vocalizations in an attempt to relocate their mother [see Fig. 1(a)]. Adults produce calls of exactly analogous structure when they get separated from the group or from particular social companions (Cheney *et al.*, 1996). These adult calls appear to be individually distinctive and recognized by others, which are characteristics that would facilitate their presumed function of promoting contact and reunion with the group at large or with specific group members (Cheney *et al.*, 1996; Fischer *et al.*, 2001, 2002).

The second focal context involves aggressive conflicts, or fights, which occur commonly in the stratified social groups of baboons and other primates. For infants, such aggression occurs when play bouts escalate into more serious fights and when conflicts with peers or adults arise over access to food, water, or other valued resources. Especially dire

aggression can also occur when immigrant adult males try to injure or kill infants when left unprotected by their mothers (Palombit *et al.*, 2000). When attacked, the victims of aggression produce loud distress screams. These calls are among the most ubiquitous in primates, occurring in every species for which vocalizations have been described and, within species, in animals of every age and sex class (reviewed in Andrew, 1963; Gautier and Gautier, 1977; Marler and Tenaza, 1977; Oppenheimer, 1977). Screams can vary substantially in form, even within single bouts of calling, and can include broad-band noisy variants, pure-tone frequency-modulated variants, and various intermediate forms (Gouzoules and Gouzoules, 2000). However, the most common are the broad-band noisy variants [see Fig. 1(b)], which are also often the first and loudest screams produced when attacked. Research on species closely related to baboons suggests that noisy screams occur more commonly when aggression is most severe with the greatest risk of injury (Gouzoules *et al.*, 1984, 1986; Fugate *et al.*, 2008). These distress screams are proposed to function in recruiting intervention and support from mothers, collateral kin, or other social allies who might be out of sight but not earshot. This recruitment function is again thought to be facilitated by individually distinctive call structures which allow listeners to readily identify the caller and then to intervene or not based on their relationship to it (Gouzoules *et al.*, 1986; Gouzoules and Gouzoules, 1990; Fugate *et al.*, 2008).

Together, infant contact calls and distress screams constitute a productive comparison. Although they vary considerably in their detailed acoustic structure and mediate very different behavioral contexts, they are nevertheless structurally and functionally similar in being loud conspicuous calls that transmit considerable distances seemingly in the service of similar functional goals, namely, to capture the attention and recruit the support of mothers or other caregivers. Hence, both calling contexts place a selective premium on individually distinctive call structures that permit ready recognition and responding by mothers and potentially others. In the first part of this paper, we report acoustic analyses quantifying the degree of individual distinctiveness in these two call types. In the second part of this paper, we test maternal recognition of both call types.

II. INDIVIDUAL DISTINCTIVENESS IN INFANT CONTACT CALLS AND DISTRESS SCREAMS

A. Methods

1. Study site, subjects, and calling contexts

Research was conducted on free-ranging chacma baboons (*Papio hamadryas ursinus*) in the Moremi Game Reserve in northern Botswana, which lies within the Okavango Delta. The habitat is a mosaic of open grassy floodplains and wooded islands that rise a few meters above the floodplains. The subjects of research were the adult females and their immature offspring, from the main study group at this site (group C). At the time of research (February 1996–March 1997), this group contained approximately 75 individuals and had been studied continuously since 1977 (Cheney and Seyfarth, 2007). As a result, the animals were fully habituated to human observers and the social and genealogical history of natal animals was known.

For this project, research focused on immature baboons between the ages of 6–15 months. Baboons in this age range are socially immature and rely heavily on their mothers for social support and physical comforting. However, they are also motorically independent and thus often stray considerable distances from, and end up out of sight of, their mothers. It is during such absences from their mothers that young baboons are especially vulnerable to predators, and to aggression from others, including infanticidal attack. As a result, this context of visual separation between infants and their mothers defines a situation in which infants should be maximally motivated to provide unambiguous cues to identity in their contact calls or distress screams, and mothers in turn should be strongly motivated to detect and act on such cues. Our analyses were therefore limited to contact calls and distress screams produced in such contexts. For distress screams, we further limited our analyses to calls produced during aggressive attacks (from immature peers or adults) that involved direct physical contact that could potentially result in injury. These attacks are more serious than ones involving only threats and chases, and therefore they represent the aggressive context for which immature animals would benefit the most from agonistic support and to which mothers should be the most motivated to respond. During this study, no instances of infanticidal attacks were captured in vocal recordings or direct behavioral data collection. Hence, this most dire context of infant screaming is not included in analyses.

2. Vocalization recording

Infant contact calls and distress screams were recorded during the course of an ongoing regimen of behavioral sampling of mothers and their infants. Behavioral sampling involved collecting a continuous record of a focal animals' general activity state (e.g., resting, moving, and foraging), social behavior, and proximity to other group members. A variety of other spatial and ecological data were collected at the same time (for additional details, see Rendall *et al.*, 2000). In addition, a continuous record was made of all vocalizations produced and the ecological and social contexts associated with calling. Where possible, audio recordings were made of these vocalizations. Recordings were made using a Sony TC-D5M cassette tape recorder, a Sennheiser ME88 directional microphone with K3U powering module, and type-IV Sony metal cassette tapes. Because the baboons had been observed for decades and had been tape recorded by numerous previous investigators, the animals were accustomed to human observers and recording equipment. As a result, high-quality recordings could therefore be made at close range (<5 m).

3. The vocalization sample

To provide the most comparable and direct tests of individual differentiation in contact calls and distress screams, we attempted to create a completely balanced sample of the two call types as produced by the same set of individual

TABLE I. Summary of the vocalization sample.

Individual	Sex	Mother's rank[a]	Matriline[b]	Contact calls	Screams
Palm	F	1	1	26	21
Persephone	F	2	1	18	
Bart	M	4	2	59	19
Bilbo	M	5	2	16	34
Cleopatra	F	7	4		16
Lizzy	F	12	5	29	36
Nat	M	17	6	59	26
Amazon	F	21	7	17	20
CP	F	23	8	35	10
Total				259	182

[a]Total number of ranked adult females in group C=23.
[b]Total number of matrilines in group C=9.

infants. However, given the exigencies of fieldwork and inevitable variation in the number of recordings we had for different individuals, this was not entirely possible. Ultimately, we had a sufficient number of high-quality recordings of contact calls to permit statistical analysis for eight infants. For seven of these same infants, we had sufficient scream recordings as well. To avoid an unbalanced number of individuals in statistical tests of individual distinctiveness in the two call types, we therefore added to the scream sample an additional infant of similar age for which we also had sufficient recordings. The sample thus constituted consisted of 259 contact calls recorded from 91 different episodes of calling from eight infants, and 182 distress screams recorded from 40 different episodes of calling from eight infants, seven of which were identical to those in the contact call sample (see Table I). These infants provided broad sampling of mothers of varying social rank and representing several different martilines. They also included both male and female infants and included some variation in the age of individuals within the 6–15 month bracket. In this range, age and sex differences in call structure are probably minimal. Nevertheless, they represent additional potential sources for variation in call structure that, in the strictest sense, might not be attributable to individual identity *per se*. At the same time, they are additional relevant dimensions of callers in naturally constituted groups that, in practical terms, should help to facilitate their identification and recognition by others. Hence, we make no further attempt to control for or eliminate their effects, but instead fold them into the naturalistic problem of vocal identification and recognition.

4. Acoustic analysis

Vocal recordings were digitized at 44.1 kHz with 16 bit accuracy after lowpass filtering at 20.0 kHz, and acoustic analyses were performed using PRAAT 4.6 (Boersma, 2001). Analysis focused on acoustic features that could be measured from single vocalizations. Both contact calls and distress screams often occur in bouts in which single vocalizations are concatenated into longer series, and it is possible that additional identity cues might be carried in the more protracted temporal organization of call sequences. However, most previous analyses of these and similar call types in primates have omitted this more complex level of organization and focused instead on the structure of single vocalizations; hence, we chose to follow this precedent to improve the comparability of our results to previous work. Furthermore, it seemed reasonable to predict, *a priori*, that if conveying cues to identity to distant listeners were an important function of these two call types then a system in which those cues are carried within each of many repeated single vocalizations would be far more functional, efficient, and redundant than one in which identity cues emerge only if and after listeners can hear a complete sequence of calls.

Acoustic analysis of contact calls and distress screams involved three broad classes of variables addressing their temporal, spectral, and intensity characteristics. For both call types, we measured the duration of each call and the relative timing of the intensity peak within the call. Although contact calls and distress screams both had broad-band spectral structures, contact calls characteristically involved a relatively stable F_0 with multiple harmonic partials, while distress screams were predominantly harsh and noisy. Therefore, different variables were required to characterize the detailed spectral structure of each call type.

For contact calls, we made multiple measurements of the F_0 contour which often contained an appreciable chevron-shaped modulation over the course of a single call. We used PRAAT to extract the F_0 ("pitch") contour for each call and from this F_0 contour we retained as variables, its beginning frequency, ending frequency, maximum frequency, and mean frequency. We also measured the relative time of the maximum F_0 within the call and used this variable to derive measures of the beginning and ending slope of the F_0 contour. We characterized potential instabilities in the voicing associated with the F_0 of each call using a measure of vocal jitter that corrects for longer-term F_0 modulation (the PPQ5 algorithm in PRAAT) and using the harmonic-to-noise ratio (HNR). Finally, to characterize the broader spectral envelope of contact calls as described by their harmonic partials, we measured the amplitude of each harmonic (up to 6) from a single 2048–point (50 ms) fast-Fourier transform (FFT) straddling the midpoint of each call. Because all calls were recorded at close range (<5 m), these amplitude measurements were unlikely to be affected by variable spectral

transmission effects or variation in ambient environmental conditions. Nevertheless, as an additional precaution, we converted our absolute measurements into relative amplitude (RelAmp) values by subtracting the amplitude of each harmonic from the mean amplitude of all harmonics (including the focal harmonic) within a given call. The resulting Rel Amp values thus express the strength of a particular harmonic relative to the other harmonics in the same call.

Because screams lacked a stable F_0, we could not make the same F_0 and harmonic measurements for these calls. It was also the case that the patterning of diffuse noisy energy within screams could vary appreciably over the course of a single call which further complicated spectral analysis of these calls. Nevertheless, to try to capture the global spectral patterning of screams, we divided the frequency spectrum into a series of nonoverlapping 500 Hz frequency bins (up to 12 kHz) and measured the amplitude of the signal spectrum within each bin. We then converted these values to RelAmp values for each bin using the same procedure that was used to generate Rel Amp values for the contact calls and for the same precautionary reasons. The resulting relative amplitude values for screams thus express the strength of the vocal signal in a particular frequency bin relative to the other frequency bins in the same call. For screams, this procedure was conducted using a single long-term average spectrum computed for each call and resulted in a set of 24 Rel Amp values characterizing the global distribution of energy in these calls over the frequency range spanning 0–12 kHz.

5. Statistical Analysis

We used multivariate discriminant analysis to evaluate individual differentiation in the structure of contact calls and screams. Because the results of such analyses can be dramatically influenced by the number of predictor variables used, we first conducted a principal components analysis (PCA) on the acoustic features measured for each call type. This preliminary step allowed us to examine natural covariation among the original acoustic features according to their associations within and between PCA factors. It also effectively reduced the number of predictor variables to be used in discriminant analysis to a small set of orthogonal multivariate dimensions that retained most of the variation in the entire set of original acoustic features. We then used the multivariate factors identified from PCA as variables in discriminant analysis of each call type. In evaluating successful call classification from discriminant analysis we followed a conservative split-sample approach (Klecka, 1980). For each call type, the sample was randomly divided in half. One-half of the cases was used as the reference sample to generate discriminant functions which were then applied to the other half of cases, the test sample. All statistics were conducted using NCSS version 5.1 (Hintze, 1999).

B. Results and discussion

1. Acoustic features of calls and possible production mechanisms

a. Contact calls. Summary statistics for the acoustic features of each call type are given in Table II. Contact calls were shorter, on average, than distress screams and were characterized by a relatively stable F_0 with relatively low jitter values and a very high mean HNR (12.35 dB) which represents an average HNR of 18:1. The F_0 contour of con-

TABLE II. Descriptive statistics for acoustic features of contact calls (a) and distress screams (b). The association of acoustic features with PCA factors and the significance (*) ($p<0.001$) of those factors in discriminant analysis tests of individual differentiation are shown.

(a) Contact calls

Variable	M	CV	PCA factors[a]
Duration (ms)	259	0.201	1
Intensity peak (as % duration)	46.3	0.305	2*, 3*
F_0-begin (Hz)	499	0.192	5*
F_0-end (Hz)	465	0.166	
F_0-maximum (Hz)	756	0.076	1
F_0-mean (Hz)	653	0.080	1
F_0-slope-begin (Hz/s)	2780	0.361	5*
F_0-slope-end (Hz/s)	1820	0.048	1
F_0-maximum (as % duration)	0.383	0.261	2*
Jitter (%)	0.034	0.179	
HNR (dB)	12.35	0.244	4*
H1 RelAmp (dB)	−8.85	0.118	1
H2 RelAmp (dB)	8.43	0.156	7*
H3 RelAmp (dB)	5.64	0.169	
H4 RelAmp (dB)	−1.03	0.141	
H5 RelAmp (dB)	−2.67	0.103	6*
H6 RelAmp (dB)	−1.52	0.108	4*
Overall mean		0.171	

(b) Distress screams

Acoustic feature	M	CV	PCA factors[b]
Duration (ms)	351	0.810	4
Intensity peak (as % duration)	0.464	0.488	4
Percent voiced (%)	0.154	0.064	7
RelAmp 0–500 Hz (dB)	−11.25	0.561	2*
RelAmp 500–1000 Hz (dB)	−12.29	0.831	2*
RelAmp 1000–1500 Hz (dB)	−9.37	0.532	1, 2*
RelAmp 1500–2000 Hz (dB)	−4.61	0.322	1
RelAmp 2000–2500 Hz (dB)	1.35	0.232	1
RelAmp 2500–3000 Hz (dB)	8.22	0.193	1
RelAmp 3000–3500 Hz (dB)	9.38	0.197	1
RelAmp 3500–4000 Hz (dB)	7.32	0.173	1
RelAmp 4000–4500 Hz (dB)	5.85	0.184	1
RelAmp 4500–5000 Hz (dB)	5.91	0.182	1
RelAmp 5000–5500 Hz (dB)	6.13	0.172	1
RelAmp 5500–6000 Hz (dB)	6.28	0.162	1
RelAmp 6000–6500 Hz (dB)	6.05	0.156	2*, 3*
RelAmp 6500–7000 Hz (dB)	5.65	0.157	2*
RelAmp 7000–7500 Hz (dB)	4.74	0.172	2*
RelAmp 7500–8000 Hz (dB)	3.71	0.179	1, 2*
RelAmp 8000–8500 Hz (dB)	2.43	0.193	1
RelAmp 8500–9000 Hz (dB)	0.71	0.222	1
RelAmp 9000–9500 Hz (dB)	−1.72	0.265	1
RelAmp 9500–10 000 Hz (dB)	−4.06	0.321	1
RelAmp 10 000–10 500 Hz (dB)	−5.80	0.363	1
RelAmp 10 500–11 000 Hz (dB)	−7.06	0.369	1
RelAmp 11 000–11 500 Hz (dB)	−8.18	0.385	1
RelAmp 11 500–12 000 Hz (dB)	−9.40	0.423	1
Overall mean		0.308	

[a]PCA factor 8 was not associated with any single specific acoustic feature.
[b]PCA factors 5, 6, and 8 were not associated with any single specific acoustic features.

tact calls showed a consistent rising and falling pattern with an average frequency excursion of approximately 150 Hz, starting from approximately 500 Hz and finishing slightly lower in frequency. Contact calls also showed a consistent pattern of harmonic emphasis in which the second and third harmonics were stronger than either the first harmonic (the F_0) or higher harmonics (H4–H6). This pattern of differential harmonic emphasis might be inherent to the laryngeal source spectrum. Alternatively, it might reflect some degree of supralaryngeal vocal-tract filtering.

Vocal-tract filtering effects have been shown to account for spectral patterning in similar harmonically rich grunt vocalizations produced by adult baboons (Owren *et al.*, 1997; Rendall, 2003) as well as in harmonically rich contact calls produced by closely related rhesus monkeys (Rendall *et al.*, 1998; Ghazanfar *et al.*, 2007). For baboons, the F_0 of adult grunts ranges between 50 and 120 Hz (for males and females, respectively: Owren *et al.*, 1997; Rendall *et al.*, 2004) and thus affords excellent harmonic sampling of vocal tract resonances which are prominent in the spectral structure of these calls. By comparison, the F_0 of infant baboon contact calls is quite high (500 Hz) and the consequently wide spacing of harmonics inherently affords much sparser sampling of vocal tract resonances, which might then argue against vocal-tract filtering effects accounting for differential harmonic emphasis in these calls. At the same time, infant baboons also have a relatively short vocal-tract compared to adults. Hence, it is possible that the second harmonic is emphasized in infant contact calls, relative to the F_0, because it coincides with the first resonance of their immature vocal tract.

We are not aware of any systematic data on vocal-tract length (VTL), and thus resonant frequency locations, in infant baboons that could be used to evaluate this possibility more directly. However, we can attempt some preliminary indirect assessment by evaluating the plausibility of the estimated (VTL) that would result from a first resonance peak coincident with the observed H2 harmonic emphasis. VTL can be estimated from resonance frequencies using the following equation: $\mathrm{VTL}=(2k+1)c/4F$, where F is the formant (or resonant) frequency, k is the formant number minus 1, and c is the speed of sound in air (34 400 cm/s). We can then use H2 frequency values of 1000 and 1300 Hz to bracket the possible range of the putative first resonance based on values from Table II. The resulting estimates of VTL for this sample of infant baboons would be 6.6–8.6 cm. These values are approximately 1/3 to $\frac{1}{2}$ of VTL estimates for adult baboons (17–19 cm; see Owren *et al.*, 1997). Given the marked difference in overall body and head size between adults and infants in this age range, our estimated VTL for infants appears reasonable. However, without more systematic VTL measurements for both age classes, it remains speculative.

b. Distress screams. By comparison to contact calls, distress screams were approximately 40% longer in duration and were characterized by a predominantly aperiodic source with detectable voicing occurring in only 15% of analysis frames. The resulting noisy spectrum of screams was extremely broadband, with detectable energy extending up to and beyond 12 kHz. However, at this frequency and beyond, signal strength had fallen off appreciably. Within the range from 0 to 12 kHz, distress screams displayed some additional substructuring across individuals. Most calls showed relatively discrete attenuation of low-frequency components (0–2500 Hz) in addition to the inherent gradual attenuation of signal strength at higher frequencies (8500–12 000 Hz), which left a broadly emphasized region of the spectrum between 2500 and 8500 Hz. The bandwidth of emphasized frequencies within this broad region could vary somewhat between calls, and, on average, the region of greatest emphasis was between 2500 and 4000 Hz. However, in general, within the broader region between 2500 and 8500 Hz there was little obvious subpatterning.

It is difficult to evaluate the likely production mechanisms responsible for these patterns in screams. One possibility is that the excess attenuation of low-frequency components in distress screams reflects some vocal-tract filtering (antiresonances) of the aperiodic laryngeal source spectrum. However, if this were true, we would also expect to see some obvious peaks in the broader spectrum between 2500 and 8500 Hz reflecting vocal-tract resonances in that region that should have been well sampled by the strong aperiodic source. The alternative is that the relatively unstructured noisy spectrum of distress screams stems primarily from the laryngeal source and reflects extremely high-amplitude driving of the vocal folds inducing chaotic (or nearly chaotic) vocal-fold activity which has been shown in closely related species (Fitch *et al.*, 2002; Riede *et al.*, 2004, 2007; Tokuda *et al.*, 2002). The absence of low-frequency components in the source spectrum could then be a by-product of the extreme amplitude of the source which would naturally drive the vocal folds at higher frequencies. This effect, combined with the relatively short vocal folds of immature baboons which naturally limit low-frequency vibratory modes to begin with, could effectively eliminate most low-frequency energy in the calls.

Although the exact role that vocal-tract filtering may play in the observed spectral patterning of either call type remains uncertain, there was little indication that it acted within individuals to impart a common resonance pattern across the two call types (see Rendall *et al.*, 1998).

2. Individual variation in calls

Coefficients of variation (CVs) were calculated for each acoustic feature and for each individual, and grand means for these values are provided in Table II. These CVs quantify the relative stability of specific acoustic features within individuals and, as a result, are often taken as a preliminary metric of the potential for reliable identity signaling. The manifest variability in these CVs for both call types suggests that some features might prove better than others in this respect. Overall, CVs for the acoustic features of contact calls were lower and less variable than were those for distress screams (contact calls: mean CV=0.171, range=0.048–0.361; distress screams: mean CV=0.308, range=0.064–0.831), perhaps suggesting greater potential for reliable identity signaling in contact calls compared to screams.

a. Contact calls. PCA on the acoustic features of contact calls identified a set of eight orthogonal factors with eigenvalues greater than 1. Together, these factors accounted for 80.81% of the variation in contact calls encompassed by the original set of 17 acoustic features. For seven of the eight factors, one or more of the original acoustic features had a factor loading of 0.5 or higher and thus could be explanatorily linked to it (see Table II).

Discriminant analysis using PCA scores for these eight factors produced an overall Wilk's lambda of 0.063, which is very close to the lower limit of 0.0 for this test statistic (which would indicate perfect discrimination among groups) and represents a statistically significant degree of differentiation in call structure between individuals ($F=15.8$, $P<0.0001$). This analysis identified five statistically significant dimensions of discrimination among individuals which involved all but one of the PCA factors and encompassed a mix of original acoustic features related to intensity, the F_0 contour, and harmonic amplitudes (see Table II). Notably, these acoustic features were not necessarily those with the lowest CVs. Some acoustic features with low CVs (e.g., H2-RelAmp, H5-RelAmp, and H6-RelAmp) did help to differentiate individuals but some others with low CVs did not (e.g., F_0-mean, F_0-max). At the same time, some features with relatively high CVs nevertheless contributed to differentiating individuals (e.g., intensity peak, F_0-slope-begin). These outcomes indicate that, despite their regular reporting in this kind of research, CVs are a quite imperfect metric of individual differentiation probably because the relative stability of features within individuals says little about the potential variability in these features between individuals.

It is also noteworthy that the only PCA factor that did not contribute to individual differentiation in contact calls was the first factor, which accounted for more variation in contact calls (23.8%) than any other single factor from PCA and more than 25% of the variation in the calls that was explained by the entire set of eight factors. This outcome indicates that a considerable amount of the measured variability in contact calls is unrelated to individual identity. Nevertheless, discriminant functions successfully classified the majority of calls to the infant that produced them. Classification success for the full sample of 259 calls was 73.3% (range: 62.5%–86.2% for different individuals). Using the more conservative split-sample approach, classification success dropped marginally to 70.5% (range: 44.4%–86.0%), which represents a 66.0% reduction in the error that would accompany random classification of the calls. Figure 2 plots each call in a bivariate space captured by the first two canonical variates from discriminant analysis and illustrates the degree of multidimensional overlap and separation of individuals' calls.

b. Distress screams. PCA on the acoustic features of screams identified only five factors with eigenvalues greater than 1. Together, these factors accounted for 80.4% of the variation in these calls encompassed by the original set of 27 acoustic features. Compared to contact calls then, the measured variation in screams was packaged into fewer significant orthogonal dimensions. To maintain comparability in our analyses of the two call types, we nevertheless retained the first eight PCA factors for screams for use in discriminant analysis tests of individual differentiation in these calls. Together, this set of eight factors accounted for 89.1% of the measured variation in screams, which actually represented approximately 10% more of the variation in these calls than was used to test individual differentiation in contact calls. Once again, for several of these factors, one or more of the original acoustic features could be explanatorily linked to it (see Table II).

Discriminant analysis using PCA scores for these eight factors produced an overall Wilk's lambda of 0.195, which is higher than that obtained for contact calls and thus suggests less individual differentiation in the calls, but this

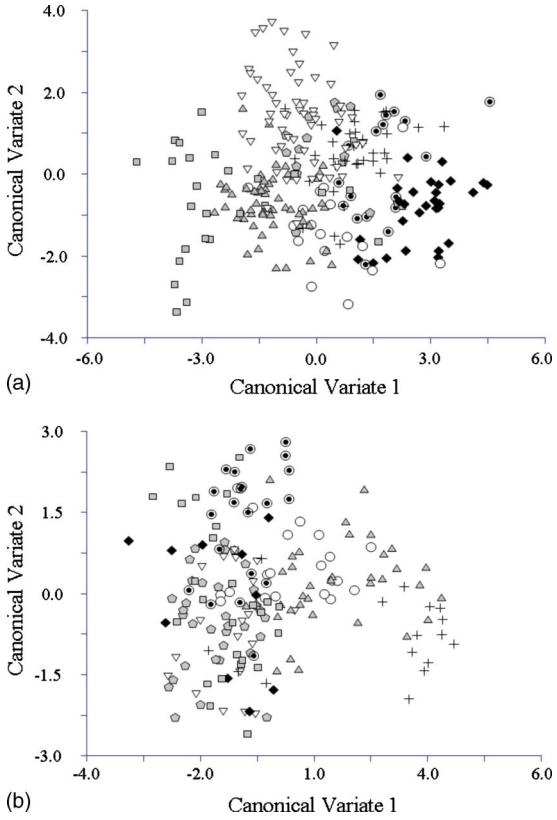


FIG. 2. (Color online) Scatterplots depicting each infant's contact calls (a) and distress screams (b) in the multidimensional space captured by the first two canonical variates from discriminant analysis tests of individual differentiation. Each individual is represented by a different symbol. The relative distinctiveness of individuals' calls is manifested in the degree of symbol separation and overlap.

value was nevertheless statistically significant ($F=5.7$, $P<0.0001$). Analysis of screams identified four statistically significant dimensions of discrimination among individuals which involved only five of the eight PCA factors and encompassed seven of the original acoustic features (see Table II). The original acoustic features contributing to individual differentiation in screams pertained only to the attenuated low-frequency spectral bins (0–1500 Hz) and a set of mid-frequency spectral bins (6000–8000 Hz). Once again, these acoustic features were characterized by some of the lowest CVs but also some of the highest CVs.

For screams it was also true that the first PCA factor did not contribute to individual differentiation in screams despite accounting for a substantial proportion of the measured variation in these calls (48.7%) and more than 50% of that encompassed by the entire set of eight PCA factors used for discriminant analysis. Three additional PCA factors (4, 5, and 7), accounting for an additional 12% of the measured variation in these calls, did not contribute to individual differentiation. Overall then, the majority of measured variation in distress screams is unrelated to individual identity. Despite this fact, discriminant functions applied to the full sample of 182 screams classified 45.1% to the infant that produced them (range: 30.0%–68.8%), and classification success using the more conservative split-sample approach was similar at 46.2% (range: 0.0%–75%). The latter classification success represents a 38.5% reduction in the error that would accompany random classification of the calls. These classification rates are substantially lower than those obtained for contact

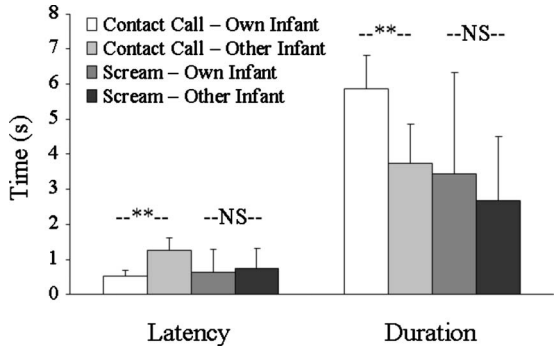


FIG. 3. Results of playback trials testing mothers' ability to discriminate the contact calls and distress screams of their own infant from those of a familiar but unrelated infant matched for age and sex. The data plotted represent the mean (+SEM) latency and duration of orienting responses by female baboons in the different experimental conditions. Females responded significantly faster and longer to contact calls in own than in other infant trials ($^{**}P<0.01$). Their responses to distress screams in own and other infant trials did not differ significantly (NS, $P>0.05$). See text for test statistics.

calls, and this can be seen graphically in Fig. 3 where multidimensional overlap in individuals' calls is greater for screams than for contact calls. Nevertheless, the level of successful classification for screams was also appreciably above the simple theoretical chance threshold for this sample (12.5%).

Taken together, acoustic results suggest a potential for individual identity signaling in both call types consistent with the hypothesis that the functions of both call types include capturing the attention and soliciting support of some kind from specific other group members, in this case most likely from their mothers. The evidence suggests somewhat greater potential for this in contact calls than screams, but no firm conclusion can be drawn. Ultimately, how distinctive or not either call type is depends on the ability of others actually to discriminate among callers. And, here, it is quite possible that the differences identified in our acoustic analyses are not actually perceptible to mothers. It is also possible that there exist differences in one or both call types that are not captured by our acoustic analysis but that nevertheless facilitate recognition. Therefore, it is also critical to test maternal recognition of both call types.

III. MATERNAL RECOGNITION OF INFANT CONTACT CALLS AND DISTRESS SCREAMS

To test maternal recognition of infant contact calls and distress screams we undertook a naturalistic field playback experiment. The experiment involved playing back prerecorded calls to mothers in standardized contexts and measuring their responsiveness to those produced by their own infant versus those produced by another familiar but unrelated infant in the group matched for age and sex. The experiment was designed to match our acoustic analyses and therefore focused on maternal recognition of single contact calls and distress screams, rather than protracted bouts of either type.

A. Methods

1. Subjects

The subjects of experimental trials were the mothers of the same infants used in acoustic analyses, thereby providing the most direct test possible of the results of our acoustic analyses of individual differentiation in these infants' calls. Limiting the experiment to only these females, however, would have yielded a precariously small experimental sample of only eight subjects for this test. Therefore, to increase the sample of subjects, the power of our test, and the potential generality of its findings, we included as subjects additional group females with infants of similar age to those used in acoustic analyses.

2. Experimental design

The experiment was conducted as a set of matched playback trials. For each call type, a given female was the subject of two trials: one trial involving the calls of her own infant ("own" infant trials) and the other trial involving the calls of a familiar, but unrelated, age- and sex-matched infant in the group ("other" infant trials). In each trial, the subject was played two different exemplars of the same call type (either contact calls or distress screams) produced by their own or another infant separated by a 1 s interval of silence. We played a pair of calls in each trial, rather than only a single vocalization, to reduce the chance that, given the exigencies of ambient noise and activity, subjects might not hear or register a single call. In effect then, the first of the two calls in a pair primed subjects for the second call, and the short interval between calls increased the likelihood that they could be coming from the same caller.

Specific vocalizations were chosen as stimuli for use in the experiment if they were of uniformly high quality, uncontaminated by background noise or other calling. The set of playback stimuli ultimately used in experiments consisted of 19 pairs of contact calls (38 different exemplars) originally recorded from 15 different infants and 14 pairs of distress screams (28 different exemplars) originally recorded from 13 different infants.

Following the logic of many similar studies testing vocal recognition abilities, we reasoned that if mothers could recognize the calls of their own infant and discriminate them from the calls of other infants in the group, then they should respond more strongly to their own infant's calls. Because social rank (particularly matrilineal rank in Old World monkeys) is an additional important dimension of primate sociality potentially affecting animals' responses to the vocalizations of others, trials were conducted to systematically control for the influence of this factor. Thus, in other infant trials, half the subjects were played the calls of an infant whose mother ranked above them in the hierarchy, and the other half were played the calls of an infant whose mother ranked below them in the hierarchy.

3. Playback trials

Playback trials were conducted during the course of focal observations on mothers as appropriate conditions arose. Hence, the choice of subject for any given trial was determined by the ongoing regimen of behavioral sampling (see Rendall *et al.*, 2000). Trials on the same mother were separated by several days. The order of own versus other infant trials was balanced across subjects. Trials were conducted only if the mother's infant (and, in the case of other infant

trials, the unrelated infant whose calls were to be played) was some distance away and out of sight and could reasonably be located in the direction from which calls were to be played. Because mothers and their infants were frequently separated from one another, these conditions occurred regularly. Trials were also only conducted when the mothers were either resting, self-grooming, moving slowly on their own, or foraging in place and therefore were undistracted by other group members or their activities. Calls were played through a Nagra DSM loudspeaker positioned approximately 20 m from the subject in the direction from which the infant was last seen. The same speaker amplitude settings were used for all playback trials and were established during a pretesting period to naturally simulate a caller at approximately 50 m.

4. Maternal response assays

Playback trials were videotaped for later scoring. Response assays included (1) whether or not the subject actively attended to call playback as evidenced by deliberate head or body orientation in the direction of the speaker, (2) how rapid was the orientation (latency to orient), and (3) for what period of time it lasted (duration of orientation) in the 10 s immediately following stimulus presentation. In conducting trials, care was taken to ensure that subjects were facing away from the speaker prior to stimulus presentation so that changes in orientation toward the speaker after stimulus presentation could be evaluated unambiguously. In the few cases (four trials) where subjects were oriented in the direction of the speaker for a brief portion of the 10 s interval preceding stimulus presentation, the duration of this incidental orientation toward the speaker in the preplayback period was subtracted from the duration of the subjects' orientation in response to the calls in the postplayback period to arrive at a conservative measure of the subject's response. The latency and duration of subjects' orienting responses were calculated from the onset of the first call in each playback stimulus.

B. Results and discussion

In total, 14 mothers were tested in paired trials involving the contact calls of their own and another infant, and 13 mothers were tested in paired trials involving distress screams. In both cases, mothers were clearly interested in the calls and motivated to respond. They showed active orientation toward the playback speaker in 20 of the 28 trials (71.4%) involving contact calls and in 16 of the 26 trials (61.5%) involving distress screams, and in almost all cases immediately upon hearing the calls (Fig. 3). The average latency to respond to contact calls was 0.79 s [$\pm$0.21 s standard error of mean (SEM)] from the onset of the first call in the stimulus pair, while the average latency to respond to distress screams was 0.76 s ($\pm$0.21 s SEM). On average, then, females started responding $\frac{1}{4}$ of a second before hearing the second of the two calls in a stimulus pair. At times, females abruptly stopped their current activity (e.g., moving, foraging in place, and self-grooming) and craned their necks, or adjusted their whole posture, to look more directly in the direction of the simulated caller. In seven of the trials involving contact calls (five involving their own infant), mothers also either got up and moved some distance toward the speaker to investigate further or climbed to an elevated position in a tree or on a stump to scan the surroundings. Such additional behaviors were not observed in response to screams.

1. Maternal responses to contact calls

In the case of contact calls, females also showed clear discrimination between calls produced by their own and other infants. Females were more likely to orient toward the speaker in trials involving their own infants' calls, 13 of 14 females orienting to own infant trials compared to only 7 of 14 orienting in other infant trials (chi-square test: $\chi_1^2=6.30$, $P=0.012$). They also responded significantly faster and longer to calls produced by their own infant than to those produced by other infants (Wilcoxon matched-pairs test: response latency, $Z=2.83$, $P<0.01$; response duration, $Z=2.57$, $P<0.01$).

2. Maternal responses to distress screams

In contrast, females' responses to distress screams provided comparatively little evidence that they discriminated between calls produced by their own and other infants. 10 of 13 females responded with deliberate orientation to the playback speaker in own infant trials and 6 of 13 also oriented in other infant trials. This difference suggests a similar tendency for mothers to respond preferentially to their own infants' screams but the difference was not statistically significant (chi-square test: $\chi_1^2=2.60$, $P=0.11$). Similarly, females responded marginally faster and longer, on average, in own than other infant trials (Fig. 3) but here again the differences were not statistically significant (Wilcoxon matched-pairs test: response latency, $Z=1.16$, $P=0.12$; response duration, $Z=1.36$, $P=0.09$).

3. Effects of social rank

Neither set of outcomes proved to be influenced by the social rank of other infants. For contact calls, there appeared to be some tendency for females to differentiate among other infants. Thus, females oriented in five of seven trials involving other infants whose mother ranked higher than them in the social hierarchy compared to only two of seven trials involving other infants whose mother ranked lower in the social hierarchy. However, this difference was not statistically significant (chi-square test: $\chi_1^{2=}2.57$, $P=0.11$). Furthermore, although females' responses to all other infants were markedly slower and shorter than they were to their own infants, they responded faster and marginally longer, on average, in other infant trials that involved infants with mothers of higher as opposed to lower maternal rank (latency: higher ranking mother, mean$=3.96\pm1.56$ s SEM; lower ranking mother, mean$=7.28\pm1.76$ s SEM; duration: higher ranking mother, mean$=2.01\pm0.91$ s SEM; lower ranking mother, mean$=1.75\pm1.26$ s SEM). However, here again, the differences were not statistically significant (latency: $Z=0.96$, $P=0.17$; duration: $Z=0.82$, $P=0.21$).

For distress screams, there were no similar hints of discrimination among other infants according to maternal social rank. Females oriented in an equal number of the trials involving other infants of either higher or lower maternal rank (three of seven trials in each case; chi-square test: $\chi_1^2 = 0.07$, $P = 0.80$). And their response latencies and durations to these two classes of other infant were nearly equal (latency: higher ranking mother, mean $= 5.34 \pm 2.09$ s SEM; lower ranking mother, mean $= 6.06 \pm 1.86$ s SEM; duration: higher ranking mother, mean $= 1.58 \pm 0.97$ s SEM; lower ranking mother, mean $= 0.93 \pm 0.48$ s SEM) and not significantly different (latency: $Z = -0.39$, $P = 0.35$; duration: $Z = 0.23$, $P = 0.41$).

4. Recognition asymmetries

Taken together, these results suggest that females can readily discriminate the contact calls of their own infant from those of other infants in the group, but that they cannot do so as clearly when it comes to distress screams. This perceptual outcome for contact calls corroborates the results of our acoustic analysis which indicated significant individual differentiation in the structure of these calls and good practical success in classifying calls to the correct infant. The perceptual outcome for distress screams is more difficult to reconcile. Acoustic analysis of these calls showed substantially lower levels of individual differentiation and successful classification than were observed for contact calls; however, statistically significant individual differentiation and moderate degrees of successful call classification were nevertheless observed for screams as well.

It is, of course, possible that females in our experiment were able to discriminate screams just as well as they did the contact calls but simply chose not to do so. It is, after all, notoriously difficult to definitively establish the lack of some kind of discriminative ability based on overt behavioral responses alone. At the same time, if females could discriminate both call types equally well, why would they have chosen to do so in trials involving contact calls but chosen not to do so in trials involving screams? This kind of motivational asymmetry is equally difficult to reconcile. Both call types mediate situations of some urgency to infants. Indeed, if anything, screams mediate contexts with more immediate dire consequences for infants (e.g., injury and infanticide). Furthermore, females' behavioral responses in experimental trials demonstrated a clear interest in both call types. They responded equally often in trials involving screams and their response latencies to screams were as fast, or faster, than they were to contact calls. Hence, the lack of clear discrimination of screams compared to contact calls is not easily explained by an asymmetry in their motivation to respond to the two call types.

Nor can it easily be accounted for by some kind of response (or motivational) ceiling in our playback trials that precluded detection of mothers' actual discrimination of the calls. Although females probably could not have responded much faster in scream trials, the 10 s postplayback analysis period certainly provided ample opportunity for them to respond for longer than they did. Indeed, in contact call trials, females responded for longer, on average, than they did in scream trials and, in the course of these more protracted responses to contact calls, showed their discrimination of calls produced by their own and other infants. The same opportunity was available for showing their ability to discriminate screams.

Ultimately, it seems that despite mediating contexts of similar need for infants and despite acoustic results indicating some individual differentiation in both call types, contact calls and distress screams differ in their effective discriminability.

IV. GENERAL DISCUSSION

A. Reconciling the recognition asymmetry

The apparent maternal asymmetry for recognizing infant contact calls and distress screams contradicts our initial predictions based on the proposed functions of these calls. At the same time, the asymmetry is not without precedent. The same asymmetry was found in analogous research on closely related rhesus monkeys (Rendall *et al.*, 1996, 1998). Like baboons, rhesus also produce harmonically rich contact calls, called "coos," during group progressions through dense vegetation where visibility is limited and also when individuals ultimately become physically separated from the group (Hauser, 1991; Rendall *et al.*, 1998). These calls, like the contact calls of baboons, are proposed to function in maintaining contact among group members and facilitating reunion of separated individuals. Rhesus monkeys also produce loud broad-band scream vocalizations when physically attacked (Rowell and Hinde, 1962; Gouzoules *et al.*, 1984). These calls are structurally similar to the noisy screams produced by baboons and are proposed to function similarly in recruiting support from kin and other social allies. Previous analyses by Rendall *et al.* (1998) showed that the contact calls (coos) and distress screams produced by adult female rhesus showed some degree of acoustic differentiation between individuals, but much greater individual distinctiveness in contact calls than screams. And, in playback experiments analogous to those conducted here with baboons, adult female rhesus readily differentiated contact calls produced by close female kin from those produced by familiar but unrelated adult females in the group, but they did not make the same distinction for those females' scream vocalizations (Rendall *et al.*, 1996, 1998; see Fig. 4).

Once again, the recognition asymmetry could not be attributed either to a motivational asymmetry or to some kind of response ceiling that might have precluded detection of the animals' discrimination of screams. Furthermore, follow-up perceptual experiments were conducted to test how well human listeners could discriminate identity in these two types of rhesus monkey calls (Owren and Rendall, 2003). Like the monkeys themselves, human listeners readily discriminated between callers' contact calls but fared much worse when the discrimination involved their screams. Similar results have been reported recently by Fugate *et al.* (2008) based on short bouts of screams.

Taken together, this corpus of work involving two closely related primate species is consistent in showing some greater capacity for discriminating individual identity from

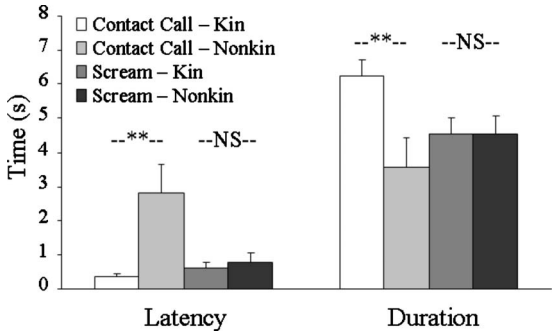


FIG. 4. Results of playback trials conducted previously on rhesus monkeys testing the ability of adult females to discriminate between the contact calls and distress screams of an adult female relative and a familiar but unrelated adult female in the group. The data plotted represent the mean (+SEM) latency and duration of orienting responses by adult females in the different experimental conditions. Females responded significantly faster and longer to the calls of their adult female relatives than they did to the calls of unrelated group females in the case of contact calls ($^{**}P<0.01$) but not screams (NS, $P>0.05$). See Rendall *et al.* (1998) for details of test statistics. Figure redrawn from Rendall *et al.* (1998).

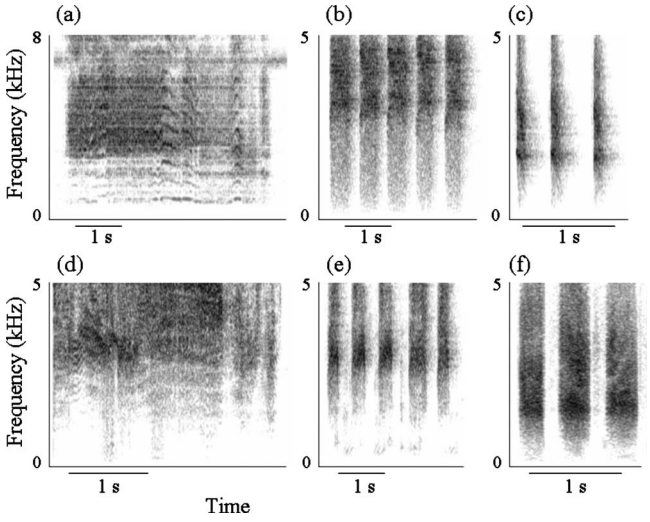


FIG. 5. Cross-species comparison of distress screams produced singly or in shout bouts of concatenated calls. Note the shared features of abrupt call onset and an extremely broad-band noisy spectral structure. The different species and contexts of call production are (a) a frog confronted by a predator, (b) a rabbit siezed by the mouth and then encircled by a constrictor snake, (c) a least weasel captured in the hand by the author (DR), and (d)–(f) infant baboons aggressively attacked by peers or adults.

species' contact calls compared to distress screams. It is important to emphasize that this conclusion does not necessarily mean that scream vocalizations are not individually distinctive, or recognizable, at any level, and hence that they cannot serve to recruit support as originally proposed. Our acoustic analyses and playback experiments, for both baboons and rhesus monkeys, focused on single vocalizations, and, in many situations, contact calls and distress screams are produced in more protracted bouts. It is, therefore, possible that some additional cues to identity might be carried in the organization of these more protracted bouts of calling (see Cheney and Seyfarth, 1980, 1999; Seyfarth and Cheney, 1984; Gouzoules *et al.*, 1986; Fugate *et al.*, 2008).

Therefore, an important caveat on the recognition asymmetry reported here, and in previous work on rhesus monkeys, is that it might apply only to single vocalizations of each type. However, that point is potentially also extremely informative. The fact that single screams are not so readily identifiable whereas single contact calls are (as are single calls of other types in these same species; Cheney and Seyfarth, 1997; Owren *et al.*, 1997; Rendall *et al.*, 1999; Fischer *et al.*, 2002) strongly suggests that some other factor might be influencing the design of screams in ways that limit or constrain salient individual differentiation at this level.

B. Alternative functions of distress screams

Previously, Owren and Rendall (1997, 2001) and Rendall and Owren (2002) suggested that victims of attack might attempt to deter aggression directly by themselves in addition to indirectly by recruiting aid from allies. They argued that, on strictly functional grounds, it would certainly be more expeditious if victims could end attacks promptly themselves rather than having to wait for help from others to arrive. By definition, though, victims' capacity to resist or repel their opponents physically might be relatively limited, which is, after all, why they are the victims rather than the aggressors to begin with. Nevertheless, they might not be entirely impotent if they can make themselves unappealing targets of continued attack. Owren and Rendall (1997, 2001) thus proposed that the loud noisy screams produced by victims of attack might be designed in part to inhibit or attenuate aggression through the calls' aversive effects on opponents.

This kind of repellent function for scream-type vocalizations has seldom been explicitly considered for primates. However, it is an acknowledged function for vocal signals in a variety of nonprimate taxa. For example, various frogs, birds, and mammals are known to produce loud distress calls when captured by a predator. These calls are variously labeled "fear," "distress," "desperation," or "death" screams and they are thought to function as a last-ditch attempt to startle or repel the predator sufficiently that it hesitates for an instant, or releases its grip marginally, or even accidentally drops its victim, allowing some opportunity for escape (Driver and Humphries, 1969; Kruuk, 1972; Schaller, 1972; Conover and Perito, 1981; Gochfeld, 1981; Conover, 1994; Wise *et al.*, 1999). Although their detailed form can vary, they are commonly described as "sudden," "powerful," "high-intensity," and "piercing" calls (see Fig. 5 for several examples).

Owren and Rendall (1997, 2001) noted that primates' distress screams share these qualities as similarly powerful high-intensity signals (often the loudest calls in the vocal repertoire) with sudden (i.e., abrupt) call onsets and often chaotic (noisy) spectral structures, which give them a jarring, plosive, and grating quality commensurate with a similar kind of repellent function. In fact, comparative research indicates that these characteristics of screams are among those that best induce the acoustic startle reflex when experienced at close range (Davis, 1984; Eaton, 1984). This reflex has been demonstrated in a wide range of animal taxa (including monkeys and apes) and is thought to occur in every hearing species. It includes a cascade of behavioral and physiological

changes in listeners that involve the immediate cessation of ongoing activity and abrupt shifts in attention toward, and in movement away from, the stimulus; a host of reflexive autonomic changes, including heart rate and blood pressure changes, increased muscle tonus and stimulation of brainstem circuits that control overall brain activity and glucose metabolism (Davis, 1984; Eaton, 1984). Together, these changes represent a broad systemic reaction readying the listener for a "fight-or-flight" response. Inducing such effects, or anything even approximating them, could be quite aversive or repellent to listeners and thus highly functional to signalers in deterring aggression.

Even if primate distress screams have not been selected specifically to induce the startle reflex *per se*, the reflex confirms the degree to which signals can contact and influence listener nervous systems and automatic response processes at foundational levels, thereby potentially returning a degree of freedom to signalers in aggressive contexts where they might otherwise be relatively impotent.

Indeed, functional effects like these might not be uncommon as demonstrated by the widespread pattern of loud crying and screaming by young infants of many species, including humans and many nonhuman primates, when denied nursing access or other physical comforts or when otherwise subdued or frustrated in their goals (e.g., Grimm, 1967; Ostwald, 1972; Lester and Boukydis, 1985; Todt, 1988; Owren *et al.*, 1992; Hammerschmidt *et al.*, 1994; Appleby *et al.*, 1999; Patel and Owren, 2007). In these situations, young individuals who are otherwise impotent to exert their will on others resort to loud vocal protests. The acoustic structures of these protests are quite consistent across taxa and involve rapidly varying combinations of loud noisy screams and piercing high-frequency tonal cries, with dramatic amplitude and frequency modulations (Owren and Rendall, 2001; Owings and Zeifman, 2004). Everyday experience and systematic perceptual studies on humans confirm that such screaming is extremely grating and aversive to listeners and very effective in catalyzing responses from them (Zeskind and Lester, 1978; Lester and Boukydis, 1985; Bradley and Lang, 2000; Owings and Zeifman 2004; Patel and Owren, 2007).

There are few perceptual tests on nonhuman primates that directly test how aversive such scream-type vocalizations might be to them. Nevertheless, the broad structural similarity of scream-type vocalizations between humans and nonhuman primates and the fundamental similarity in the auditory systems and hearing abilities of the two groups (Hienz and Brady, 1988; Sommers *et al.*, 1992; Fichtel *et al.*, 2001; Fitch and Fritz, 2006; reviewed in Ghazanfar, 2002; Heffner, 2004) together suggest that scream-type calls could be as aversive to nonhuman primates as they are to human listeners.

This proposed additional repellent function of screams thus has considerable intuitive appeal and is consistent with a variety of comparative acoustic and perceptual research. It also helps to make sense of the otherwise puzzling fact that similarly structured scream-type vocalizations occur during aggressive conflicts in a wide range of other primate and nonprimate species whose social behavior seldom involves recruiting agonistic aid from others (Owings and Morton, 1998). In fact, such screams are produced in aggressive conflicts that occur in otherwise solitary species where there is not even the potential for recruiting such support (e.g., *primates*: Aye-aye, Stanger and Macedonia, 1994; *orangutan*, Marler and Tenaza, 1977; *rodents*: Haney and Miczek, 1993; *shrews*: Andrew, 1963). Hence, inhibiting attackers might well represent the original ancestral function of screams that has been elaborated subsequently to include recruiting agonistic support in social species where that potential exists.

The additional repellent function of screams compared to contact calls might also help to account for the observed recognition asymmetry in the two call types if the design requirements associated with one function for screams influence or constrain the design requirements of the other. Such signal-design tradeoffs in the face of competing signal functions are certainly well known from other species and from multiple signaling modalities, particularly in contexts where the signals involved can have multiple potential audiences (e.g., Cade, 1975; Tuttle and Ryan, 1981; Endler, 1988). In the case of distress screams then, it is possible that the noisy spectral structure of these calls makes them functionally aversive to opponents close at hand specifically because of their unpatterned quality, and this in turn precludes, by definition, any consistent patterning associated with carrying reliable identity cues to distant recruits. This possibility remains speculative, but it is an important signal-design issue to resolve. If there is not some additional function like this at work for screams, and vocalizing when attacked is therefore only an attempt to recruit support from distant others by conveying to them clear cues to caller identity, then why, in fact, do infants (and adults) actually scream in these circumstances? Why do they not instead produce other calls which are demonstrably more individually distinctive and recognized by others? For example, contact calls.

Ultimately, then, we tentatively suggest that individual contact calls and distress screams differ in their effective discriminability because they vary in the range of functions they have been selectively designed for. Agonistic screams may serve a repellent function in addition to a recruitment function, and the balance of these functional requirements appears to emphasize aversiveness at the level of individual scream calls and identity cueing potentially in features that emerge from the organization of more protracted sequences of calling. Contact calls appear not to be similarly constrained; hence, their structure is comparatively freed to promote individual differentiation and recognition of single vocalizations.

ACKNOWLEDGMENTS

Thanks are due to the Office of the President and the Department of Wildlife and National Parks of the Republic of Botswana for permission to conduct research in the Moremi Game Reserve and to Robert Seyfarth and Dorothy Cheney for the opportunity to work with baboons at their field site. Thanks also to Robert Seyfarth and Sergio Pellis for valuable comments on an early version of this paper.

Special thanks to Karen Rendall for her help with all aspects of fieldwork. Research was supported in part by NIMH Grant No. MH62249 to D. L. Cheney and R. M. Seyfarth and by grants from the Natural Sciences and Engineering Research Council of Canada, the L.S.B. Leakey Foundation, and the University of Lethbridge to D. Rendall. The research was reviewed and approved by the Animal Care and Use Committee at the University of Pennsylvania.

Altmann, J. (**1980**). *Baboon Mothers and Infants* (Harvard University Press, Cambridge, MA).

Andrew, R. J. (**1963**). "The origin and evolution of the calls and facial expressions of the primates," Behaviour **20**, 1–109.

Appleby, M. C., Weary, D. M., Taylor, A. A., and Illmann, G. (**1999**). "Vocal communication in pigs: Who are nursing piglets screaming at?," Ethology **105**, 881–892.

Boersma, P. (**2001**). "Praat, a system for doing phonetics by computer," Glot Internat. **5**, 341–345.

Bradley, M. M., and Lang, P. J. (**2000**). "Affective reactions to acoustic stimuli," Psychophysiology **37**, 204–215.

Cade, W. (**1975**). "Acoustically orienting parasitoids: Fly phonotaxis to cricket song," Science **190**, 1312–1313.

Chapais, B., and Berman, C. (**2004**). *Kinship and Behavior in Primates* (Oxford University Press, Oxford).

Cheney, D. L., and Seyfarth, R. M. (**1980**). "Vocal recognition in free-ranging vervet monkeys," Anim. Behav. **28**, 362–367.

Cheney, D. L., and Seyfarth, R. M. (**1997**). "Reconciliatory grunts by dominant female baboons influence victim's behaviour," Anim. Behav. **54**, 409–418.

Cheney, D. L., and Seyfarth, R. M. (**1999**). "Recognition of other individuals' social relationships by female baboons," Anim. Behav. **58**, 67–75.

Cheney, D. L., and Seyfarth, R. M. (**2007**). *Baboon Metaphysics: The Evolution of a Social Mind* (University of Chicago Press, Chicago).

Cheney, D. L., Seyfarth, R. M., and Palombit, R. A. (**1996**). "The function and mechanisms underlying baboon 'contact' barks," Anim. Behav. **52**, 507–518.

Conover, M. R. (**1994**). "Stimuli eliciting distress calls in adult passerines and response of predators and birds to their broadcast," Behaviour **131**, 19–37.

Conover, M. R., and Perrito, J. J. (**1981**). "Response of starlings to distress calls and predator models holding conspecific prey," Ethology **57**, 163–172.

Davis, M. (**1984**). "The mammalian startle response," in *Neural Mechanisms of Startle Behavior*, edited by R. C. Eaton (Plenum, New York), pp. 287–351.

Driver, P. M., and Humphries, D. A. (**1969**). "The significance of the high-intensity alarm call in captured passerines," Ibis **111**, 243–244.

Eaton, R. C. (**1984**). *Neural Mechanisms of Startle Behavior* (Plenum, New York).

Endler, J. A. (**1988**). "Sexual selection and predation risk in guppies," Nature (London) **332**, 593–594.

Engh, A. E., Hoffmeier, R. R., Cheney, D. L., and Seyfarth, R. M. (**2006**). "Who, me? Can baboons infer the target of vocalisations?," Anim. Behav. **71**, 381–387.

Fichtel, C., Hammerschmidt, K., and Jurgens, U. (**2001**). "On the vocal expression of emotion: A multi-parametric analysis of different states of aversion in the squirrel monkey," Behaviour **138**, 97–116.

Fischer, J. (**2004**). "Emergence of individual recognition in young macaques," Anim. Behav. **67**, 655–661.

Fischer, J., Hammerschmidt, K., Cheney, D. L., and Seyfarth, R. M. (**2001**). "Acoustic features of female chacma baboon barks," Ethology **107**, 33–54.

Fischer, J., Hammerschmidt, K., Cheney, D. L., and Seyfarth, R. M. (**2002**). "Acoustic features of male baboon loud calls: Influences of context, age, and individuality," J. Acoust. Soc. Am. **111**, 1465–1474.

Fitch, W. T., and Fritz, J. (**2006**). "Rhesus macaques spontaneously perceive formants in conspecific vocalizations," J. Acoust. Soc. Am. **120**, 2132–2141.

Fitch, W. T., Neubauer, J., and Herzel, H. (**2002**). "Calls out of chaos: The adaptive significance of nonlinear phenomena in mammalian vocal production," Anim. Behav. **63**, 407–418.

Fugate, J. M. B., Gouzoules, H., and Nygaard, L. C. (**2008**). "Recognition of rhesus macaque (*Macaca mulatta*) noisy screams: Evidence from conspecific and human listeners," Am. J. Primatol **70**, 594–604.

Gautier, J.-P., and Gautier, A. (**1977**). "Communication in Old World monkeys," in *How Animals Communicate*, edited by T. A. Sebeok (Indiana University Press, Bloomington, IN), pp. 890–964.

Ghazanfar, A. A. (**2002**). *Primate Audition: Ethology and Neurobiology* (CRC, Boca Raton, FL).

Ghazanfar, A. A., and Santos, L. (**2004**). "Primate brains in the wild: The sensory bases for social interactions," Nat. Rev. Neurosci. **5**, 603–616.

Ghazanfar, A. A., Turesson, H. K., Maier, J. X., van Dinther, R., Patterson, R. D., and Logothetis, N. K. (**2007**). "Vocal-tract resonances as indexical cues in rhesus monkeys," Curr. Biol. **17**, 425–430.

Gochfeld, M. (**1981**). "Responses of young black skimmers to high-intensity distress notes," Anim. Behav. **29**, 1137–1145.

Gouzoules, H., and Gouzoules, S. (**1990**). "Matrilineal signatures in the recruitment screams of pigtail macaques, *Macaca nemestrina*," Behaviour **115**, 327–347.

Gouzoules, H., and Gouzoules, S. (**2000**). "Agonistic screams differ among four species of macaques: The significance of motivation-structural rules," Anim. Behav. **59**, 501–512.

Gouzoules, S., Gouzoules, H., and Marler, P. (**1984**). "Rhesus monkey (*Macaca mulatta*) screams: Representational signaling in the recruitment of agonistic aid," Anim. Behav. **32**, 182–193.

Gouzoules, H., Gouzoules, S., and Marler, P. (**1986**). "Vocal communication: A vehicle for the study of social relationships," in *The Cayo Santiago Macaques: History, Behavior and Biology*, edited by R. G. Rawlins and M. J. Kessler (State University of New York Press, Albany, NY), pp. 111–129.

Grimm, R. J. (**1967**). "Catalogue of sounds of the pigtailed macaque," J. Zool. Soc. London **152**, 361–373.

Hammerschmidt, K., Ansorge, V., Fischer, J., and Todt, D. (**1994**). "Dusk calling in Barbary macaques (*Macaca sylvanus*): Demand for social shelter," Am. J. Primatol **32**, 277–289.

Hammerschmidt, K., and Fisher, J. (**1998**). "Maternal discrimination of offspring vocalizations in Barbary macaques (*Macaca sylvanus*)," Primates **39**, 231–236.

Haney, M., and Miczek, K. A. (**1993**). "Ultrasounds during agonistic interactions between female rats (*Rattus norvegicus*)," J. Comp. Psychol. **107**, 373–379.

Hansen, E. W. (**1976**). "Selective responding by recently separated juvenile rhesus monkeys to the calls of their mother," Dev. Psychobiol. **9**, 83–88.

Hauser, M. D. (**1991**). "Sources of acoustic variation in rhesus macaque (*Macaca mulatta*) vocalizations," Ethology **89**, 29–46.

Hauser, M. D. (**1992**). "Articulatory and social factors influence the acoustic structure of rhesus monkey vocalizations: A learned mode of production?," J. Acoust. Soc. Am. **91**, 2175–2179.

Heffner, R. S. (**2004**). "Primate hearing from a mammalian perspective," Anat. Rec. Part A **281A**, 1111–1122.

Hienz, R. D., and Brady, V. (**1988**). "The acquisition of vowel discrimination by nonhuman primates," J. Acoust. Soc. Am. **84**, 186–194.

Hintze, J. L. (**1999**). *Number Cruncher Statistical System, 2000* (Kaysville, UT).

Jovanic, T., Megna, N. L., and Maestripieri, D. (**2000**). "Early maternal recognition of offspring vocals in rhesus," Primates **41**, 421–428.

Kaplan, J. N., Winship-Ball, A., and Sim, L. (**1978**). "Maternal discrimination of infant vocalizations in squirrel monkeys," Primates **19**, 187–193.

Klecka, W. R. (**1980**). *Discriminant Analysis*, Quantitative Applications in the Social Sciences (Sage, London).

Kruuk, H. (**1972**). *The Spotted Hyena* (University of Chicago Press, Chicago, IL).

Lester, B. M., and Boukydis, C. F. Z. (**1985**). *Infant Crying: Theoretical and Research Perspectives* (Plenum, New York).

Lieberman, P., Klatt, D. H., and Wilson, W. H. (**1969**). "Vocal tract limitations on the vowel repertoires of rhesus monkeys and other nonhuman primates," Science **164**, 1185–1187.

Macedonia, J. M. (**1986**). "Individuality in a contact call of the Ringtailed lemur (*Lemur catta)*," Am. J. Primatol **11**, 163–179.

Marler, P., and Hobbett, L. (**1975**). "Individuality in a long-range vocalization of wild chimpanzees," Z. Tierpsychol **38**, 97–109.

Marler, P., and Tenaza, R. (**1977**). "Signaling behavior of apes with special reference to vocalization," in *How Animals Communicate*, edited by T. A. Sebeok (Indian University Press, Bloomington, IN), pp. 965–1033.

Nunn, C. L. (**2000**). "Maternal recognition of infant calls in ring-tailed lemurs," Folia Primatol. **71**, 142–146.

Oppenheimer, J. R. (**1977**). "Communication in New World monkeys," in

How Animals Communicate, edited by T. A. Sebeok (Indian University Press, Bloomington, IN), pp. 851–889.
Ostwald, P. F. (**1972**). "The sounds of infancy," Dev. Med. Child Neurol. **14**, 350–361.
Owings, D., and Morton, E. S. (**1998**). *Animal Vocal Communication: A New Approach* (Cambridge University Press, Cambridge, MA).
Owings, D., and Zeifman, D. M. (**2004**). "Human infant crying as an animal communication system: Insights from an assessment-management approach," in *Evolution of Communication Systems: A Comparative Approach*, edited by D. K. Oller and U. Griebel (MIT Press, Cambridge, MA), pp. 151–170.
Owren, M. J., Dieter, J. A., Seyfarth, R. M., and Cheney, D. L. (**1992**). "'Food' calls produced by adult female rhesus (*Macaca mulatta*) and Japanese (*Macaca fuscata*) macaques, their normally raised offspring, and offspring cross-fostered between species," Behaviour **120**, 218–231.
Owren, M. J., and Rendall, D. (**1997**). "An affect-conditioning model of nonhuman primate vocal signaling," in *Perspectives in Ethology*, 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: Returning form-function relations to the forefront of acoustic primatology," Evol. Anthropol. **10**, 58–71.
Owren, M. J., and Rendall, D. (**2003**). "Salience of caller identity in rhesus monkey (*Macaca mulatta*) coos and screams: Perceptual experiments with human listeners," J. Comp. Psychol. **117**, 380–390.
Owren, M. J., Seyfarth, R. M., and Cheney, D. L. (**1997**). "The acoustic features of vowel-like grunt calls in chacma baboons (*Papio cynocephalus ursinus*): Implications for production processes and functions," J. Acoust. Soc. Am. **101**, 2951–2963.
Palombit, R., Cheney, D., Seyfarth, R., Rendall, D., Silk, J., Johnson, S., and Fischer, J. (**2000**). "Male infanticide and defense of infants in chacma baboons," in *Infanticide by Males and Its Implications*, edited by C. van Schaik and C. Janson (Cambridge University Press, Cambridge), pp. 123–152.
Palombit, R. A., Seyfarth, R. M., and Cheney, D. L. (**1997**). "The adaptive value of 'friendships' to female baboons: Experimental and observational evidence," Anim. Behav. **54**, 599–614.
Patel, E. R., and Owren, M. J. (**2007**). "Acoustics and behavioral contexts of 'gecker' vocalizations in young rhesus macaques (*Macaca mulatta*)," J. Acoust. Soc. Am. **121**, 575–585.
Pereira, M. E. (**1986**). "Maternal recognition of juvenile offspring coo vocalizations in Japanese macaques," Anim. Behav. **34**, 935–937.
Rendall, D. (**2003**). "Acoustic correlates of caller identity and affect intensity in the vowel-like grunt vocalizations of baboons," J. Acoust. Soc. Am. **113**, 3390–3402.
Rendall, D. (**2004**). "'Recognizing' kin: Mechanisms, media, minds, modules and muddles," in *Kinship and Behaviour in Primates*, edited by B. Chapais and C. Berman (Oxford University Press, Oxford), pp. 295–316.
Rendall, D., Cheney, D. L., and Seyfarth, R. M. (**2000**). "Proximate factors mediating 'contact' calls in adult female baboons and their infants," J. Comp. Psychol. **114**, 36–46.
Rendall, D., and Owren, M. J. (**2002**). "Animal vocal communication: Say what?," in *The Cognitive Animal: Empirical and Theoretical Perspectives on Animal Cognition*, edited by M. Bekoff, C. Allen, and G. Burghardt (MIT Press, Cambridge, MA), pp. 307–313.
Rendall, D., Owren, M. J., and Rodman, P. S. (**1998**). "The role of vocal tract filtering in identity cueing in rhesus monkey (*Macaca mulatta*) vocalizations," J. Acoust. Soc. Am. **103**, 602–614.
Rendall, D., Owren, M. J., Weerts, E., and Hienz, R. J. (**2004**). "Sex differences in the acoustic structure of vowel-like grunt vocalizations in baboons and their perceptual discrimination by baboon listeners," J. Acoust. Soc. Am. **115**, 411–421.
Rendall, D., Rodman, P. S., and Emond, R. E. (**1996**). "Vocal recognition of individuals and kin in free-ranging rhesus monkeys," Anim. Behav. **51**, 1007–1015.
Rendall, D., Seyfarth, R. M., Cheney, D. L., and Owren, M. J. (**1999**). "The meaning and function of grunt variants in baboons," Anim. Behav. **57**, 583–592.
Riede, T., Arcadi, A. C., and Owren, M. J. (**2004**). "Nonlinear acoustics in pant hoots of common chimpanzees (*Pan troglodytes*): Frequency jumps, subharmonics, biphonation, and deterministic chaos," Am. J. Primatol **64**, 277–291.
Riede, T., Arcadi, A. C., and Owren, M. J. (**2007**). "Nonlinear acoustics in the pant hoots of common chimpanzees (*Pan troglodytes*): Vocalizing at the edge," J. Acoust. Soc. Am. **121**, 1758–1767.
Riede, T., and Zuberbühler, K. (**2003**). "The relationship between acoustic structure and semantic information in Diana monkey alarm vocalization," J. Acoust. Soc. Am. **114**, 1132–1142.
Rowell, T. E., and Hinde, R. A. (**1962**). "Vocal communication by the rhesus monkey (*Macaca mulatta*)," Proc. Zool. Soc. Lond. **138**, 279–294.
Schaller, G. B. (**1972**). *The Serengeti Lion* (University of Chicago Press, Chicago, IL).
Semple, S. (**2001**). "Individuality and male discrimination of female copulation calls in the yellow baboon," Anim. Behav. **61**, 1023–1028.
Seyfarth, R. M., and Cheney, D. L. (**1984**). "Grooming, alliances and reciprocal altruism in vervet monkeys," Nature (London) **308**, 541–543.
Smuts, B. B. (**1985**). *Sex and Friendship in Baboons* (Harvard University Press, Cambridge, MA).
Snowdon, C. T., and Cleveland, J. (**1980**). "Individual recognition of contact calls by pygmy marmosets," Anim. Behav. **28**, 717–727.
Sommers, M. S., Moody, D. B., Prosen, C. A., and Stebbins, W. C. (**1992**). "Formant frequency discrimination by Japanese macaques (*Macaca fuscata*)," J. Acoust. Soc. Am. **91**, 3499–3510.
Stanger, K. F., and Macedonia, J. M. (**1994**). "Vocalizations of aye-ayes (*Daubentonia madagascariensis*) in captivity," Folia Primatol. **62**, 160–169.
Strum, S. C. (**1987**). *Almost Human: A Journey Into the World of Baboons* (University of Chicago Press, Chicago, IL).
Symmes, D., and Biben, M. (**1985**). "Maternal recognition of individual infant squirrel monkeys from isolation call playbacks," Am. J. Primatol **9**, 39–46.
Todt, D. (**1988**). "Serial calling as a mediator of interactions processes: Crying in primates," in *Primate Vocal Communication*, edited by D. Todt, P. Godeking, 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.
Tuttle, M. D., and Ryan, M. J. (**1981**). "Bat predation and the evolution of frog vocalizations in the Neotropics," Science **214**, 677–678.
Waser, P. M. (**1977**). "Individual recognition, intragroup cohesion and intergroup spacing: Evidence from sound playback to forest monkeys," Behaviour **60**, 28–74.
Widdig, A. (**2007**). "Paternal kin discrimination: The evidence and likely mechanisms," Biol. Rev. **82**, 319–334.
Wise, K. K., Connover, M. R., and Knowlton, F. F. (**1999**). "Response of coyote to avian distress calls: Testing the startle-predator and predator-attraction hypotheses," Behaviour **136**, 935–949.
Zeskind, P. S., and Lester, B. (**1978**). "Acoustic features and auditory perceptions of the cries of newborns with prenatal and perinatal complications," Child Dev. **49**, 580–589.

Manatee (*Trichechus manatus*) vocalization usage in relation to environmental noise levels

Jennifer L. Miksis-Olds[a)]
Applied Research Laboratory, The Pennsylvania State University, P.O. Box 30, State College, Pennsylvania 16804

Peter L. Tyack
Department of Biology, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543

(Received 5 August 2008; revised 8 December 2008; accepted 12 December 2008)

Noise can interfere with acoustic communication by masking signals that contain biologically important information. Communication theory recognizes several ways a sender can modify its acoustic signal to compensate for noise, including increasing the source level of a signal, its repetition, its duration, shifting frequency outside that of the noise band, or shifting the timing of signal emission outside of noise periods. The extent to which animals would be expected to use these compensation mechanisms depends on the benefit of successful communication, risk of failure, and the cost of compensation. Here we study whether a coastal marine mammal, the manatee, can modify vocalizations as a function of behavioral context and ambient noise level. To investigate whether and how manatees modify their vocalizations, natural vocalization usage and structure were examined in terms of vocalization rate, duration, frequency, and source level. Vocalizations were classified into two call types, chirps and squeaks, which were analyzed independently. In conditions of elevated noise levels, call rates decreased during feeding and social behaviors, and the duration of each call type was differently influenced by the presence of calves. These results suggest that ambient noise levels do have a detectable effect on manatee communication and that manatees modify their vocalizations as a function of noise in specific behavioral contexts. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3068455]

PACS number(s): 43.80.Nd [WWA] Pages: 1806–1815

I. INTRODUCTION

Environmental noise has the potential to interfere with acoustic communication by masking signals containing biologically important information. Communication theory recognizes several ways a sender can modify its acoustic signal to compensate for noise. These include increasing the source level of a signal (Brumm, 2004; Scheifele *et al.*, 2005), its repetition (Penna *et al.*, 2005), its duration (Foote *et al.*, 2004; Miller *et al.*, 2000), shifting frequency outside of the noise band (Lesage *et al.*, 1999; Brumm and Slabbekoorn, 2005; Parks *et al.*, 2007), and altering the timing of signals to correspond with periods of less noise (Brumm and Slabbekoorn, 2005; Sun and Narins, 2005; Parks *et al.*, 2007; Sousa-Lima and Clark, 2008). All of these phenomena have been demonstrated in birds exposed to noise from natural sources (Lengagne *et al.*, 1999; Slabbekoorn and Peet, 2003; Brumm, 2004; Brumm and Slater, 2006) and to marine mammals exposed to anthropogenic noise (Lesage *et al.*, 1999; Miller *et al.*, 2000; Parks, 2003; Foote *et al.*, 2004; Scheifele *et al.*, 2005; Parks *et al.*, 2007).

The effective range of acoustic communication (or acoustic active space) in the marine environment depends on the acoustic propagation loss characteristics of the area, the frequency and amplitude of the vocalizations being emitted, the hearing sensitivity of the animal, and the ambient noise. The importance of maintaining contact, and the required range of communication, depends on behavioral context. For example, it may be critical for a mother to maintain contact with her young, but this may be over a relatively short range if mother and young do not separate far. Members of a loose group of adults might separate over greater distances, but missed reunion cues may be less costly compared to interactions between mothers and calves.

The effective range of vocalizations may also be affected by environmental noise level. If the frequency of a vocalization overlaps with that of noise, this may mask significant signal information within the call, consequently interfering with communication. However, compensation mechanisms used to reduce the effect of noise likely involve a cost. Calls with a high probability of detection may be intercepted by predators, although this is not a significant cost to manatees because they have no natural nonhuman predators. Increased harassment of females by roving males may be another potential cost resulting from high call rates between mothers and calves (Hartman, 1979; Bengston, 1981; O'Shea and Hartley, 1995; Sousa-Lima *et al.*, 2008). Louder and longer calls take more energy to produce, and if producing a call out of the normal frequency band is less efficient for the sound production apparatus, this may take more energy as well (Ryan, 1986). This balance of cost and benefit suggests that animals are likely to modulate signaling behavior to maintain effective communication out to the likely range of expected receivers, but that they may not

[a)]Electronic mail: jlm91@psu.edu

always signal to achieve the maximum possible range. The extent to which animals engage costly mechanisms to compensate for noise may also depend on the cost of failing to communicate (e.g., when a mother and dependent young are separated). Thus signaling behavior including compensation for noise should depend on the behavioral context, the cost of failure to communicate, and the distribution of receivers.

Vocalizations are assumed to form the basis of most long range communication in marine mammals, as sound has the potential to travel rapidly over long distances in water, providing a reliable medium for manatees to communicate beyond visual range in murky coastal or riverine waters (Sousa-Lima *et al.*, 2002). The impact of noise may become even more crucial for manatee acoustic communication during the summer months when vision is reduced due to high turbidity. Previous studies describing West Indian manatee (*Trichechus manatus*) vocalizations allude to the presence of two different vocalization types: tonal harmonic calls and broader-band, less tonal calls (Schevill and Watkins, 1965; Bengtson and Fitzgerald, 1985; Nowacek *et al.*, 2003; Sousa-Lima *et al.*, 2008). However, this differentiation has not been quantified, and authors either analyzed only one call type or combined all call types in their rate and source level analyses. The present study quantitatively classified the vocalizations into two separate categories based on call structure. Analyses of usage were then performed separately for each call type. The identification of two acoustically distinct categories of manatee vocalizations may provide some insight into their function. The presence of individually stereotyped vocalizations in the repertoire of captive Amazonian and Antillean manatees provides evidence in support of the idea that manatee vocalizations function to establish and maintain contact between individuals (Sousa-Lima *et al.*, 2002, 2008).

The goal of this paper is to study how manatees vary their vocal output to compensate for noise in different kinds of social groups and in different behavioral contexts. This species inhabits the shallow coastal waters near Sarasota, FL year round, although greatest numbers are observed in the nonwinter months when the water is most turbid. The grass-bed habitats and shallow dredged basins that manatees utilize are environments where low frequency sound does not propagate as far as in the deep ocean (Miksis-Olds and Miller, 2006; Urick, 1983). The frequencies that propagate best are between 1 and 2 kHz (Miksis-Olds and Miller, 2006), a band which coincides with the dominant frequencies of manatee vocalizations (Nowacek *et al.*, 2003). Thousands of boats cause considerable variation in noise levels, which vary systematically on an hourly basis during the day and on a weekly basis including weekends and weekdays. Miksis-Olds *et al.* (2007) demonstrated that this variation affects how manatees utilize specific habitats. Here we study whether manatees alter their vocalizations based on ambient noise levels, and whether they show context dependent compensation for noise in their extremely shallow coastal environment.

II. METHODS

A. Acoustic recordings

Vocalizations and ambient noise were recorded continuously throughout all group sightings and individual focal follows from a towed hydrophone system. The observation vessel maintained a 25–50 m distance to the individual or group of animals being recorded. The recording system contained an HTI-99-HF hydrophone with built-in preamplifier. This hydrophone had an operational frequency range of 2–125 000 Hz and a sensitivity of −178 dB re 1 mV/Pa. The hydrophone was calibrated prior to and following each field season via comparison to a reference transducer. The analog signal was transferred to a Dell Inspiron 8110 laptop computer via a National Instruments PCMCIA DAQ Card-6062E with 12 bit resolution. The input range of the analog-to-digital converter (ADC) was set to ±1 V for all recordings. The 12 bit data were then converted and stored at 16 bit resolution using the CHICKADEE multichannel recorder version 1.9b software program. The system noise was dominated by discretization noise (Oppenheim *et al.*, 1999) and was well below the minimum detectable signal. Only vocalizations emitted at times when the trolling motor on the observation vessel was off were used in analyses. This was done in order to eliminate any potential response of the animals to the sound of the motor that was used to follow them during observation periods. Vocalization analysis was not designed to account for potential vocalization responses to transitions of the trolling motor being turned on and off during the focal follow.

Broadband recordings of ambient noise were made at the beginning and end of each sighting and/or follow when all motors were off and the boat was drifting. The sampling rate was 200 kHz. Noise levels were calculated in three 1/3 octave bands with center frequencies of 500 Hz, 4 kHz, and 32 kHz. System noise of the recording system was below all ambient noise levels at all 1/3 octave bands. Three 1/3 octave band categories were selected to represent low, mid, and high frequency noises (LFN, MFN, and HFN, respectively) in manatee habitats. From the perspective of the manatee, the noise categories were designed to correspond to specific environmental signals and hearing capabilities. The LFN band (500 Hz) was selected to match the dominant frequencies of watercraft and other human activities that have the potential to cause bodily injury and mask vocalizations. MFN (4 kHz) directly overlapped with the dominant frequency of manatee vocalizations and had a high potential for masking. Finally the HFN band (32 kHz) was reported to be outside the most sensitive areas of hearing for manatees (Ketten *et al.*, 1992; Gerstein *et al.*, 1999). A detailed description of the noise components and spectral content of manatee habitats in Sarasota Bay can be found in Miksis-Olds *et al.* (2007).

B. Behavioral states

Manatees were observed to engage in five behavioral states throughout the study: social, mill, travel, rest, and feed. Social behaviors were defined as the direct interaction between one or more manatees and include mating, herding, and playing. Milling behavior was characterized by undi-

rected movement, and traveling was characterized by directed movement. Resting behavior was identified by repeated surfacing in the same location in the absence of feeding. Feeding was identified by the observation of food ingestion evident by grass in mouth or cropped grasses floating in the vicinity.

Analysis of vocalizations included only those produced when an entire group of manatees under observation was engaged in the same behavior or only a single animal was being observed. This is consistent with the methodology of Bengtson and Fitzgerald (1985). An additional criterion for inclusion in the behavior analysis was that the recording lasted a minimum of 20 min. Ten of the 99 recordings were shorter than 20 min and were eliminated from analysis. The 20 min time period was selected because it was the minimum amount of time needed to approach a group and accurately determine behavioral state, number of animals, and calf presence. Estimating the number of manatees and behavior classification are both potential sources of error in the analyses. The number and behavior of manatees may have been misidentified in turbid waters although the number of animals obtained from the observation vessel compared well with simultaneous aerial survey counts from a low flying aircraft. No discrepancies occurred for groups of five animals or fewer. In 13 comparisons, the maximum difference between the aerial survey and on-water counts was two animals for groups with a minimum of six individuals and ranging from 6 to 20 total animals. The on-water counts underestimated the number of manatees present in all comparisons. Another possible source of error was that animals nearby or approaching the observed group could have introduced vocalizations while in a different behavioral state.

C. Vocalization rate

The number of vocalizations used from each recording was determined according to the following protocol. The protocol was implemented to account for the context dependent usage of vocalizations during different behavioral states. Recordings made when animals were continuously observed engaging in the same behavior were divided into 20 min time periods; a 2 h recording made while animals were resting would contain six nonoverlapping time periods. The 20 min block selected to initiate the vocalization rate analysis for each behavioral state was randomly identified to prevent bias of rate changes associated with behavioral transitions (such as always starting at the beginning of an observed behavior). Vocalizations continued to be identified in 20 min time periods within the same behavioral state until a total of 200 vocalizations were reached, the behavioral state changed, or the recording ended. Because of the great number of vocalizations produced during social interactions, a separate criterion was applied to this state. For social behaviors only, if more than 200 vocalizations were obtained within the first randomly selected 20 min time period, no further vocalizations were identified from that behavioral state on a single recording. Vocalization rates were calculated in 20 min blocks and were presented as the number of vocalizations per manatee per 5 min period to maintain consistency with previous studies (Bengtson and Fitzgerald, 1985; Phillips *et al.*, 2004). Mean vocalization rate was calculated from multiple 20 min time periods within the same behavioral state from the same recording and averaged to produce a single rate value per manatee for each behavioral state within the recording; hence the recording was the unit of analysis.

TABLE I. Kolmogorov–Smirnov normality test statistics.

Category	Covariate parameter	*D* value	*p* value
	Average rate	0.092	>0.15
	LFN	0.074	>0.15
	MFN	0.104	0.06
	HFN	0.091	>0.15
	Group size	0.120	<0.01
Chirp	Average source level	0.110	0.08
	Average duration	0.089	>0.15
	Average freq. range	0.090	>0.15
	Average min. freq.	0.095	>0.15
	Average max. freq.	0.110	0.06
	Average peak freq.	0.080	>0.15
	LFN	0.080	>0.15
	MFN	0.100	0.1
	HFN	0.110	0.06
	Group size	0.130	0.01
Squeak	Average source level	0.110	0.13
	Average duration	0.120	0.1
	Average freq. range	0.093	>0.15
	Average min. freq.	0.094	>0.15
	Average max. freq.	0.120	0.06
	Average peak freq.	0.090	>0.15
	LFN	0.110	>0.15
	MFN	0.120	0.1
	HFN	0.070	>0.15
	Group size	0.150	<0.01

D. Vocalization structure

Vocalizations included in the analysis of vocalization structure were subjected to the same selection criteria as those for the vocalization rate analysis. Subsets of these vocalizations were selected for structure analysis based on the following additional criteria: (1) a clear start and end to the vocalization was identified, (2) there were no overlapping signals, and (3) the received level (rms) was a minimum of 3 dB above the noise level over the same bandwidth. Parameters measured were number of inflection points, duration, minimum frequency, maximum frequency, frequency range, and frequency of peak energy. All parameters were measured by hand from a 512 point Fast Fourier Transform (FFT) spectrogram. Absolute level of the signal was not used as part of the structure analysis.

E. Source level

Vocalizations included in the source level analysis were subjected to the same selection criteria as those for the vocalization rate analysis. A subset of these vocalizations was selected for source level analysis based on the following ad-

TABLE II. Effects tested in initial mixed model ANCOVA.

Effect		Variable	Categories
Behavior	(main effect)	Categorical	Social, mill, feed, rest
Calf presence	(main effect)	Categorical	Present, absent
Group size	(main effect)	Categorical	A=1, B=2, C=3–4, D=5–6, E=7 or more
LFN	(main effect)	Continuous	
MFN	(main effect)	Continuous	
HFN	(main effect)	Continuous	
Behavior × calf presence			
Behavior × group size			
Calf presence × group size			
Behavior × LFN			
Behavior × MFN			
Behavior × HFN			
Calf presence × LFN			
Calf presence × MFN			
Calf presence × HFN			
Group size × LFN			
Group size × MFN			
Group size × HFN			
Calf presence × behavior × group size			
Calf presence × behavior × LFN			
Calf presence × behavior × MFN			
Calf presence × behavior × HFN			
Calf presence × group size × LFN			
Calf presence × group size × MFN			
Calf presence × group size × HFN			
Behavior × group size × LFN			
Behavior × group size × MFN			
Behavior × group size × HFN			

ditional criteria: (1) distance from a single animal or compact group of animals to the hydrophone was specified, (2) no other animals were observed within 500 m, (3) a clear start and end to the vocalization was identified, and (4) there were no overlapping signals. Laser range finding binoculars were used to obtain the distance to the most likely vocalizing animal during surfacings. A compact group was defined as a group of animals (mother-calf pair, social group, etc.) within a 5 m radius which translated into a 5 m error in source level calculations. A 500 m distance was selected for criterion 2, as sound energy from vocalizing animals beyond 500 m would not be audible above the background noise (Miksis-Olds and Miller, 2006; Miksis-Olds *et al.*, 2007).

A fifth order Butterworth band pass filter was constructed and applied to each vocalization over the band of minimum and maximum signal frequencies selected by hand. A rms received level was then calculated over the duration of the vocalization. Final source level calculations were obtained by adding a transmission loss component which was specific to each animal distance and geographical site. Average transmission loss values generated from the Monterey–Miami parabolic equation model (Smith, 2001) within each manatee habitat were obtained from Miksis-Olds and Miller (2006).

F. Statistical analysis

Kolmogorov–Smirnov normality tests were performed for all vocal response parameters (rate, source level, frequency range, minimum frequency, maximum frequency, and peak frequency of the dominant harmonic) and continuous covariates (group size, LFN, MFN, and HFN) (Table I). Group size was the only parameter shown not to be normally distributed. As a result, group size was converted to a categorical variable with five categories: (1) single animal, (2) pair of animals, (3) three to four animals, (4) five to six animals, and (5) seven or more animals.

A mixed model analysis of covariance (ANCOVA) was run to determine if noise level had an effect on vocal parameters. The unit of analysis was the recording. Manatees are not reported to be territorial, so recordings of vocalizations made at the same site on different days were highly unlikely to be from the same individuals. This was confirmed by photographs taken of each animal and compared to known animals in the Manatee Individual Photoidentification System (MIPS) maintained by the State of Florida. One recording per day was made at a site, and subsequent recordings on the same day were geographically separated by a minimum of 1 km. Photoidentification was used to ensure that the same manatee was not recorded more than once on the same day. The probability of repeated recording of the same individual

TABLE III. Noise levels for each behavior category in decibels re 1 μPa. Min is the minimum level recorded during the designated behavior. Max is the maximum level recorded during the designated behavior. Average is the noise level over the range of the Q2 and Q3 quartiles.

	Feed			Mill			Rest			Social		
	Min	Average	Max	Min	Average	Max	Min	Average	Max	Min	Average	Max
LFN	47	56–68	80	38	51–59	67	38	47–53	67	47	51–65	70
MFN	45	52–66	79	43	48–56	65	40	48–52	64	46	49–58	68
HFM	46	54–66	80	45	53–64	65	45	54–58	66	44	53–59	68

was thus assumed to be low. Over the course of the 2 year study, only six animals were observed to be recorded more than once. One manatee was the focal animal in two follows, and five manatees were observed in different sightings on different days over the course of 1 year.

A separate model was run for each call parameter in the two call categories and included three categorical factors (behavior, calf presence, and group size) and three covariates (LFN, MFN, and HFN) (Table II) (Littell *et al.*, 2006). The initial mixed model included 6 main effects, 12 two-factor interactions, and 10 three-factor interactions (Table II). Interactions between noise levels at different frequencies were not the focus of this study, so only one noise variable was included in each model interaction. The data set did not support four-factor interactions due to the low sample size in each statistical cell. The initial mixed model was then refined by a three-step elimination of nonsignificant factors ($p > 0.60$ for three-factor interactions and $p > 0.80$ for two-factor interactions). Interactions exceeding the designated significance thresholds were removed prior to the next mixed model run. Eliminations occurred three times, and the models were run four times to identify the final significant interactions.

Pairwise comparisons across states (feeding vs milling or calf absent vs calf present) within a category of significant interactions were made using least squares means with a Tukey–Kramer adjustment (Littell *et al.*, 2006). In the least squares means comparisons, noise levels within each frequency category were tested at low noise levels (lower quartile), average noise levels (middle quartiles), and high noise levels (upper quartile) (Table III). Significant relationships within a categorical state and involving a continuous covariate interaction were identified with linear regression.

III. RESULTS

Over the course of the study, 103 h of vocalization recordings were obtained during 128 sightings, and 50 focal follows. Of the total recordings, 62.5 h met all the analysis criteria and were reviewed in detail. These were the hours when the manatees were observed either as a single animal or all animals within the observation group were engaged in the same behavior. The number of usable hours in each behavioral category, as well as the number of vocalizations identified, is outlined in Table IV.

A preliminary analysis was conducted to determine whether manatee vocalizations can be classified into two distinct vocalization types. The main classification parameter was the number of inflection points. Calls containing one or more inflection points were designated chirps (Fig. 1). Calls with no inflection points were defined as squeaks (Fig. 1). A classification and regression tree was not needed, as the vocalization categories were split cleanly from the root based on one parameter, which accounted for 100% of the variability. Chirps tended to have a clear tonal sound to them, whereas squeaks sounded raspy. Additionally, five traditional acoustic parameters were measured for each vocalization: duration, frequency range, minimum frequency, maximum frequency, and frequency of peak energy. A series of Bonferroni corrected *t*-tests showed that chirps and squeaks differed acoustically based on the five parameters measured (Table V). Chirps were longer in duration, had a broader frequency range, and were higher in frequency compared to the more raspy sounding squeaks. Based on these findings, chirps and squeaks were analyzed separately for patterns associated with behavior, calf presence, group size, and ambient noise. The classification in this study supports the verbal descriptions of different calls made previously by Schevill and Watkins (1965), Bengtson and Fitzgerald (1985), and Nowacek *et al.* (2003) where vocalizations were described as either as tonal harmonic calls or broader-band, less tonal, nonharmonically related calls that sounded squeaky or raspy. Sousa-Lima *et al.* (2008) also identified the two vocalization types and presented the idea of a graded repertoire between the two call types. However, the study of Sousa-Lima *et al.* (2008)

TABLE IV. Behavior category breakdown of hours and identified vocalizations.

Behavior	Behavior (h)	Usable (h)	Total vocalizations	Trolling motor off	Trolling motor on
Social	20	14.5	3083	2668	415
Mill	10.5	10	677	580	97
Rest	23	18	935	887	48
Feed	23.5	20	1326	1083	243
Total	77	62.5	6021	5218	803
% of total				(86.7)	(13.3)

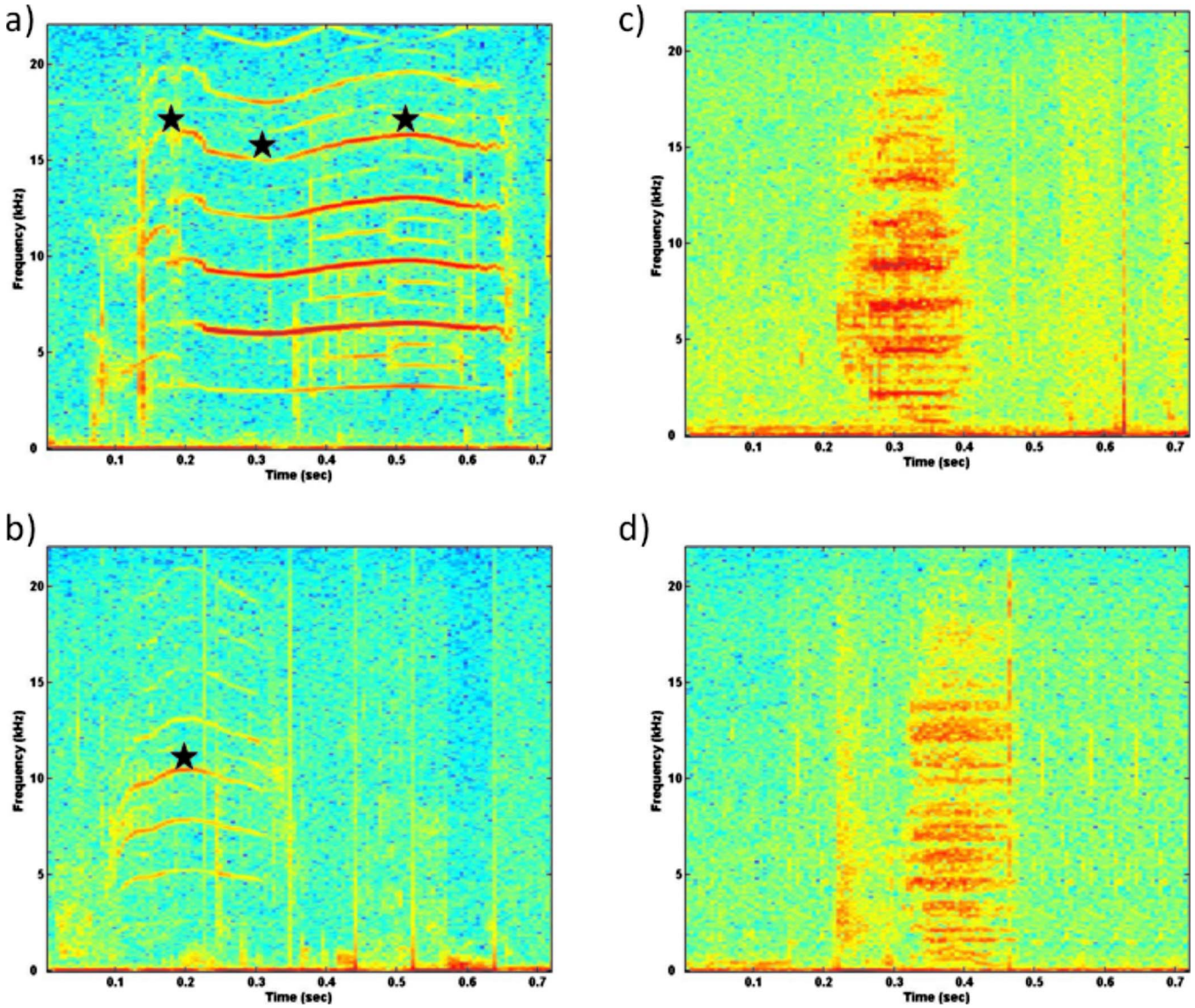


FIG. 1. (Color online) Spectrogram of manatee vocalization exemplars: (a) chirp, (b) chirp, (c) squeak, and (d) squeak. Spectrograms are plotted on a relative decibel scale. Black stars indicate inflection points.

did not measure inflection points as one of the vocalization parameters, which was the primary categorization parameter in this study.

The range of source levels obtained in this study (105–150 dB rms re 1 μPa at 1 m) overlapped with the range of source levels previously reported (90–138 dB rms re 1 μPa at 1 m) (Nowacek *et al.*, 2003; Phillips *et al.*, 2004). Average source level for chirps was 122 ± 6.5 dB rms re 1 μPa at 1 m. Average squeak source level was 120 ± 6.8 dB rms re 1 μPa at 1 m. These average source levels were greater than the means of 100 and 112 dB re 1 μPa at 1 m previously reported (Nowacek *et al.*, 2003; Phillips *et al.*, 2004). This difference could result from the selection in this study for source level calculations of vocalizations that exceeded the 3 dB threshold criteria. Quieter vocalizations tended to have less sharp onset and offset, were more likely to be masked by background noise, and were therefore more likely to be eliminated from further analysis. Source levels reported here therefore reflect a sample that was biased to represent relatively high signal-to-noise levels for manatee vocalizations, which is likely also biased for high source levels.

Alternative explanations as to why this study obtained greater source levels include (1) that vocalizations emitted during the summer season in very turbid waters may be louder than those emitted in less turbid waters or during the winter season and (2) that there was a systematic error in the

TABLE V. Summary table of the average parameter values for chirps and squeaks. The reported *p*-value should be compared with a Bonferroni adjusted alpha value of 0.01 for all parameters.

	Chirps		Squeaks		*t* stat	*p*-value
Sample size (*n*)	1168		375			
Duration (ms)	221.8	(93)	198.4	(69)	4.47	<0.001
Freq range (Hz)	15,033	(3915)	12,776	(4614)	9.28	<0.001
Min freq (Hz)	1,804	(704)	1,358	(868)	9.7	<0.001
Max freq (Hz)	18,026	(3953)	14,135	(4647)	11.02	<0.001
Peak freq (Hz)	5,097	(2721)	3,341	(1884)	11.63	<0.001

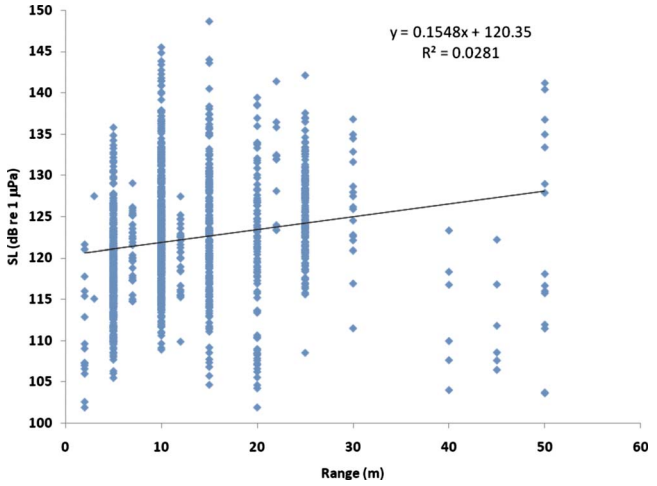


FIG. 2. (Color online) Source level plotted as a function of range for all vocalizations. The low r^2 value and p-value of 0.29 indicate no significant relationship.

transmission loss estimation when calculating source level. Source levels of Florida manatees reported by Nowacek *et al.* (2003) were obtained from a spring in Crystal River, FL, and those levels reported by Phillips *et al.* (2004) were obtained during the late winter/early spring season in Homosassa Springs Wildlife State Park. Environmental conditions associated with the turbid summer conditions in Sarasota may induce louder vocalizations in order to compensate for decreased visual range. Systematic error in the transmission loss estimates for calculating source level was unlikely, as there was no relationship between source level and range to the vocalizing animals ($p=0.29$) (Fig. 2).

A. Vocalization rate

Average vocalization rates (including chirps and squeaks) ranged from 1.04 to 4.39 vocalizations/manatee/5 min period. These values correspond well to the rates of 0.25–4.75 and 1–2 vocalization/manatee/5 min period previously reported by Bengtson and Fitzgerald (1985) and Phillips *et al.* (2004), respectively. A mixed model ANCOVA tested the hypothesis that all fixed effects regression slopes were equal to zero ($H_0: B_1=B_2=B_3=B_n=0$). Model results showed significant interactions between behavior and both LFN and MFN, indicating that the slopes were not all equal to zero (Table VI). Pairwise comparisons indicated that rates under average noise conditions in the LFN band were lower during feeding than resting and lower during resting than socializing (Table VI). Rates solely within the social behavior state also decreased with increasing LFN level (Table VI). Patterns seen for rates in association with MFN were similar. Resting vocal rates were lower than social vocal rates under average noise levels, whereas feeding rates were lower than social rates under high MFN levels. Rates solely within the feeding behavioral state decreased with increasing MFN levels (Table VI).

B. Chirp usage

A mixed model ANCOVA was run on the chirp vocalization parameters of source level, duration, frequency range, minimum frequency, maximum frequency, and peak frequency of the dominant harmonic. Table VI shows all the significant fixed effects and pairwise comparisons of significant interactions. In summary, manatees did not show modification of chirp source level, frequency range, minimum frequency, or maximum frequency in relation to behavioral context, group composition, or noise level. Chirp call durations were altered based on calf presence and noise level. Chirps had longer durations when calves were present under conditions of low and average LFN levels. Chirps were also longer when calves were present under conditions of average and high MFN levels. The peak frequency of the dominant harmonic of chirp vocalizations was context dependent. There was a significant interaction between group size and behavior. Multiple comparisons revealed that the peak frequency of chirps for manatees resting in groups equal to or greater than five animals was higher than equivalent groups of socializing manatees. A significant interaction between calf presence and MFN was also observed in relation to peak frequency. Under low MFN levels, chirps had lower peak frequencies when calves were present compared to when they were absent. Interestingly, when calves were absent, peak frequencies decreased with increasing MFN levels.

C. Squeak usage

A mixed model ANCOVA was used to investigate the effect of noise on squeak vocalization parameters. Table VI shows all the significant fixed effects and pairwise comparisons of significant interactions. In summary, modification of squeaks was not observed in relation to noise level, group composition, or behavioral context for any frequency parameter. Squeak source levels were modified as a function of HFN and behavioral state. Under conditions of low HFN levels, squeaks emitted during feeding were louder than during milling. The opposite was true under high levels of HFN. In high levels of HFN, squeaks emitted during milling were louder than during feeding. Durations of squeaks were shorter under high levels of LFN when calves were present. This was opposite of the trend observed for chirp durations. Lastly, squeak durations increased with increased in LFN when no calves were present.

IV. DISCUSSION

Whether most marine mammals can adjust the frequencies, duration, rate, and source levels of their various call types to increase communication range in the presence of noise has not been extensively studied (Richardson *et al.*, 1995). This study investigated whether manatees modify vocal parameters to compensate for environmental noise. It was discovered that manatees exhibited a number of vocal compensation techniques and that the vocal parameters modified were context dependent. Compensation techniques observed in response to LFN levels included a decrease in vocalization rate which was dependent on behavioral state, an increase in chirp durations when calves were present, and increased squeak durations as noise level increased when calves were absent. Similarly, behavior-dependent decreases in vocalization rate and increases in chirp durations with calf presence

TABLE VI. Significant ANCOVA effects and pairwise comparisons. The × denotes type 3 test for fixed effects. Across categorical state comparisons were made with least squares means with a Tukey–Kramer adjustment. Covariate relationships within categorical states were tested with linear regression. In the Significant Relationships descriptions, the values in parentheses are mean values of the measured parameter.

Category	Response	Significant Effects	Effect F value (p value)*	Significant relationships	Relationship statistics
	Rate (calls/indv/5 min)	LFN × behavior	$F_{3,27}=3.17$ (0.040)	Feeding (1.08) < resting (2.12) for average noise levels (across)	$t=3.13$; $p=0.02$; df=27
				Resting (2.12) < social (2.61) for average noise levels (across)	$t=-3.83$; $p=0.004$; df=27
				Social rates decreased with increasing noise (within)	$y=-0.16x+12.6$; $r^2=0.25$; $F_{1,14}=4.8$; $p=0.04$
		MFN × behavior	$F_{3,27}=3.71$ (0.023)	Resting (2.06) < social (2.52) for average noise levels (across)	$t=-3.23$; $p=0.02$; df=27
				Feeding (1.63) < social (3.31) for high noise condition (across)	$t=-3.40$; $p=0.01$; df=27
				Feeding rates decreased with increasing noise (within)	$y=-0.10x+8.2$; $r^2=0.20$; $F_{1,22}=5.4$; $p=0.03$
Chirp	SL (dB)	None			
	Duration (ms)	LFN × calf presence	$F_{1,39}=7.35$ (0.010)	Present (219.8) > absent (194.4) for low noise levels (across)	$t=-3.38$; $p=0.002$; df=39
				Present (231.5) > absent (201.8) for average noise levels (across)	$t=-2.20$; $p=0.03$; df=39
		MFN × calf presence	$F_{1,39}=8.77$ (0.005)	Present (224.1) > absent (201.5) for average noise levels (across)	$t=-2.66$; $p=0.01$; df=39
				Present (266.2) > absent (191.8) for high noise levels (across)	$t=-3.50$; $p=0.001$; df=39
	Freq range (Hz)	None			
	Min freq (Hz)	None			
	Max freq (Hz)	None			
	Peak freq (Hz)	Group size × behavior	$F_{3,32}=4.93$ (0.006)	For groups ≥5, resting (6785) > social (5071)	$t=2.87-3.53$; $p=0.007-0.04$; df=32
		MFN × calf presence	$F_{1,32}=7.34$ (0.011)	Present (4702) < absent (5250) for low noise levels (across)	$t=3.27$; $p=0.003$; df=32
				Absent: peak frequency decreased with increasing noise levels (within)	$y=-70.4x+8691$; $r^2=0.15$; $F_{1,32}=5.5$; $p=0.03$
Squeak	SL (dB)	HFN × behavior	$F_{3,21}=3.60$ (0.032)	Feeding (123.6) > milling (119.0) for low noise levels (across)	$t=2.95$; $p=0.04$; df=21
				Feeding (120.8) < milling (130.3) for high noise levels (across)	$t=-2.94$; $p=0.04$; df=21
	Duration (ms)	LFN × calf presence	$F_{3,19}=5.33$ (0.030)	Present (156.6) < absent (236.2) for high noise levels (across)	$t=2.45$; $p=0.02$; df=19
				No calves: durations increased in high noise levels (within)	$y=4.9x-76.7$; $r^2=0.22$; $F_{1,23}=6.6$; $p=0.02$
	Freq range (Hz)	None			
	Min freq (Hz)	None			
	Max freq (Hz)	None			
	Peak freq (Hz)	None			

were observed in response to MFN. Peak frequency of chirp vocalizations decreased with increasing MFN, and peak frequencies were higher when calves were absent and MFN was low. Vocal responses to HFN included a source level pattern shift.

Further discussion places the observed patterns of manatee vocal compensation in a broader framework of costs and benefits relating to effective communication under conditions of different group composition and behavioral contexts. Vocalization effort is the term used here to represent the physical cost of acoustic communication. Vocalization effort refers to the energy expenditure of producing a call and is a function of source level, vocalization rate, signal duration, and/or vocalization frequency. Vocalization effort is increased by increasing call amplitude, rate, and duration. Shifting to higher frequencies may take more energy for a manatee to produce, thus increasing vocalization effort, if the sound production mechanism of the signaler is tuned to the frequencies usually used. It has been argued that high frequency sounds are more costly to produce because it takes more energy to produce sounds that are closer in wavelength to the size of the sound-producing and resonating structures (Ryan, 1986).

Increasing vocalization effort is one mechanism to increase or maintain effective communication in the presence

of environmental noise. Manatees exhibited marked changes in vocalization effort by selectively decreasing vocalization rate, increasing duration, decreasing peak frequencies, and altering source levels in response to noise levels. The specific vocal parameter altered was a function of behavioral state, calf presence, noise level, and frequency of noise. Manatees emitted longer chirps when calves were present, whereas the effect on squeaks was the opposite. The difference in the results between chirps and squeaks hints at a mother-calf proximity maintenance function of chirps, as suggested by Sousa-Lima *et al.* (2002, 2008). Increased vocal effort in this circumstance may aid in more effective communication geared toward maintaining mother-calf contact. Alternatively, vocal effort decreased with increasing noise levels due to a decrease in vocalization rate during social and feeding behaviors. This pattern may indicate that manatees wait until it is quiet to vocalize while socializing and feeding, exhibiting only minimal vocalization effort during periods of high noise. Such a pattern of increased signaling during periods of low noise has been documented to avoid interference from conspecific vocalizations (e.g., Cody and Brown, 1969; Ficken *et al.*, 1974) or anthropogenic noise (Sun and Narins, 2005; Parks *et al.*, 2007; Sousa-Lima and Clark, 2008). Another way manatees increased their vocal effort was by increasing source level. Under low noise levels the source level of squeaks emitted during feeding was less than that emitted during milling. However, under high noise levels, the source level of squeaks recorded during milling was higher than when feeding. Manatees engaged in milling behaviors tend to separate more than when feeding, resting, or socializing, so it is possible that manatees need to increase vocalization effort under conditions of elevated noise levels to compensate for noise effects given the greater ranges between signaler and receiver at these times.

According to communication theory, manatees should only increase their vocalization effort when the benefits of effective vocal communication outweigh the cost of increased energy expenditure. Some benefits of effective communication include maintaining mother-calf contact, maintaining group cohesion, and signaling intentions. One important factor driving vocalization effort is the range over which the signaler and receiver must effectively communicate. Typical ranges between animals will vary as a function of context, and vocalization effort would be expected to vary accordingly. Consequently, a common reason for increasing the energy involved in vocalizing would be to maintain the required range of effective communication in increased noise. Increased vocal effort resulting from increases in source level in elevated noise conditions is a well known response known as the Lombard effect in humans, and has been demonstrated for other animals including nightingales (*Luscinia megarhynchos*), beluga whales (*Delphinapterus leucas)*, common marmoset (*Callithrix jacchus*), and cats (Brumm, 2004; Egnor *et al.*, 2006; Nonaka *et al.*, 1997; Scheifele *et al.*, 2005). Results presented here indicate that manatees can now be included in this list.

Our findings demonstrate that manatees do alter their vocalization effort, as indicated by changes in usage and structure, as a function of behavior and group composition. However, there was no simple or clear pattern that emerged from the results. Clear-cut results are often rare in biology and variation can be indicative of animals using different strategies to compensate for the same stressor (i.e., one animal increases source level while another increases duration to compensate for noise). More extensive studies are needed to provide details of how these changes affect the probability of detecting and classifying the signals, how net change in vocalization effort relates to energetic expenditure, and whether or not the mode of compensation is indicative of call function. The findings presented here pertain only to the Florida manatee during the nonwinter months and in very turbid environments. Whether the observed patterns can be generalized to all seasons and habitats also requires additional studies.

ACKNOWLEDGMENTS

Special thanks are extended to the staff in the Manatee Research Program at Mote Marine Laboratory for their support of this study. This research would not have been possible without the aid of numerous Mote interns, and the sharing of knowledge and time from the staff. We also wish to thank anonymous reviewers for their insightful comments on a previous version of this manuscript. This project was conducted under permits issued by the U.S. Federal Fish and Wildlife Service to David Mann at the University of South Florida (MA051709–0) and the Fish and Wildlife Research Institute (MA773494–7), and it compiled with all the current federal and state laws. This research was supported by a P.E.O. Scholar Award and National Defense Science and Engineering Fellowship awarded to Jennifer Miksis.

Bengston, J. L. (**1981**). "Ecology of manatees (*Trichechus manatus*) in the St. Johns River, Florida," Ph.D. dissertation, University of Minnesota, Minneapolis, MN.

Bengtson, J. L., and Fitzgerald, S. M. (**1985**). "Potential role of vocalizations in West Indian manatees," J. Mammal. **66**, 816–819.

Brumm, H. (**2004**). "The impact of environmental noise on song amplitude in a territorial bird," J. Anim. Ecol. **73**, 434–440.

Brumm, H., and Slabbekoorn, H. (**2005**). "Acoustic communication in noise," Adv. Stud Behav. **35**, 151–209.

Brumm, H., and Slater, P. J. B. (**2006**). "Ambient noise, motor fatigue, and serial redundancy in chaffinch song," Behav. Ecol. Sociobiol. **60**, 475–481.

Cody, M. L., and Brown, J. H. (**1969**). "Song asynchrony in neighbouring bird species," Nature (London) **222**, 778–780.

Egnor, S. E. R., Iguina, C. G., and Hauser, M. D. (**2006**). "Perturbation of auditory feedback causes systematic perturbation in vocal structure in adult cotton-top tamarins," J. Exp. Biol. **209**, 3652–3663.

Ficken, R. W., Ficken, M. S., and Hailman, J. P. (**1974**). "Temporal pattern shifts to avoid acoustic interference in singing birds," Science **183**, 762–763.

Foote, A. D., Osborne, R. W., and Hoelzel, A. R. (**2004**). "Whale-call response to masking boat noise," Nature (London) **428**, 910.

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.

Hartman, D. S. (**1979**). "Ecology and behavior of the manatee," *Trichechus manatus*, in Florida (American Society of Mammalogy, Special Publication 5.

Ketten, D. R., Odell, D. K., and Domning, D. P., **1992**, "Structure, function, and adaptation of the manatee ear," in *Marine Mammal Sensory Systems*, edited by J. A. Thomas, R. A. Kastelein, and A. Ya. Supin (Plenum, New York), pp. 77–95.

Lengagne, T., Aubin, T., Lauga, J., and Jouventin, P. (**1999**). "How do king

penguins (*Aptenodytes patagonicus*) apply the mathematical theory of information to communicate in windy conditions?," Proc. R. Soc. London, Ser. B **266**, 1623–1628.
Lesage, V. C., Barrette, C., and Kingsley, M. C. S. (**1999**). "The effect of noise from an outboard motor and ferry on the vocal activity of beluga (*Delphinapterus leucas*) in the St. Lawrence Estuary, Canada," Abstracts of the Tenth Biennial Conference on the Biology of Marine Mammals, Galveston, TX.
Littell, R. C., Milliken, G. A., Stroup, W. W., Wolfinger, R. D., and Schabenberger, O. (**2006**), *SAS for Mixed Models*, 2nd ed. (SAS, Cary, NC).
Miksis-Olds, J. L., Donaghay, P. L., Miller, J. H., Tyack, P. L., and Nystuen, J. A. (**2007**). "Noise level correlates with manatee use of foraging habitats," J. Acoust. Soc. Am. **121**, 3011–3020.
Miksis-Olds, J. L., and Miller, J. H. (**2006**). "Transmission loss in manatee habitats," J. Acoust. Soc. Am. **120**, 2320–2327.
Miller, P. J. O., Biassoni, N., Samuels, A., and Tyack, P. L. (**2000**). "Whale songs lengthen in response to sonar," Nature (London) **405**, 903.
Nonaka, S., Takahashi, R., Enomoto, K., Katada, A., and Unno, T. (**1997**). "Lombard reflex during PAG-induced vocalizations in decerebrate cats," Neurosci. Res. **29**, 283–289.
Nowacek, D. P., Casper, B. M., Wells, R. S., Nowacek, S. M., and Mann, D. A. (**2003**). "Intraspecific and geographic variation of West Indian manatee (*Trichechus manatus spp.*) vocalizations (L)," J. Acoust. Soc. Am. **114**, 66–69.
Oppenheim, A. V., Schafer, R. W., and Buck, J. R. (**1999**). *Discrete-Time Signal Processing*, 2nd ed. (Prentice-Hall, Englewood Cliffs, NJ).
O'Shea, T. J., and Hartley, W. C. (**1995**). "Reproduction and early-age survival of manatees at Blue Spring, upper St. Johns River, Florida," in *Population Biology of the Florida Manatee*, edited by T. J. O'Shea, B. B. Ackerman, and H. F. Percival (National Biological Service Information and Technology, Gainesville, FL), pp. 157–170.
Parks, S., Clark, C., and Tyack, P. L. (**2007**). "Short and long-term changes in right whale calling behavior: The potential effects of noise on acoustic communication," J. Acoust. Soc. Am. **122**, 3725–3731.
Parks, S. E. (**2003**). "Acoustic communication in the North Atlantic right whale, *Eubalaena glacialis*," Ph.D. dissertation, MIT, Woods Hole, MA.
Penna, M., Pottstock, H., and Velasquez, N. (**2005**). "Effect of natural and synthetic noise on evoked vocal responses in a frog of the temperate austral forest," Anim. Behav. **70**, 639–651.
Phillips, R., Niezrecki, C., and Beusse, D. O. (**2004**). "Determination of West Indian manatee vocalization levels and rate," J. Acoust. Soc. Am. **115**, 422–428.
Richardson, W., Greene, C., Malme, C., and Thomson, D. (**1995**). *Marine Mammals and Noise* (Academic, San Diego, CA).
Ryan, M. J. (**1986**). "Factors influencing the evolution of acoustic communication: Biological constraints," Brain Behav. Evol. **28**, 70–82.
Scheifele, P. M., Andrew, S., Cooper, R. A., Darre, M., Musiek, F. E., and Max, L. (**2005**). "Indication of a Lombard vocal response in the St. Lawrence River beluga," J. Acoust. Soc. Am. **117**, 1486–1492.
Schevill, W. E., and Watkins, W. A. (**1965**). "Underwater calls of *Trichechus* (Manatee)," Nature (London) **205**, 373–374.
Slabbekoorn, H., and Peet, M. (**2003**). "Birds sing at a higher pitch in urban noise," Nature (London) **424**, 267.
Smith, K. B. (**2001**). "Convergence, stability, and variability of shallow water acoustic predictions using a split-step Fourier parabolic equation model," J. Comput. Acoust. **9**, 243–285.
Sousa-Lima, R. S., and Clark, C. W. (**2008**). "Modeling the effect of boat traffic on singing activity of humpback whales (*Megaptera novaeangliae*) in the Abrolhos National Marine Park, Brazil," Can. Acoust. **36**, 174–181.
Sousa-Lima, R. S., Paglia, A. P., and da Fonseca, G. A. B. (**2002**). "Signature information and individual recognition in the isolation calls of the Amazonian manatees, *Trichechus inunguis* (Mammalia: Sirenia)," Anim. Behav. **63**, 301–310.
Sousa-Lima, R. S., Paglia, A. P., and da Fonseca, G. A. B. (**2008**). "Gender, age, and identity in the isolation calls of Antillean manatees (*Trichechus manatus manatus*)," Aquat. Mamm. **34**, 109–122.
Sun, J. W. C., and Narins, P. M. (**2005**). "Anthropogenic sounds differentially affect amphibian call rate," Biol. Conserv. **121**, 419–427.
Urick, R. J. (**1983**). *Principles of Underwater Sound*, 3rd ed. (Peninsula, Los Altos, CA).

Predicting temporary threshold shifts in a bottlenose dolphin (*Tursiops truncatus*): The effects of noise level and duration

T. Aran Mooney[a)]
Department of Zoology and Hawaii Institute of Marine Biology, University of Hawaii, P.O. Box 1106, Kailua, Hawaii 96734

Paul E. Nachtigall, Marlee Breese, Stephanie Vlachos, and Whitlow W. L. Au
Hawaii Institute of Marine Biology, University of Hawaii, P.O. Box 1106, Kailua, Hawaii 96734

(Received 21 June 2008; revised 11 December 2008; accepted 13 December 2008)

Noise levels in the ocean are increasing and are expected to affect marine mammals. To examine the auditory effects of noise on odontocetes, a bottlenose dolphin (*Tursiops truncatus*) was exposed to octave-band noise (4–8 kHz) of varying durations (<2–30 min) and sound pressures (130–178 dB re 1 μPa). Temporary threshold shift (TTS) occurrence was quantified in an effort to (i) determine the sound exposure levels (SELs) (dB re 1 μPa^2 s) that induce TTS and (ii) develop a model to predict TTS onset. Hearing thresholds were measured using auditory evoked potentials. If SEL was kept constant, significant shifts were induced by longer duration exposures but not for shorter exposures. Higher SELs were required to induce shifts in shorter duration exposures. The results did not support an equal-energy model to predict TTS onset. Rather, a logarithmic algorithm, which increased in sound energy as exposure duration decreased, was a better predictor of TTS. Recovery to baseline hearing thresholds was also logarithmic (approximately −1.8 dB/doubling of time) but indicated variability including faster recovery rates after greater shifts and longer recoveries necessary after longer duration exposures. The data reflected the complexity of TTS in mammals that should be taken into account when predicting odontocete TTS.
 [DOI: 10.1121/1.3068456]

PACS number(s): 43.80.Nd, 43.80.Lb, 43.66.Ed [ADP] Pages: 1816–1826

I. INTRODUCTION

Anthropogenic noise in the ocean stems from a variety of sources including shipping, naval sonar, scientific study, oil exploration and drilling, and construction. As usage of the oceans increases, marine noise levels are also expected to rise (National Academy of Sciences, 2005). Serious concern regarding the effects of this noise on marine mammals, as major utilizers of sound in the ocean, has been emerging during the past decade (Richardson *et al.*, 1995; National Academy of Sciences, 2003; Wartzok *et al.*, 2004; National Academy of Sciences, 2005). Excessive noise exposure in marine mammals can induce a variety of adverse behavioral and physiological consequences including changes in hearing sensitivity. In order to mitigate these effects in wild populations, it is necessary to better understand their causes.

Excessive sound exposures may induce both permanent and temporary alterations in hearing thresholds. Permanent threshold changes occur when hearing capabilities do not return to baseline levels. These changes are indicative of hearing damage and are referred to as permanent threshold shifts (PTSs). Temporary threshold shifts (TTSs) demonstrate full recovery of hearing abilities and are generally assumed to be innocuous. These shifts have been demonstrated across vertebrates including fish, reptiles, birds, and mammals (Ward *et al.*, 1958; Saunders and Dooling, 1974; Popper and Clarke, 1976; Mulroy, 1986). Characterizing and understanding how TTS is induced may allow the extrapolation and prediction of PTS levels (Southall *et al.*, 2008). Further, TTS exposure conditions may be considered as a reasonable limit of excessive noise exposure. There is much concern, for obvious reasons, for determining the levels of noise that induce TTS or PTS in humans. As a result, the subject is well studied in some terrestrial mammals, and the variables that relate to TTS (intensity, duration and frequency) are relatively well understood. Thus, models have been developed to predict situations that would induce human TTS and PTS (Ward *et al.*, 1959; Kryter *et al.*, 1966; Ward *et al.*, 1976).

TTS has long been demonstrated in other taxa, but it was not until recently that it was shown that cetaceans are also susceptible to threshold shifts (Schlundt *et al.*, 2000). Further research has shown a relatively robust and resilient marine mammal hearing system and has demonstrated shifts using broadband noise, tones, and seismic waterguns (Finneran *et al.*, 2002; Nachtigall *et al.*, 2003; Finneran *et al.*, 2005). Yet there is much we do not know regarding TTS occurrence and noise effects in cetaceans. For example, do short loud sounds have the same effects as longer quieter sounds of equivalent energy? If we know the intensity and duration of a noise exposure, can we predict TTS?

Answers to these questions require the investigation of a wide range of fatiguing noise levels and durations to develop a predictive model based on empirical evidence. In pinni-

[a)]Present address: Biology Department, Woods Hole Oceanographic Institution, Woods Hole, MA 02543. Electronic mail: amooney@whoi.edu.

peds, several studies have done just that (Kastak *et al.*, 2005; 2007). However, in odontocetes, predictive TTS models have been developed based on comparisons across studies that used quite different methodologies and fatiguing stimuli, and the models developed are consequently straightforward but general (Finneran *et al.*, 2005). Thus there is a need for a comprehensive study that encompasses a range of noise level and duration conditions in order to accurately predict the effects of noise on a representative odontocete.

The goal of this study was to examine the relationship between fatiguing noise amplitude and duration in inducing TTS in an odontocete cetacean, and in doing so, to develop a model that predicts the noise levels and durations that would cause TTS and determine if a simple time-intensity trade-off (equal-energy) rule could be applied to these predictions. A secondary goal was tracking the recovery from TTS to establish the recovery rates. To achieve this, the auditory evoked potential (AEP) technique was utilized which allowed for rapid and repeated auditory threshold measurements. Data gathered using the AEP method compare favorably with those obtained using behavioral tests (Yuen *et al.*, 2005; Finneran and Houser, 2006) and have been applied previously in other marine mammal audiometric work including TTS investigations (Nachtigall *et al.*, 2004; Finneran *et al.*, 2007; Nachtigall *et al.*, 2007).

II. MATERIALS AND METHODS

A. Subject and experimental procedure

The subject used in this experiment was an 18-year-old male Atlantic bottlenose dolphin, *Tursiops truncatus,* named Boris. Boris was born and raised in the dolphin breeding colony at the marine mammal research facility in Kaneohe Bay, Oahu, HI. The animal has had substantial cooperative experience with hearing research experiments, including AEP and TTS work (Nachtigall *et al.*, 2003; Nachtigall *et al.*, 2004).

All threshold testing was conducted in the floating open-water sea pens of the Hawaii Institute of Marine Biology, moored off Coconut Island in Kaneohe Bay. The experiment began in May 2004 with the establishment of a baseline audiogram of the subject. Controlled noise exposures for this experiment were introduced in February 2005 and conducted through September 2006 with a total of 57 noise exposures. Exposures were permitted once every 4 days; however, more typically exposures occurred once per week, and often there would be several weeks without exposure sessions. The animal's hearing was monitored and always returned to baseline levels prior to a subsequent exposure. Control sessions were paired with noise exposure sessions and were experimentally identical, excluding the presentation of a fatiguing sound. Because noise exposure sessions were usually presented once per week, a greater number of control and training threshold sessions were conducted, resulting in 82 control sessions (thus some controls were repeated) and 201 days and nearly 300 hearing thresholds measured that were not associated with a fatiguing noise exposure. Noise exposures were permitted with only one subject; thus measurements were often repeated, and data were compared to those of other marine and terrestrial mammals. This ensured that our data were not strongly influenced by a one time occurrence.

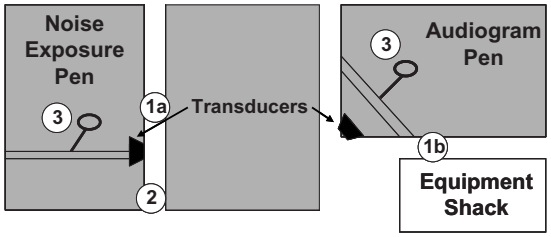


FIG. 1. Diagram of dolphin audiogram threshold testing and fatiguing noise pen experimental setup. (1a, 1b) Trainer positions, (2) assistant position, (3) hoop stations for noise and threshold tests. The equipment shack that housed the AEP and noise exposure equipment is also indicated.

Noise sessions consisted of three phases: (1) a pre-exposure threshold measurement to ensure that the subject's threshold was similar to its "baseline," or average threshold, (2) the noise exposure session, and (3) follow-up threshold measurements designed to determine the amount of threshold shift and track the subject's recovery until the subject's hearing returned to the normal threshold range. A threshold shift was determined as the postexposure thresholds, which exceeded +1 standard deviation (SD) of the subject's average threshold. The amount of TTS was measured as a TTS postexposure threshold minus the animal's average threshold. We based all TTS measurements off the subject's mean threshold because between measurements, hearing thresholds often vary slightly (Johnson, 1967; Finneran and Houser, 2006). Using the average threshold allowed us to apply a standard threshold value and TTS level (+1 SD) for each respective frequency examined. All threshold measurements were conducted in the threshold measurement pen using AEPs (Fig. 1). Five AEP records, of 50 s each, were collected per threshold determination; thus a threshold could be estimated in 4–5 min. After the initial threshold measurements, the dolphin voluntarily moved to a separate pen for the fatiguing noise exposure. Immediately after the noise exposure, Boris returned to the threshold measurement pen for the postexposure threshold measurements. These measurements would begin 1–2 min after the cessation of the noise exposure and were conducted with their middle point at 5, 10, 20, 40, and 80 min after noise exposure to thoroughly track the subject's hearing recovery.

B. Evoked potential measurements

At the beginning of each hearing threshold measurement, Boris entered the hearing test pen, and he immediately stationed in front of the trainer. The trainer then gently placed two standard 10-mm gold electroencephalogram (EEG) electrodes, embedded in latex suction cups on the animal. The recording electrode was located 4–10 cm behind the blow-hole, just to the left or right of the animal's midline. The reference electrode was placed on the subject's dorsal fin, which minimized electrical noise from extraneous muscle or nerve movements. Signal conduction was enhanced by the use of standard conductive gel placed between the skin and the electrodes. The animal then returned to station. Upon the trainer's cue, Boris restationed in a hoop 1 m

below the surface and faced a sound projector 2 m away. An acoustic baffle was hung at the surface, 1 m from both the dolphin and the transducer, to prevent extraneous acoustic surface reflections from interfering in the hearing threshold area. The dolphin remained in the hoop for one to three AEP trials (~1–3 min), after which he was recalled to the surface to breathe and receive fish rewards. Boris was then quickly resent to the hoop for the remaining AEP trials. If the hearing thresholds were part of the postexposure threshold measurements, the dolphin was given a break, either resting at the station or being taken out of the test pen, until it was time to begin the next threshold trials.

The sound stimuli were sinusoidally amplitude modulated (SAM) tone bursts digitally generated with a custom LABVIEW program. The tones were converted from digital to analog using an update rate of 200 kHz and a National Instruments PCI-MIO-16E-1 data acquisition (DAQ) card implemented into a desktop computer. Individual tone bursts were 20 ms in duration, presented at a rate of 20 s^{-1}, and for 1000 bursts per trial. On the basis of prior established dolphin modulation rates (Supin and Popov, 1995; Mooney *et al.*, 2006), carrier frequencies were modulated at 1000 Hz and with a modulation depth of 100%. The analog signals were sent from the computer to a custom-built attenuator that could decrease the tones in 1 dB steps. From the signal shaping box, outgoing signals were sent to the projecting transducer, an ITC-1032, and concurrently monitored on a Tektronix TDS 1002 oscilloscope. For each session, thresholds were collected at one of five frequencies: 5.6, 8, 11.2, 16, or 22.5 kHz. Signals typically started at a predetermined start sound pressure level (SPL) of 15–20 dB above the mean threshold and were reduced in 5 dB steps for the remainder of the five trials. If TTS was apparent and thresholds were reached before four to five trials, the SPLs were increased to best track the threshold.

A 30-ms AEP record was collected simultaneously with each stimulus presentation. The received AEPs were amplified and filtered using an Iso-Dam Isolated Biological Amplifier and a Krohn-Hite 3102 filter, both with a bandpass of 300–3000 Hz. They were then digitized by the DAQ card at a rate of 16 kHz. To extract the AEP from noise, 1000 of these records were averaged for each AEP trial.

C. Noise exposure

The noise exposure pen was equipped with a hoop fixed where the subject was required to station for the exposure 2 m from the fatiguing noise transducer. When the subject entered the noise exposure pen, he immediately stationed in front of the trainer. Upon a visual cue from the trainer, the dolphin swam and stationed in the hoop. Typically, several warm-up trials were initiated where the subject was sent to the hoop for 1–2 min with the noise off. When the equipment, animal, and trainer were ready, the trainer directed the fatiguing noise equipment operator to turn the noise on. Because the subject initially demonstrated a startle response that might have inhibited the experiment when an intense fatiguing noise was abruptly turned on, the experimental procedure required the sound to be ramped up from 130 dB re 1 μPa to 15 dB below the deemed exposure level or a maximum value of 150 dB over the course of a 30-s trial. The animal was then recalled back to the trainer, and the sound was turned up to the planned experimental intensity. The dolphin was then sent back to the hoop to begin what was considered the timed exposure. The warm-up trials were on average 140–145 dB SPL of noise for 30 s and well below TTS levels discussed here and elsewhere (Nachtigall *et al.*, 2004; Finneran *et al.*, 2007). Further, these trials did not substantially increase sound exposure levels (<0.01 dB) re 1 $\mu Pa^2 s$. For these reasons, they were not likely to affect measured TTS and were not considered part of the noise exposure.

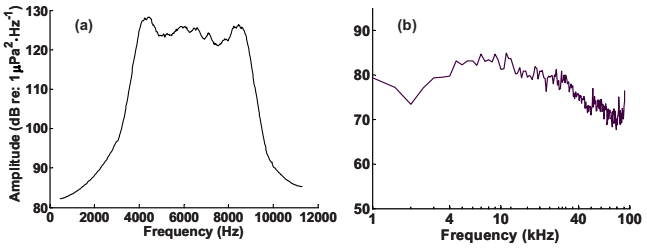


FIG. 2. (Color online) (a) Fatiguing noise spectral density recorded from the hoop calibration position in the noise pen. The octave-band noise is 4–8 kHz with a center frequency of 5.6 kHz. The SPL in this case is approximately 160 dB re 1 μPa. (b) Ambient sound at experimental pens in Kaneohe Bay, Oahu, HI measured with a Biomon 8235 and plotted as noise spectral density using a 1024-point FFT. Ten noise samples were averaged to create both plots.

To examine the effects of noise duration and intensity on dolphin hearing, both exposure duration and SPL were varied. Hoop sound exposure duration was set at 30, 15, 7.5, 5.6 (337 s), 3.75, or 1.88 min. All exposure durations were monitored by the trainer who cumulatively noted and controlled the animal's exact time in the hoop for each trial. In doing so, the trainer relayed to the equipment operator the exact time the fatiguing sound could be turned off to get a precise exposure duration (e.g., 1.88 min or 112 s). Fatiguing noise intensities ranged from 130 to 178 dB re 1 μPa where, irrespective of the initial ramp-up, SPLs were kept constant throughout the exposure. The fatiguing noise was an octave-band noise of 4–8 kHz (Fig. 2). It was generated by a custom-built white noise generator and then filtered using a custom bandpass filter. Noise was then amplified using a Hafler P3000 amplifier, monitored on the oscilloscope and played through a Massa TR-61A transducer (peak frequency of 5.5 kHz). The noise level was calibrated 11 times throughout the experiment to ensure that no drift occurred in sound levels. To calibrate, a Biomon 8235 hydrophone (−173 dB sensitivity and ±1 dB from 1 to 40 kHz) was placed in the center of the hoop when the animal was not present. The hydrophone was connected to the oscilloscope to monitor the noise levels, a Fluke 8922A rms voltmeter (1 MHz bandwidth, ±0.01 dB up to 200 kHz, 1 s integration time), the same DAQ card, and a custom LABVIEW program to record received levels. The noise was then measured with the voltmeter in SPL based on the total rms over the 4–8 kHz band. Received sound pressure varied by ±0.35 dB (SD) at the 160 dB level across the 11 calibrations. From the SPL (re 1 μPa), it was then possible to calculate the noise spectral

density (dB re 1 $\mu Pa^2\ Hz^{-1}$) from 4 to 8 kHz. Sound exposure level (SEL) (dB re 1 μPa^2 s) was also calculated from the received SPLs for each exposure sound pressure and duration. SPL and SEL are hereafter referred to with the above units.

The exposure schedule was semirandom. Initial exposures were 30 min, and SPL was increased until threshold shifts were induced. Duration was then decreased in order, from 30 to 1.88 min. We then increased the duration (similar to an ABBA session format). Exposures were then remeasured at the shortest durations (1.88 and 3.75 min) using higher SELs than the previous sessions. These sessions were repeated twice. We then increased exposure duration again, this time at the initial SELs. Thresholds shifts were usually investigated at all five frequencies before exposure duration was changed. The time between exposure sessions as well as the variation of exposure duration and SEL likely experimentally precluded some processes that could affect TTS (such as cochlear toughening; Hamernik *et al.*, 2003). Further, the animal was well experienced in TTS studies (Nachtigall *et al.*, 2003; Nachtigall *et al.*, 2004) and likely did not develop unique protective auditory mechanisms for this study.

The dolphin's behavior was continuously monitored during the noise exposure by both the trainer and an assistant. While the trainer was responsible for interacting with the dolphin, the assistant recorded behavioral alterations including number of respirations during surface intervals, latency of time from surface station to the hoop (delay), excessive head or body movement, or any apparent reactions to the fatiguing sound (or reactions within control trials). Because the dolphin had significant experience in TTS, AEP, and other psychoacoustic experiments, his "normal" behavior was well known, and behavioral changes were easily noted. The animal also had significant previous training required to maintain participation in research and husbandry activities, and thus reactions to avoid such activities were expected to be minimized. The dependent variables of respirations, latency, and behavioral modification were measured and analyzed in respective one-way analyses of variance (ANOVAs) to determine if behavioral changes were observed. The assistant also informed the trainer of the duration of the surface intervals and time in the hoop, which were varied somewhat from trial to trial to prevent the dolphin from predicting when each trial and hoop session was over. However, the total time in the hoop for each noise exposure or control session was pre-established and maintained as the exposure time for that session (i.e., 30, 15,… min). While the animal needed to surface and breathe during exposure sessions, this did not increase SELs because the surface intervals were kept short (45 s) and SPLs were lower at the surface station (10 dB).

D. Data analysis

A 16-ms window of each averaged AEP response to the SAM tones was fast Fourier transformed (FFT) to view the response in the frequency domain. At higher amplitude SPLs (~20 dB above threshold), the spectra reflected a well-pronounced peak of 0.1–0.2 μV at the frequency of stimulus modulation—1000 Hz. At higher sound levels, these peaks were easily distinguishable from background noise; this is because the background noise levels of these spectra were distributed among 50 harmonics across a range of 3 kHz. Thus the magnitude of this noise was a few nV. At SPLs close to threshold, the amplitude of the 1000-Hz spectrum peak was dependent on stimulus sound pressure. Thus, the peak values could be plotted against SPL, and an approximated linear regression could be obtained addressing the peak values. The point where the regression line crossed the abscissa was taken as the theoretical sound level at which no AEP response would be induced and thus was considered the animal's threshold (Nachtigall *et al.*, 2004; Nachtigall *et al.*, 2007). While the entire EFR input-output function of the odontocete was not linear, the oblique portion near threshold was reasonably approximated by regression methods, thus aiding to determine an arbitrary but well established method to estimate response threshold (e.g., Supin *et al.*, 2001; Popov *et al.*, 2005; Nachtigall *et al.*, 2008). Five records were collected for each threshold set, and the points with the highest r^2 value were used in determining the thresholds. While actual threshold measurements were determined offline, EFRs and FFTs were monitored in real time to assess the animal's AEP response. If both values reached background noise levels for two subsequent trials, SPL was increased 5 dB above the start SPL. This was to ensure the greatest number of reliable data points per threshold estimation within a short period of time.

To determine TTS, postexposure thresholds were compared to the subject's mean (baseline) thresholds for each of the five frequencies tested. These thresholds and their respective standard deviations were calculated from measurements not associated with a noise exposure. More than 40 thresholds were acquired for each frequency. A threshold shift was then determined as a threshold, after a control or actual noise exposure, where the threshold exceeded 1 SD above the mean threshold. These criteria allowed for comparisons to an established threshold and variance. Variance beyond this could then likely be explained by the dependent variable, noise exposure. A TTS was defined as a demonstrated "recovery" back to within ± 1 SD of the mean. These shifts were deemed greater than the day-to-day variation that was found in the baseline thresholds and were distinguishable by comparison to over 300 thresholds measured during the course of the experiment. The consequent recoveries provided confirmation of the shift. All analyses were completed with Excel, MATLAB, and MINITAB software.

III. RESULTS

Baseline hearing thresholds for the dolphin subject were relatively consistent, varying by 2–3 dB SD (mean =2.8 dB) at a particular frequency [Fig. 3(a) and 3(b)]. Hearing was most sensitive at the higher frequencies (16 and 22.5 kHz) and followed a typical mammalian curve. Baseline thresholds and TTS were primarily explored at 11.2 kHz. The subject demonstrated hearing recovery from all threshold shifts, and baseline thresholds did not significantly increase over the duration of the experiment (Fig. 4).

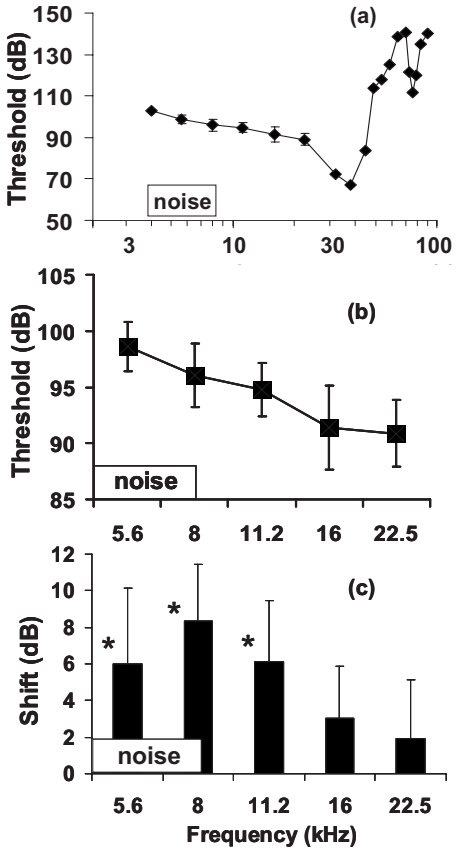


FIG. 3. (a) Audiogram of Boris. (b) Mean hearing thresholds (±1 SD) of Boris for 5.6–22.5 kHz in SPL (dB re 1 μPa). The fatiguing noise band, from 4 to 8 kHz, is depicted at the bottom, relative to the hearing thresholds tested. (c) Mean (±1 SD) amount of TTS (dB re 1 μPa) at each of the five frequencies tested. The * indicates that 5.6, 8, and 11.2 kHz demonstrated mean shifts that were significantly greater than their average threshold (one-way ANOVA and Tukey's pairwise comparison, $p<0.001$, $F_{5,82}=16.69$).

No TTS was measured after control conditions in which no sound was presented. The experimental matrix for all exposures ⩾160 dB and occurrences of TTS are listed in Table I.

A. Threshold shifts and frequency

Temporary threshold shifts were recorded in 26 of 57 noise exposure sessions and at all frequencies tested [Fig. 3(b)]. However, TTS at the higher more sensitive frequencies (16 and 22.5 kHz) did not occur frequently, and no significant increases in the mean threshold postnoise exposure were recorded at those frequencies. Mean postexposure thresholds for 5.6, 8, and 11.2 kHz demonstrated significant increases in hearing thresholds, relative to the baseline hearing threshold (one-way ANOVA, $F_{5,82}=16.69$, $p<0.001$; subsequent Tukey's pairwise comparison). These frequencies fell immediately within the center frequency of the fatiguing noise (5.6 kHz) and one-half (8 kHz) and one octave above the noise center frequency. The greatest mean threshold shift oc-

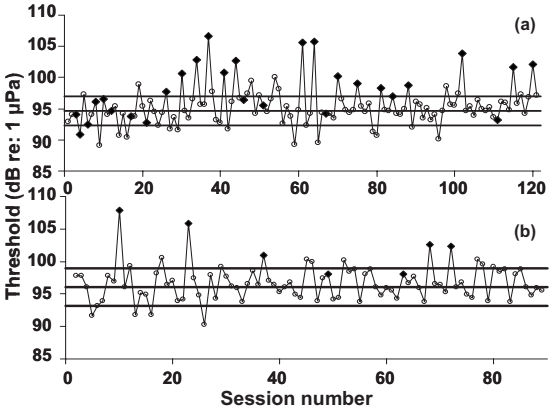


FIG. 4. Thresholds over the duration of the experiment for (a) 11.2 kHz and (b) 8 kHz. Black diamonds indicated thresholds measured after a fatiguing noise exposure. The middle line indicates mean threshold, and top and bottom lines indicate ±1 SD from the mean.

TABLE I. Noise exposure experimental matrix where the fatiguing sound was octave-band white noise of 4–8 kHz. Parameters listed include audiogram test frequency (kHz), exposure duration (min), noise SPL, SEL, and number of exposures and number of shifts at corresponding test conditions.

Frequency (kHz)	Duration (min)	SPL (dB re 1 μPa)	SEL (dB re 1 μPa^2 s)	(No. of exposures-no. of shifts)
5.6	30	160	192.5	(1-1)
	15	163	192.5	(1-1)
	7.5	166	192.5	(1-1)
	3.75	169	192.5	(1-0)
	3.75	175	198.5	(1-0)
	1.875	172	192.5	(1-0)
	1.875	178	198.5	(1-0)
8	30	160	192.5	(1-1)
	15	163	192.5	(1-1)
	7.5	166	192.5	(1-1)
	3.75	169	192.5	(1-0)
	3.75	175	198.5	(1-1)
	1.875	172	192.5	(1-0)
	1.875	178	198.5	(1-1)
11.2	30	160	192.5	(1-1)
	15	163	192.5	(3-2)
	7.5	166	192.5	(3-3)
	5.625	172	197	(1-1)
	3.75	169	192.5	(1-0)
	3.75	172	195.5	(1-0)
	3.75	175	198.5	(3-1)
	1.875	172	192.5	(1-0)
	1.875	175	195.5	(1-0)
	1.875	178	198.5	(3-3)
16	30	160	192.5	(1-1)
	15	163	192.5	(1-1)
	7.5	166	192.5	(1-0)
	3.75	169	192.5	(1-0)
	3.75	175	198.5	(1-0)
	1.875	172	192.5	(1-0)
	1.875	178	198.5	(1-0)
22	30	160	192.5	(1-0)
	15	163	192.5	(1-1)
	7.5	166	192.5	(1-0)
	3.75	169	192.5	(1-0)
	3.75	175	198.5	(1-0)
	1.875	172	192.5	(1-0)
	1.875	178	198.5	(1-0)

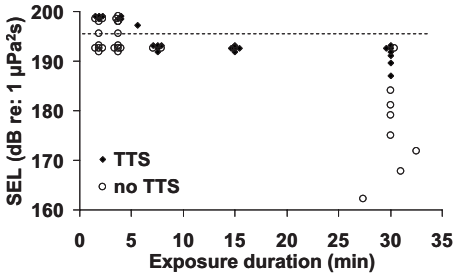


FIG. 5. Sound exposure levels required to induce TTS as the duration of exposure changes at all threshold frequencies tested. Shorter duration exposures required greater SEL to induce TTS. Dotted line indicates an equal-energy line of 195 dB SEL. Black diamonds indicate TTS occurrence; open circles indicate no TTS.

curred at 8 kHz (8.3 dB SPL) although 5.6 and 11.2 kHz both demonstrated an average TTS of 6 and 6.1 dB SPL, respectively.

Of all the thresholds measured at each selected frequency, TTS was found at 11.2 kHz after 51% of postnoise thresholds ($n=29$) and in 71% ($n=21$) of the postexposures when the SEL was >185 dB. Thresholds after noise exposure were measured in seven instances for each of the other frequencies tested, and TTS occurrences were found to be 43%, 71%, 29%, and 14% for 5.6, 8, 16, and 22.5 kHz, respectively.

B. Threshold shifts and sound exposure levels

Fatiguing sound exposure levels could be adjusted in at least two ways: either by altering the exposure duration or by varying the amplitude of the sound. Generally, as either exposure duration or sound pressure increased, if the other variable remained constant, a greater incidence of TTS could be expected. For example, at the start of the experiment, exposure duration was held at 30 min, but noise SPL and concomitant SELs were gradually increased (Fig. 5). Postnoise exposure thresholds did not demonstrate TTS until approximately 155 dB SPL, or 187 dB SEL. These sound levels induced a relatively small shift (3 dB SPL) at 11.2 kHz. In the following several sessions, as SEL was gradually increased to 192.5 dB (160 dB SPL), the amount of TTS at 11.2 kHz also increased [Fig. 6(a)].

One initial goal of the experiment was to examine the equal-energy hypothesis by keeping SEL constant and by varying exposure duration and sound pressure to determine the effects on TTS. Thus, after TTS was measured at all relevant threshold frequencies (5.6–22.5 kHz) using a 30-min, 160-dB SPL noise exposure, exposure duration was halved and sound pressure was increased by 3 dB, keeping SEL constant, and exposures and thresholds were measured again. However, TTS did not stay constant as exposure duration decreased. Threshold shift occurrence decreased from 80%–86% for 15–30 min exposures to 71% at 7.5 min and to zero significant TTS occurrences at 3.75 and 1.88 min. The amount of TTS also decreased with exposure duration [Fig. 6(a)] and could be predicted by a significant linear relationship ($r^2=0.77$; $p<0.001$; $y=0.34x+2.57$; $F_{1,12}=36.29$). Thus, for the same SELs, exposures of 30 min produced nearly 12 dB TTS, whereas 1.88 min of exposure did not generate a shift.

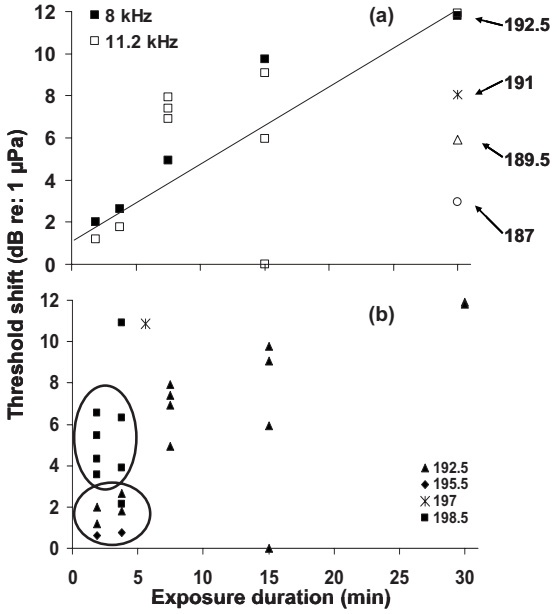


FIG. 6. (a) Amount of threshold shift measured at 8 kHz (closed symbols) and 11.2 kHz (open symbols) for SELs of 192.5 dB and below. Increasing SELs and corresponding symbols are labeled at the right, demonstrating that at constant exposure duration but increasing SEL, TTS increased. A regression of the 192.5 dB SEL data was used to illustrate that with decreasing noise exposure duration, TTS decreased when SEL was held constant ($r^2=0.77$, $p<0.001$, $y=0.34x+2.57$, $n=13$). (b) Amount of TTS for SELs of 192.5 dB and higher. Threshold frequencies are not discriminated, but SEL symbols are labeled to the right. Note that at shorter durations the amount of shift is clustered and higher SELs (top circle) are required to induce significant TTS.

In order to induce TTS at durations of 1.88–3.75 min, fatiguing noise SEL had to be increased from 192.5 to 198.5 dB. Intermediate levels at 195.5 dB SEL did not induce significant amounts of TTS. At 198.5 dB SEL, significant TTS was induced at 8 and 11.2 kHz in seven of eight exposures, with a mean shift of 5.4 dB and one shift of 11 dB [Fig. 6(b)]. Overall, no significant relationship was found between the amount of shift (measured at 5.6, 8, and 11.2 kHz) and SEL ($r^2=0.02$; $p=0.42$). Using the situations in which SEL was increased to induce TTS, threshold shift growth was predicted, revealing a strong positive relationship between SELs and amount of TTS ($r^2=0.96$; $p<0.001$; TTS=−0.702+1.36 (SEL); $F_{2,18}=112.2$).

C. Recovery from threshold shifts

Following noise exposures, AEP measurements were recorded for up to 80 min afterward to track the subject's recovery. Following all noise exposure sessions, the subject fully recovered to baseline values within 80 min. When shifts occurred, recovery to within ±1 SD of the baseline thresholds was typically seen in 20 min (15/26). In only three instances was the subject not within 1 SD of the baseline threshold values by 40 min after the noise exposure. These were either 15- or 30-min exposures at a SEL of 192.5 dB. The shifts were 12, 9.3, and 9.6 dB for 11.2, 11.2, and 5.6 kHz, respectively. Total recovery was rare within

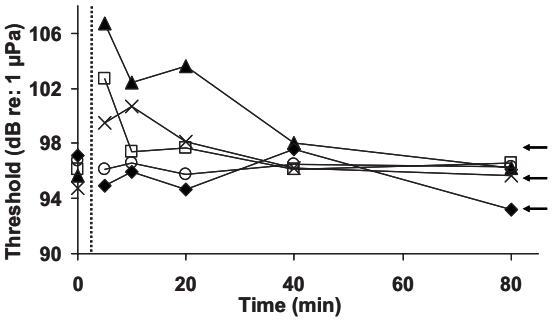


FIG. 7. Thresholds at 11.2 kHz before and after exposures for 192.5 dB SEL for five exposure durations (min): 30 (triangle), 15 (X), 7.5 (open square), 3.75 (open circle), and 1.875 (diamond). Thresholds were measured 5, 10, 20, 40, and 80 min after noise ended. The dotted line indicates noise exposure. Arrows indicate mean threshold ±1 SD.

10 min of the noise exposure following shifts, only occurring on three occasions, with TTSs of 3.0, 4.7, and 5.8 dB.

Examples of recovery functions after noise exposure are plotted in Fig. 7. In the figure, despite a constant SEL of 192.5, the recovery functions varied for each respective exposure duration and sound pressure session. While the greatest shift was found at the 30-min exposure, the 7.5-min exposure demonstrated greater TTS than the 15 min exposure. However recovery from the 7.5-min exposure was most rapid. Neither the 3.75- nor the 1.88-min exposures induced threshold shifts.

Recovery rates followed a logarithmic function, which was held relatively constant across various methods of analysis (Fig. 8). Greater shifts demonstrated initially steeper slopes of recovery, and lesser shifts reflected more gradual recovery rates. This was best seen by breaking the recovery functions into separate groups (>7 dB shifts, <7 dB shifts, ⩾15 min exposure, and <15 min exposure). The slopes of these groups varied somewhat but were roughly similar and were calculated as −7.4, −5.6, −6.3, and −6.7 log(min), respectively, and all were linear in log time. The somewhat steeper slopes of >7 dB shifts and ⩾15 min exposure may reflect the greater mean shifts of those groups, 8.3 and 7.2 dB SPL, respectively. The shallower slope of the <7 dB shift group was a result of the relatively lower TTS values (mean=4.6 dB SPL) of the grouping. On average, all recoveries could be approximated by a function of −1.8 dB/doubling of time.

D. Behavioral reactions

The subject's behavior during both the fatiguing noise and the control sessions was monitored and recorded by an assistant; for analysis, behavioral comparisons were made between the control and exposure sessions. Sessions were also divided into groupings of those with 30, 15, ⩽7.5 min exposures, which were rough proxies for long, moderate, and short duration exposures. No significant difference in overt behavioral changes was found between the exposure and control sessions. However, the subject did exhibit a significant increase in respirations during noise sessions as compared to the control sessions for the longer duration exposures and during noise exposures overall [Fig. 9(a)] (one-

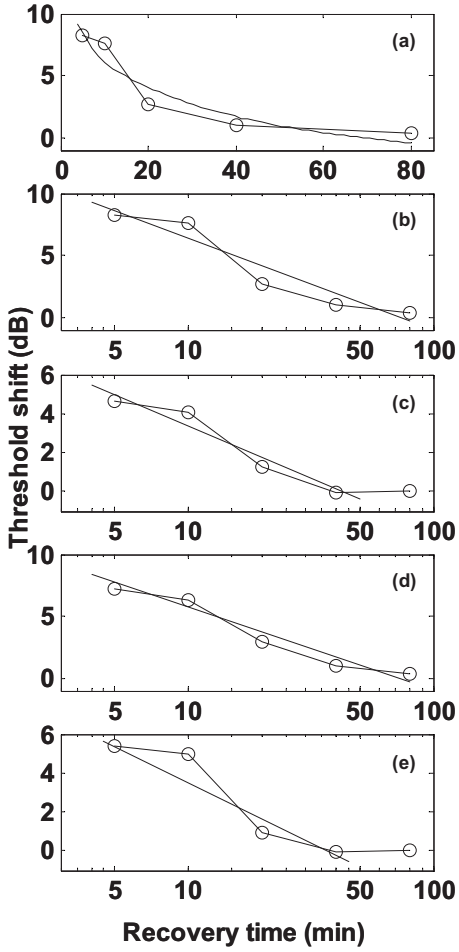


FIG. 8. Threshold shift (dB re 1 μPa) recovery functions demonstrating linear logarithmic recoveries in log time across multiple analysis methods: (a) TTS>7 dB (linear time), (b) TTS>7 dB (log time), (c) TTS<7 dB, (d) TTS from longer duration exposures (30–15 min), and (e) TTS from shorter duration exposures (7.5–1.875 min). The rate of recovery in all cases was approximately −1.8 dB/doubling of time.

way ANOVA and subsequent Tukey's pairwise comparison, $p<0.001$, $F_{7,4557}=31.98$). Mean respirations (±SD) for these longer durations and all control exposures were 6.14 (±3.21) and 5.98 (±3.10), respectively, and 8.81 (±2.99) and 7.31 (±3.59) for the equivalent noise exposures. The delay from the intertrial station to the hoop/noise exposure station was also significantly greater during noise exposure sessions, and this was across the groupings of exposure duration ($p<0.001$, $F_{7,4559}=25.31$). Mean values for these differences for the groups of 30, 15, ⩽7.5 min, and all control trials were 6.61 (±4.48), 6.01 (±2.85), 5.92 (±3.64), and 6.41 s (±4.01 s), respectively. For the noise sessions, delay means were 9.28 (±5.50), 8.02 (±3.51), 8.00 (±3.97), and 8.77 s (±4.95 s) for the respective groups.

IV. DISCUSSION

The data presented here provide a broad examination of the interaction of fatiguing noise duration and amplitude on TTS in a bottlenose dolphin. Shorter duration exposures were found to require greater amounts of energy (higher SELs) to induce similar amounts of TTS relative to longer duration exposures, indicating that a simple equal-energy ap-

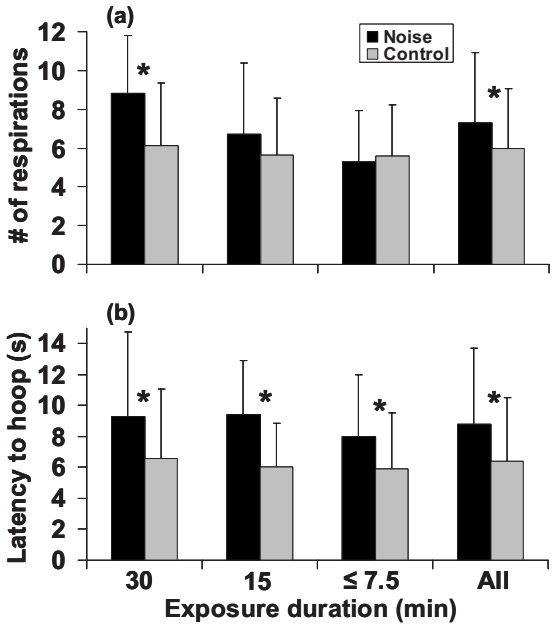


FIG. 9. Mean number of respirations (a) and delay from intertrial stationing pad to hoop station (b) for various noise exposure and control trials. For comparison, we grouped the sessions by the number of trials as well as summed data from all trials of all sessions. There were significantly greater respirations, indicated by the *, during noise sessions with >10 trials and all the sessions grouped together (one-way ANOVA and Tukey's pairwise comparison, $p<0.001$, $F_{7,4557}=31.98$). The hoop delay was significantly greater for all groups of noise exposure trials ($p<0.001$, $F_{7,4559}=25.31$).

proach does not adequately predict TTS in dolphins. Small behavioral changes in the number of respirations and delay to the exposure station were evident in noise exposure conditions versus control. Mean TTS levels induced here were relatively small compared to those often demonstrated in terrestrial mammals and some other marine mammals (Ward *et al.*, 1959; Finneran *et al.*, 2007; Kastak *et al.*, 2007). However, the data presented here demonstrate the SELs required for TTS onset across a range of exposure durations, and TTS onset can certainly be considered helpful in assessing the effects of noise on wild populations.

The frequency for which the greatest TTS was observed was one-half octave above the center frequency of the noise. Other significant levels of mean TTS were found at the center frequency of the noise and one octave above the center frequency. These trends reflect what has been demonstrated previously with the same animal in similar noise exposure conditions (Nachtigall *et al.*, 2004) as well as results collected with terrestrial mammals and pinnipeds (Ward *et al.*, 1959; Ward, 1962; Kastak *et al.*, 2005), indicating that the frequency trend of odontocete noise-induced TTS appears relatively conserved in marine mammals. The experiment was designed using frequencies below the regions of best auditory sensitivity for a bottlenose dolphin but similar to that of introduced anthropogenic noise. It is likely that at higher frequencies (20–100 kHz) and regions of better sensitivity, TTS levels might have been greater (Mills, 1982). Further, snapping shrimp produce high levels of background noise in Kaneohe Bay, creating an environment for masking [Fig. 2(b)]. In a quieter situation, we would have likely seen greater levels of TTS (Humes, 1980); thus these results may be somewhat conservative in the amount of TTS induced for the sound levels used. In quieter situations, TTS onset may actually occur at lower exposure levels than predicted by the present study. The dolphin also has relatively poor high frequency hearing (>50 kHz), which seems typical of adult male dolphins (Houser and Finneran, 2006). The animal's hearing within the range of the noise exposure (4–8 kHz) and test conditions (5.6–22.5 kHz) was relatively normal (Nachtigall *et al.*, 2000). There are no data to suggest that the hearing loss outside of the range of the TTS test conditions would affect the amount of TTS at these much lower frequencies examined here.

Exposure duration was also a consideration here. Sound exposure levels were calculated using hoop exposure times. However, for most sessions, the animal had to leave the experimental sound field to breathe. This procedure was not quite an intermittent exposure protocol as the animal was still exposed to noise while he was at the surface, although noise levels were 10 dB SPL lower at the surface station. Because these surface interval durations were minimized and surface sound levels were lower, this protocol did not significantly change SELs (e.g., 192.5 versus 192.7 dB SEL for 30 min exposures, 192.5 versus 192.7 dB for 15 min, 192.5 versus 192.6 dB for 7.5 min, 192.5 versus 192.6 dB for 3.75 min). Thus, it was possible that some very minor TTS recovery occurred during surface intervals, although this was unlikely based on the consistency of SELs and maintenance of relatively high SPLs (150–165 dB) (Ward, 1991).

Based on the SELs that induced TTS, the data were evaluated to determine a model that would predict TTS onset. Several algorithms have been suggested to predict noise levels that induce TTS including the general equal-energy rule, which assumes that as long as noise exposure energy levels are constant, similar threshold shifts will be induced regardless of the noise temporal pattern (Ward *et al.*, 1959). This is often termed the "3-dB rule" as a halving of sound exposure duration and a 3-dB increase in sound intensity maintains a constant energy level and should theoretically induce similar shifts (Kryter *et al.*, 1966). This rule has been employed for human standards (NIOSH, 1998) and has recently been proposed for use in predicting TTS in odontocetes (Finneran *et al.*, 2005). However, despite the fact that it is applied as a general rule and fits in many situations, empirical studies often do not support the equal-energy hypothesis as an accurate means to predict TTS, demonstrating that the trade-off between time and energy is not necessarily linear (Buck *et al.*, 1984; Ward, 1991; Hamernik and Qui, 2001).

Our data generally follow an equal-energy line of 195 dB SEL; however, the TTS instances more often split the line rather than fall upon it (Fig. 5). At shorter duration exposures, such as might occur with a single sonar ping exposure, greater SELs were required to induce TTS relative to longer duration exposures. Further, when SEL was held constant (sound duration decreased but SPLs increased), the amount of TTS did not hold constant as would be predicted by an equal-energy hypothesis [Fig. 6(a)]. Rather, TTS levels also decreased. At shorter duration exposures, increased SELs were required to induce significant levels of TTS [Fig.

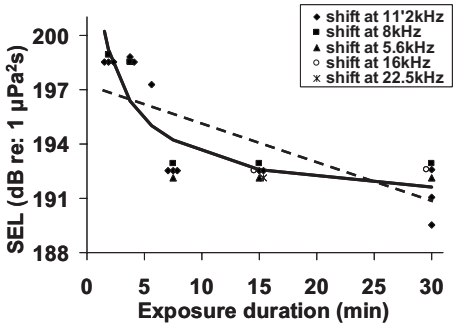


FIG. 10. The SELs required to induce TTS at higher noise levels where threshold frequencies are discriminated. Just TTS occurrence is plotted. Two methods to predict TTS are also graphed: a linear estimation of TTS (dotted line) (r^2=0.57, $p<0.001$, SEL=−0.21T+197.14, n=23) and a logarithmic estimation (r^2=0.78, $p<0.001$, SEL=−6.17 log(T)+200.21, n=23). Note that in both cases, the slope is positive, indicating that for shorter time durations, greater energy is required to induce TTS.

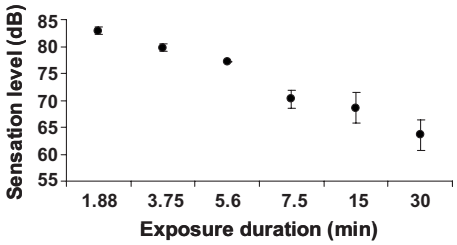


FIG. 11. SL (dB re threshold) at which TTS was induced for various exposure durations. TTS could be induced at much lower SLs for longer duration exposures. Note that the abscissa is categorical.

6(b)]. This noncoherence to the equal energy rule was previously demonstrated in pinnipeds (Kastak *et al.*, 2005; Kastak *et al.*, 2007). These data also indicate that equal-energy levels do not induce similar TTS levels in odontocete cetaceans.

To better predict TTS onset, a model of increasing energy as exposure level decreases appears to fit these data more closely (Fig. 10). Described is both a linear relationship to the threshold shift data, as well as a logarithmic relationship. The linear model does approximate the trend sufficiently well (r^2=0.57, $p<0.001$, SEL=−0.21(T)+197.14, $F_{1,22}$=28.65), where T=duration of the exposure in minutes. However, the logarithmic estimation does a much better job of predicting the threshold shifts and would apparently make a better model for predicting TTS (r^2=0.78, $p<0.001$, SEL=−6.17 log(T)+200.21, $F_{2,21}$=78.71).

Interestingly, although the equal-energy model is often proposed to predict the occurrence of human TTS, there is often a contradiction to this rule, and this is found across taxa, including in humans (Mills *et al.*, 1981), guinea pigs (Buck *et al.*, 1984), and chinchillas (Ward, 1991). Threshold shift occurrence depends on many factors in addition to fatiguing noise energy; these include frequency, intensity, duration, and time intervals of exposure (Bohne and Clark, 1990). Therefore using energy and SEL alone is an insufficient metric for predicting TTS. Although SEL combines both exposure SPLs and duration, these two factors do not necessarily contribute equally to the TTS onset. For example, it is possible that mechanisms that work to reduce TTS for short duration exposures fatigue are reduced for longer duration exposures. It is vital to present both SPL and duration in reporting and predicting TTS. Further, as more information is collected regarding odontocete TTS, it becomes increasingly obvious that the subject of TTS is quite complex. The best way to predict TTS may be to investigate it using a range of variables to determine what exposures produce the same TTS and address an equal-TTS based approach to reducing deleterious noise exposures (Kryter *et al.*, 1966; Ward, 1991).

When TTS onset was examined relative to sensation level (SL), or the difference in decibel between the fatiguing noise SPL and the average threshold, two points were relatively apparent (Fig. 11). First, there was a relationship between SL and exposure duration. As exposure duration decreased, lower SLs were needed to induce TTS. The clarity of the trend may be a means of predicting TTS based on SLs and duration and deserves further investigation. As the trend was negative, it may be important to determine the SLs that induce TTS at much longer duration exposures and if or at what SLs the trend asymptotes. Such predictions may further the understanding of effective quiet (the level at which no duration of exposure will induce TTS) for marine mammals. Second, the SLs of TTS onset for longer duration exposures were similar to SLs of shipping and snapping shrimp noise found in some natural habitats (Johnson, 1967; Erbe, 2002; Aguilar de Soto *et al.*, 2006; Lammers *et al.*, 2008). In other words, both shipping and snapping shrimp sounds may have enough energy to induce TTS if the animal is exposed for a long enough duration. This result implies caution for introducing a constant of noise into the marine environment, but we stress that such sound levels may be naturally encountered by some animals, in some instances.

When considering the behavioral data, it is important to realize that these experiments were primarily designed as hearing investigations, not as investigations of behavioral effects of noise. Thus some caution must be taken when interpreting behavior during these studies and extrapolating to other conditions. We assume that behavioral changes associated with noise exposure indicate an aversion to the fatiguing noise, but this is *only* an assumption. There were no overt behavioral changes that were significantly associated with the presence of the fatiguing noise. However, the dolphin was reinforced throughout his extensive research training and husbandry experience to limit any deviations from expected procedures, and in the interest of the present study, such considerations were taken here. Thus, major behavioral changes were unexpected. However, more subtle changes such as significant increases in respiration rates and delay from the intertrial station to the hoop station were observed (Fig. 9). Fatiguing noise levels at the surface intertrial station were 10 dB lower in SPL then at the hoop. Presumably, this may have been a passive method of deferring noise exposure by the dolphin. Interestingly, only the longer duration exposures reflected higher respiration rates, perhaps implying that potentially the shorter duration exposures, although higher in SPL, may have been less adverse. The only occasion on which the dolphin exhibited an obvious reaction to fatiguing noise exposure was when an amplifier electrically shorted during the exposure, creating an unplanned, unusual, and

relatively loud sound. The dolphin immediately pulled out of the hoop and swam vigorously around the noise exposure pen, burst-pulsing and jaw popping at the transducer and not heeding the trainer for several minutes. However, this was an isolated event and typically the animal stationed properly and observed the trainers' cues. The animal reacted to the unexpected in an unusual way.

Previous TTS studies have often demonstrated more obvious behavioral reactions to noise exposure (Kastak *et al.*, 1999; Schlundt *et al.*, 2000). However, it seems more often that there are only small behavioral changes as these animals are exceptionally well trained. Thus, motivation regarding certain tasks can override presumed reactions to potentially adverse stimuli. Such overriding motivations have been found in wild individuals, for example, pinnipeds raiding fish farms that are equipped with acoustic harassment devices (Quick *et al.*, 2004). The prevalence of such behaviors both in wild and captive animals deserves further attention. Further, our documented behavioral changes were subtle, indicating the importance of detailed observations when examining the effects of noise on marine mammals.

Hearing recovery rates generally followed a logarithmic trend of −1.8 dB recovery/doubling of time. This was similar to some previous results, particularly for shifts that are not greater than 10–15 dB (Nachtigall *et al.*, 2004; Finneran *et al.*, 2007). However, there was variability in the slopes, and greater shifts reflected faster recovery rates. A similar trend was found by Finneran *et al.* (2007), which is not only a confirmation of the logarithmic recovery functions but indicates robustness of both data sets. An interesting note is that the shifts that took 40 min to demonstrate recovery were all of longer duration exposures. This may support the trend that longer duration exposures will often induce greater amounts of TTS, which concurrently requires a greater amount of time for recovery. Similarly, in humans, recovery functions depend somewhat on the exposure situation, and longer duration exposures have demonstrated longer recovery times (Mills, 1982). This may also be demonstrated here. Further, these longer recoveries were recorded on shifts measured at 5.6 and 11.2 kHz, frequencies that did not produce the highest mean TTS. If these shifts were recorded at 8 kHz, greater shifts and longer recoveries might have been found. This underlies the importance of measuring multiple frequencies simultaneously, an important AEP advancement when investigating TTS (Finneran *et al.*, 2007).

V. CONCLUSIONS

This work demonstrates TTS onset in an odontocete across a range of exposure durations and sound levels, and the results indicate that shorter duration exposures often require greater sound energy to induce TTS than longer duration exposures. Recovery functions were relatively consistent but did show some indications that different exposure situations may relate to different recovery rates. The sample size was limited, but repetitive exposures and comparisons between studies increased the robustness of the data. These results are inconsistent with an equal-energy model of TTS supporting the notion that, as in terrestrial mammals, predicting odontocete TTS is quite complicated. It is suggested that TTS onset is considered sufficient to conclude physiological effects of noise exposure in marine mammals. Future investigations should continue to explore the range of variables that relate to threshold shifts to develop an equal-TTS model to better predict and mitigate situations in which anthropogenic noise may affect marine mammals.

ACKNOWLEDGMENTS

The authors would like to thank the numerous assistants that participated in this project including Jen Schultz, Vanessa Redigulo, Nina Spontak, Chiyo Nagata, Jacquline Krause-Nerhing, Sara Stieb, Nadav Shashara, Ariel Rivera, and Eric Litmann. Amy Apprill provided a much needed extra set of hands at crucial times. Additional animal training was conducted by Dera Look. Adam Pack also made helpful training suggestions. The prior work and advice of James Finneran and David Kastak were important. Help from the students and co-workers of the MMRP Laboratory were vital, including that of Vinny DePaolo, Michelle Yuen, Alison Stimpert, and Linda Choy. This work was funded by the Office of Naval Research Grant No. 00014-098-1-687 to P.E.N. and the support of Bob Gisiner and Mardi Hasting is noted. Additional support came from SeaSpace to T.A.M.

Aguilar de Soto, N., Johnson, M., Madsen, P. T., Tyack, P. L., Bocconcelli, A., and Borsani, J. F. (**2006**). "Does intense ship noise disrupt foraging in deep-diving Cuvier's beaked whales (*Ziphius cavirostris*)?," Marine Mammal Sci. **22**, 690–699.

Bohne, B., and Clark, W. (**1990**). "Studies of noise-induced hearing loss using an animal model," Hear. Instrum. **41**, 13–16.

Buck, K., Dancer, A., and Franke, R. (**1984**). "Effect of temporal pattern of a given noise dose on TTS in guinea pigs," J. Acoust. Soc. Am. **76**, 1090–1097.

Erbe, C. (**2002**). "Underwater noise of whale-watching boats and potential effects on killer whales (*Orcinus orca*), based on an acoustic impact model," Marine Mammal Sci.huma **18**, 394–418.

Finneran, J. J., Carder, D. A., Schlundt, C. E., and Ridgeway, S. H. (**2005**). "Temporary threshold shift in bottlenose dolphins (*Tursiops truncatus*) exposed to mid-frequency tones," J. Acoust. Soc. Am. **118**, 2696–2705.

Finneran, J. J., and Houser, D. S. (**2006**). "Comparison of in-air evoked potential and underwater behavioral hearing thresholds in four bottlenose dolphins (*Tursiops truncatus*)," J. Acoust. Soc. Am. **119**, 3181–3192.

Finneran, J. J., Schlundt, C. E., Branstetter, B. K., and Dear, R. (**2007**). "Assessing temporary threshold shift in a bottlenose dolphin (*Tursiops truncatus*) using multiple simultaneous auditory evoked potentials," J. Acoust. Soc. Am. **122**, 1249–1264.

Finneran, J. J., Schlundt, C. E., Dear, R., Carder, D. A., and Ridgway, S. H. (**2002**). "Temporary shift in masked hearing thresholds in odontocetes after exposure to single underwater impulses from a seismic watergun," J. Acoust. Soc. Am. **111**, 2929–2940.

Hamernik, R. P., and Qui, W. (**2001**). "Energy-independent factors influencing noise-induced hearing loss in the chinchilla model," J. Acoust. Soc. Am. **110**, 3163–3168.

Hamernik, R. P., Qui, W., and Davis, B. (**2003**). "Cochlear toughening, protection, and potentiation of noise-induced trauma by non-Gaussian noise," J. Acoust. Soc. Am. **113**, 969–976.

Houser, D. S., and Finneran, J. J. (**2006**). "Variation in the hearing sensitivity of a dolphin population determined through the use of evoked potential audiometry," J. Acoust. Soc. Am. **120**, 4090–4099.

Humes, L. E. (**1980**). "Temporary threshold shift for masked pure tones," Audiology **19**, 335–345.

Johnson, C. S. (**1967**) "Sound detection thresholds in marine mammals," in *Marine Bioacoustics*, edited by W. N. Tavolga (Pergamon, New York), pp. 247–260.

Kastak, D., Reichmuth, C. J., Holt, M. M., Mulsow, J., Southall, B. L., and Schusterman, R. J. (**2007**). "Onset, growth, and recovery of in-air tempo-

rary threshold shift in a California sea lion (*Zalophus californianus*)," J. Acoust. Soc. Am. **122**, 2916–2924.
Kastak, D., Schusterman, R. J., Southall, B. L., and Reichmuth, C. J. (**1999**). "Underwater temporary threshold shift induced by octave-band noise in three species of pinniped," J. Acoust. Soc. Am. **106**, 1142–1148.
Kastak, D., Southall, B. L., Schusterman, R. J., and Kastak, C. R. (**2005**). "Underwater temporary threshold shift in pinnipeds: Effects of noise level and duration," J. Acoust. Soc. Am. **118**, 3154–3163.
Kryter, K. D., Ward, W. D., Miller, J. D., and Eldredge, D. H. (**1966**). "Hazardous exposure to intermittent and steady-state noise," J. Acoust. Soc. Am. **39**, 451–464.
Lammers, M. O., Brainard, R. E., Au, W. W. L., Mooney, T. A., and Wong, K. (**2008**). "An ecological acoustic recorder (EAR) for long-term monitoring of biological and anthropogenic sounds on coral reefs and other marine habitats," J. Acoust. Soc. Am. **123**, 1720–1728.
Mills, J. H. (**1982**). "Effects of noise on auditory sensitivity, psychophysical tuning curves, and suppression," in *New Perspectives on Noise-Induced Hearing Loss*, edited by R. P. Hamernik, D. Henderson, and R. J. Salvi (Raven, New York), pp. 249–263.
Mills, J. H., Adkins, W., and Gilbert, R. (**1981**). "Temporary threshold shifts produced by wideband noise," J. Acoust. Soc. Am. **70**, 390–396.
Mooney, T. A., Nachtigall, P. E., and Yuen, M. M. L. (**2006**). "Temporal resolution of the Risso's dolphin, *Grampus griseus*, auditory system," J. Comp. Physiol., A **192**, 373–380.
Mulroy, M. J. (**1986**). "Permanent noise-induced damage to stereocilia: A scanning electron microscopy study of the lizard's cochlea," Scan Electron Microsc. **4**, 1451–1457.
Nachtigall, P. E., Lemonds, D. W., and Roitblat, H. L. (**2000**). "Psychoacoustic studies of dolphin and whale hearing," in *Hearing by Whales and Dolphins*, edited by W. W. L. Au, A. N. Popper, and R. J. Fay (Springer-Verlag, New York), pp. 330–363.
Nachtigall, P. E., Mooney, T. A., Taylor, K. A., Miller, L. A., Rasmussen, M., Akamatsu, T., Teilmann, J., Linnenschidt, M., and Vikingsson, G. A. (**2008**). "Shipboard measurements of the hearing of the white-beaked dolphin, *Lagenorynchus albirostris*," J. Exp. Biol. **211**, 642–647.
Nachtigall, P. E., Mooney, T. A., Taylor, K. A., and Yuen, M. M. L. (**2007**). "Hearing and auditory evoked potential methods applied to odontocete cetaceans," Aquat. Mamm. **33**, 6–13.
Nachtigall, P. E., Pawloski, J. L., and Au, W. W. L. (**2003**). "Temporary threshold shifts and recovery following noise exposure in the Atlantic bottlenosed dolphin (*Tursiops truncatus*)," J. Acoust. Soc. Am. **113**, 3425–3429.
Nachtigall, P. E., Supin, A. Y., Pawloski, J. L., and Au, W. W. L. (**2004**). "Temporary threshold shifts after noise exposure in the bottlenose dolphin (*Tursiops truncatus*) measured using evoked auditory potentials," Marine Mammal Sci. **24**, 673–687.
National Academy of Sciences. (**2003**). *Ocean Noise and Marine Mammals* (National Academies, Washington, DC).
National Academy of Sciences. (**2005**). *Marine Mammal Populations and Ocean Noise: Determining When Noise Causes Biologically Significant Effects* (National Academies, Washington, DC).
NIOSH. (**1998**). *Criteria for a Recommended Standard: Occupational Noise Exposure, Revised Criteria 1998* (NIOSH, Cincinnati, OH).
Popov, V. V., Supin, A. Y., Wang, D., Wang, K., Xiao, J., and Li, S. (**2005**). "Evoked-potential audiogram of the Yangtze finless porpoise *Neophocaena phocaenoides asiaeorientalis* (L)," J. Acoust. Soc. Am. **117**, 2728–2731.
Popper, A. N., and Clarke, N. S. (**1976**). "The auditory system of goldfish (*Carassius auratus*): Effects of intense acoustic stimulation," Comp. Biochem. Physiol. A **53A**, 11–18.
Quick, N. J., Middlemas, S. J., and Armstrong, J. D. (**2004**). "A survey of antipredator controls at marine salmon farms in Scotland," Aquaculture **230**, 169–180.
Richardson, W. J., Greene, C. R., Jr., Malme, C. I., and Thomson, D. H. (**1995**). *Marine Mammals and Noise* (Academic, San Diego).
Saunders, J. C., and Dooling, R. (**1974**). "Noise-induced threshold shifts in the parakeet (*Melopsittacus undulatus*)," Proc. Natl. Acad. Sci. U.S.A. **71**, 1962–1965.
Schlundt, C. E., Finneran, J. J., Carder, D. A., and Ridgway, S. H. (**2000**). "Temporary shift in masked hearing thresholds of bottlenose dolphins, *Tursiops truncatus*, and white whales, *Delphinapterus leucas*, after exposure to intense tones," J. Acoust. Soc. Am. **107**, 3496–3508.
Southall, B. L., Bowles, A. E., Ellison, W. T., Finneran, J. J., Gentry, R. L., Greene, C. R., Jr., Kastak, D., Ketten, D. R., Miller, J. H., Nachtigall, P. E., Richardson, W. J., Thomas, J. A., and Tyack, P. L. (**2008**). "Marine mammal noise exposure criteria: Initial scientific recommendations," Aquat. Mamm. **33**, 411–509.
Supin, A. Y., and Popov, V. V. (**1995**). "Envelope-following response and modulation rate transfer function in the dolphin's auditory system," Hear. Res. **92**, 38–45.
Supin, A. Y., Popov, V. V., and Mass, A. M. (**2001**). *The Sensory Physiology of Aquatic Mammals* (Kluwer Academic, Boston).
Ward, W. D. (**1962**). "Damage-risk criteria for line spectra," J. Acoust. Soc. Am. **34**, 1610–1619.
Ward, W. D. (**1991**). "The role of intermittence in PTS," J. Acoust. Soc. Am. **90**, 164–169.
Ward, W. D., Cushing, E. M., and Burns, E. M. (**1976**). "Effective quiet and moderate TTS: Implications for noise exposure standards," J. Acoust. Soc. Am. **58**, 160–165.
Ward, W. D., Glorig, A., and Sklar, D. L. (**1958**). "Dependence of temporary threshold shift at 4 kcon intensity and time," J. Acoust. Soc. Am. **30**, 944–954.
Ward, W. D., Glorig, A., and Sklar, D. L. (**1959**). "Temporary shift from octave-band noise: Applications to damage-risk criteria," J. Acoust. Soc. Am. **41**, 522–528.
Wartzok, D., Popper, A. N., Gordon, J., and Merrill, J. (**2004**). "Factors affecting the responses of marine mammals to acoustic disturbance," Mar. Technol. Soc. J. **37**, 6–15.
Yuen, M. M. L., Nachtigall, P. E., Breese, M., and Supin, A. Y. (**2005**). "Behavioral and auditory evoked potential audiograms of a false killer whale (*Pseudorca crassidens*)," J. Acoust. Soc. Am. **118**, 2688–2695.

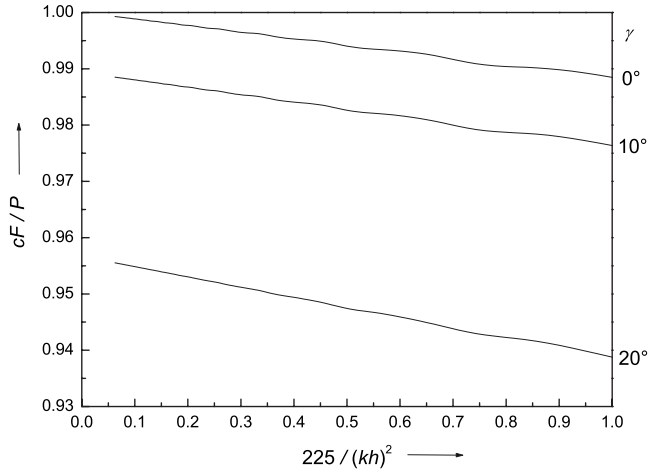
Erratum: Radiation force calculations for focused fields [J. Acoust. Soc. Am. 124, 1941–1949 (2008)]

K. Beissner
Physikalisch-Technische Bundesanstalt, 38116 Braunschweig, Germany

(Received 22 December 2008; accepted 24 December 2008)
[DOI: 10.1121/1.3075592]

PACS number(s): 43.25.Qp, 43.80.Vj, 43.10Vx

A printing error appears on page 1946 in the printed edition of JASA (the error does not appear in the online version). Jagged lines were introduced into Fig. 9 and above the figure itself during the printing process. The correct figure is printed below.



 0001-4966/2009/125(3)/1827/1/$25.00

ACOUSTICAL NEWS—USA

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

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

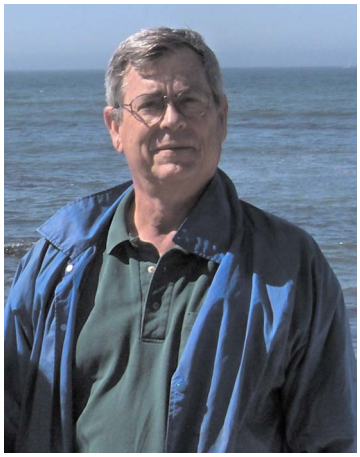
New Fellows of the Acoustical Society of America

Marshall Long—For contributions to acoustic design for the built environment

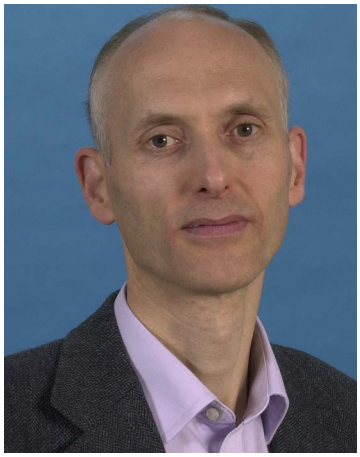
Christian Lorenzi—For contributions to the perception of amplitude modulation

Luc Mongeau—For contributions to the aeroacoustics of phonation

D. Lloyd Rice—For contributions to the assessment of technical advancement in acoustics



Patrick W. Moore—For contributions to dolphin bioacoustics and bio-inspired sonar



Michael A. Stone—For contributions to hearing aid gain control

 0001-4966/2009/125(3)/1828/5/$25.00

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.

2009

18–22 May	157th Meeting of the Acoustical Society of America, Portland, OR [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; Email: asa@aip.org; WWW: http://asa.aip.org].
24–28 June	5th International Middle-Ear Mechanics in Research and Otology (MEMRO), Stanford University, Stanford, CA [http://memro2009.stanford.edu].
26–30 October	158th Meeting of the Acoustical Society of America, San Antonio, TX [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; Email: asa@aip.org; WWW: http://asa.aip.org].

2010

19–23 April	158th Meeting of the Acoustical Society of America, Baltimore, MD [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; Email: asa@aip.org; WWW: http://asa.aip.org].
15–19 November	2nd Iberoamerican Conference on Acoustics (Joint Meeting of the Acoustical Society of America, Mexican Institute of Acoustics, and Iberoamerican Federation on Acoustics), Cancun, Mexico [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; Email: asa@aip.org; WWW: http://asa.aip.org].

Cumulative Indexes to the Journal of the Acoustical Society of America

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

Some indexes are out of print as noted below.

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

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

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

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

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

Volumes 36–44, 1964–968: Contemporary Literature. Classified by subject and indexed by author. Pp. 1060. Out of Print

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

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

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

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

Volumes 859–4, 1989–1993: JASA and Patents. Classified by subject and indexed by author and inventor. Pp. 736. Price: ASA members $30 (paperbound); Nonmembers $80 (clothbound)

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

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

Revision List

New Associates

Aldana, Guillermo 250 Beanviller Rd., Randolph, VT 05060

Andersen, Per, Lloyds Register ODS, Noise and Vibration, Titangade 15, 2200 Copenhagen, Denmark

Arbogast, Tanya 125 Nassau Ct, San Ramon, CA 94582

Aubee, Andrew 27 Beacon Dr., North Kingstown, RI 02852

Avis, Richard, Avis Design, 13648 SW Lauren Ln., Tigard, OR 97223

Bailey, Helen, NOAA, Environmental Research Division, 1352 Lighthouse Ave., Pacific Grove, CA 93950

Balla-Holden, Andrea 5988 SE Kelsey Court, Port Orchard, WA 98367

Barker, Thomas, 136 Sunset Ave, Vestal, NY 13850

Beane, Michelle 102 Lyme Rd., Hanover, NH 03755

Berninger, Erik, Karolinska University Hospital, Audiology, B 58, Huddinge, 141 86 Stockholm, Sweden

Bradley, Jeffrey P., 11 Hortense Place, St. Louis, MO 63102

Brette, Romain, Departement D'Etudes Cognitives, 29, Rue D'ulm, Paris 75005, France

Brumleve, John 912 Forest Ave, Milford, OH 45150

Butcher, Jon, St. Andrew's College, Physics, 15800 Yonge St., Aurora, ON L4G 3H7, Canada

Butko, Daniel 19171 W 207th Terrace, Spring Hill, KS 66083

Campbell, Ewan, Morgan Electroceramics, Engineering Development, Bursledon Rd., Thornhill, Southampton, SO15 2GQ, UK

Canady, Wyatt, Halliburton, WPS, 3000 N. Sam Houston Pkwy. E., J2T27, Houston, TX 77450

Christian, Andrew 293 11th St., Brooklyn, NY 11215

Chu, Raymond, American Inst. of Physics, One Physics Ellipse, College Park, MD 20740

Clair, Shaun, 205 East Second Ave., Lititz, PA 17543

Clifford, Paul, Program Management, Ultra Electronics Ocean Systems, 115 Bay State Dr., Braintree, MA 02184

Cochran, John, Raytheon, Advanced Technology, 1847 West Main Rd., Portsmouth, RI 02871

Cody, Kristin 11 Wood Duck Pl., Waterford, NY 12188

Coleman, Janie, 7302 Blue Springs Rd., Cleveland, TN 37311

Cooper, Chuck 25 Fairfield St., Newton, MA 02460

Cooper, Alan, Univ. of Texas, Applied Research Laboratories, P.O. Box 8029, Austin, TX 78713-8029

Coorman, Geert, Gen. Eisenhowerstr. 55 bus 1, 8520 Kuurne, Belgium

Cui, Weili 2628 Quail Ridge Rd., Endicott, NY 13760

Darling, Richard, 55 Railroad Row, White River Junction, VT 05001

De Santis, Eric #49—31450 Spur Ave., Abbotsford, BC V2 T 5M3, Canada

Di Betta, Anna Maria, Univ. of Sheffield, Human Communication Sciences, 31 Claremont Cresc., Sheffield, S10 1SW, UK

Duchamp, Daniel, Stevens Inst. of Technology, Computer Science, Castle Point on Hudson, Hoboken, NJ 07030

Dufour, Nicolas, DBXPL:smart pro sound solution, Rostoviou 10, 11526 Athens, Greece

Eguia, Manuel, Universidad Nacional deQuilmes, Laboratorio de Acustica, R S Pena 352, 1876 Bernal, Buenos Aires, Argentina

Etter, Walter 18 Maidenstone Dr., Ocean, NJ 07712

Every, Arthur, Univ. of Witwatersrand, Physics, 1 Jan Suts Ave., Johannesburg, WITS 2050, Republic of South Africa

Fox, John, Air King Ventilation Products, 820 Lincoln Ave, West Chester, PA 19380

Garner, Jimmy, Gulf Coast Community College, Visual and Performing Arts, 5230 West US 98, Panama City, FL 32401
Gavin, Joseph, Gulfstream Aerospace, Flight Sciences, 45 Sabal Dr., Richmond Hill, GA 31324
Gilbert, John 69 Naples Rd., Brookline, MA 02446
Griffin, Dana, Griffin Acoustical Engineering, 403 Chantecler Ave., Richmond, VA 23226
Guida, Edward, Engineering, Ultra Electronics Ocean Systems, 115 Bay State Dr., Braintree, MA 02184
Harvey, Scott, Phoenix Noise and Vibration, LLC, 930 North East St., Ste. #4, Frederick, MD 21707
Heavener, Don, Vis-A-Vis, Inc., 7936 Camino Cir., Miami, FL 33143
Hoshi, Kazuma 2-16-30-305, Funabashi, 2730021, Japan
Hosoda, Kosuke, Foster Electric Co., RandD Dept., 512, Miyazawa-cho, Akishima, Tokyo, 196-8550, Japan
Iannacci, Matthew 505 Aura Dr., Monroeville, PA 15146
Jaeger, Kaj Gronnegarden 558 B, 2670 Greve, Denmark
Kloser, Rudolf 144 Bayview Rd., Lauderdale 7021 Tasmania, Australia
Kulig, Arkadiusz, DJ COOLIG, 1053 64th St., Apt. A2, Brooklyn, NY 11219
Kurze, Philip, Whelen Engineering Company, Mass Notification Products, 51 Winthrop Rd., Chester, CT 06412
Lappas, Simon, Acoustic Logic Consultancy, 9 Sarah St., Mascot, Nsw, 2020, Australia
Lee, Richard 1526 SE North Blackwell Dr., Port St. Lucie, FL 34990
Lee, Keunhwa, Ocean Engineering, Bldg. 36, Rm. 306, San 56-1, Sillim, Kwanak, Seoul 151-744 South Korea
Lemmert, Ernest 2417 2nd St. S., Arlington, VA 22204
Lenz, Richard, RealAcoustix LLC, Design, 903 N. Marshall Way, Ste. A, Layton, UT 84041
Lewis, Matthew, UT Southwestern Medical Ctr., Advanced Radiological Sciences, 5323 Harry Hines Blvd., MC 9058, Dallas, TX 75390-9058
Logan, Tamara, Cisco Systems, Inc, IPCBU, 170 West Tasman Dr., SJC 32/1, San Jose, CA 95134
Macherey, Olivier, Medical Research Council, Cognition and Brain Sciences, 15 Chaucer Rd., Cambridge, CB27EF, UK
Macirowski, Thomas 799 Crandon Blvd., Apt. 1501, Key Biscayne, FL 33149
Mallory, Robert, Southwest Noise Control LLC, 615 Alvarado Dr. NE, Albuquerque, NM 87108-1645
Marquardt, Torsten, Univ. College London, UCL Ear Inst., 332 Gray's Inn Rd., London, WC1X 8EE, UK
Marsh, Bob, Technical Dept., 420 Corporate Blvd., Newark, DE 19702
Mc Donagh, Liam 271 Richmond Rd., Morningside, Brisbane, Qld, 4170, Australia
Mckee, Floyd 2150 Highway Nine, Newmarket, ON L3Y 4V9, Canada
McNally, Kevin, McKay Conant Hoover Inc., 5655 Lindero Canyon Rd., Ste. 325, Westlake Village, CA 91362
Misra, Ananya, Princeton Univ., Dept. of Computer Science, 35 Olden St., Princeton, NJ 08540
Moose, Robert, 509 CES/CECP, 6600 10th St., Ste. 211, Whiteman AFB, MO 65305
Munro, Miles, Exponent, Human Factors, 230 Goddard, Ste. 200, Irvine, CA 92618
Narasimhan, Rangan 1681 West 820 North, Provo, UT 84601
Newton, Jessica, 9600 Victoria Ln. #302, Naples, FL 34109
O'Connor, Robert, Midwestern Univ., Media Resources, 555 31st St., Downers Grove, IL 60515
Ortkiese, Tina 10846 Paladin Dr., Hampton, GA 30228
Pann, Heather 1809 Chandler, Ann Arbor, MI 48105
Perlman, Elizabeth 303 99th St. Apt. 5E, Brooklyn, NY 11209
Petersen, Cornelis, Bassett, Acoustics, Level 28, 91 King William Rd., Adelaide 5000 SA, Australia
Plishka, Lloyd, Lloyd Plishka Architect, 1595 West 68th Ave., Vancouver, BC V6P 2V4, Canada
Plumbley, Mark 82 Prebend St., London N1 8PR, UK
Prairie, Michael PO Box 453, Northfield, VT 05663
Pu, Youchun, MEDISON, PROBE RnD Center, Disscusser and MEDISON Bldg B1, 1003 Daechi-dong Gangnam-gu, Seoul, 135280, Korea
Radford, Craig, Univ. of Auckland, Leigh Marine Laboratory, PO Box 349, Warkworth 0941, New Zealand
Riley, Brian, Margenau Associates, Inc., 4473 Scarsdale Place, Boulder, CO 80301
Robinson, Threcia, 625 NW 60th St., Ste. C, Gainesville, FL 32607
Roland, William, Gracenote Consulting, Inc., 2120 4th St., Ste. 205, Meridian, MS 39301
Sanavi, Sharo, ABC Acoustics, Inc., Management, 990 Highland Dr., Ste. 110-C, Solana Beach, CA 92075
Scarton, Henry, Dept. of Mechanical, Aerospace and Nuclear Engrg., JEC 4008, San 56-1, Sillim, Kwanak, Troy, NY 12180
Shrestha, Mona, Far Sounder Inc., Research and Development, 43 Jefferson Blvd., Warwick, RI 02888
Simpson, Adrian, Universitaet Jena, Inst. f. Germ. Sprachw., Fuerstengraben 30, 07743 Jena, Germany
Smith, Heather 827 E 200 N, Springville, UT 84663
Snure, Harold, CBC Broadcasting, Inc., 523 East Sibley St., Hammond, IN 46320
Spaulding, Robert, Spaulding Associates, 12 Boothby Rd., Kennebunk, ME 04043
Swager, Bill, Brass on Ivory Music, 117B Mayo Rd., Edgewater, MD 21037
Todd, Victoria, Ocean Science Consulting Ltd., Ocean House, 4 Brewery Ln., Belhaven, Dunbar, East Lothia, EH42 1PD, UK
Van Luxemburg, Laurentius, Level Acoustics BV, De Rondom 10, 5612 AP Eindhoven, The Netherlands
Van Den Boogaart, C. Gregor, Univ. of Augsburg, Computer Science, Eichleitnerstrasse 30, 86159 Augsburg, Germany
Von Hunerbein, Sabine, Acoustic Research Centre, The Crescent, San 56-1, Sillim, Kwanak, Salford M54WT, UK
Widell, Alfred Nya Vattentornet, 22188, Lund, Sweden
Willems, Stefan, Interleuvenlaan 74-76, B-3001 Leuven, Belgium
Williams, Paul, Samaxus Ltd, Managing Director, PO Box 74073, Market Rd., Auckland 1534, New Zealand
Yu, Zhaohui, Purdue Univ., Mechanical Engineering, Herrick Labs, 140 S. Martin Jischke Dr., W. Lafayette, IN 47907
Zhou, Shengli, Univ. of Connecticut, Electrical and Computer Engineering, 371 Fairfield W. Way, Storrs, CT 06268

New Students

Allen, Phoebe, 4723 Sheboygan Ave., Apt. 117, Madison, WI 53705
An, Young-Ran, Stony Brook Univ., Linguistics, S-205, Social and Behavioral Sciences, Stony Brook, NY 11794-4376
Antfire, Dord, CHU, CSIE, Shinchu, Shin-Chu, Taiwan
Assink, Jelle, Univ. of Mississippi, NCPA, 1 Coliseum Dr., Office 1135, University, MS 38677
Aubin, Nicholas, 4600 SW 67 Ave., Apt. 243, Miami, FL 33155
Barringer, Jessica, Univ. of Hartford, Box 1993, 200 Bloomfield Ave, West Hartford, CT 06117
Battaglia, Michael, 1317 S. 10th St., Philadelphia, PA 19147
Behroozmand, Roozbeh, 1915 Maple Ave. Apt. 608, Evanston, IL 60201
Betancourt, Mavie, 17065 Heart of Palms Dr., Tampa, FL 33647
Birney, Jonathan, 516 N 900 Rd., Lawrence, KS 66047
Bradley, Daniel J., Penn State Acoustics Program, 201 Applied Science Bldg., University Park, PA 16802
Burns, Tom, 3071 Sunny Ayre Dr., Lansdale, PA 19446
Byrd, Andrew, 2200 SW 57th Ct, Miami, FL 33155
Caillat, Marjolaine, SMRU, Univ. of St. Andrews, St. Andrews, KY16 8LB, UK
Calderon, Jose, 6851 SW 94th St., Miami, FL 33156
Cano, Estefania, 8600 SW 67 Ave., Apt. 905, Miami, FL 33143
Chakraverty, Arindom, Purdue Univ., Electrical Engineering Student, 708 N Chauncey Ave., Apt. B, West Lafayette, IN 47906
Che Din, Nazli, Oita Univ., Architecture and Mechatronics, Dannoharu 700 Banchi, Oita, 870-1192, Japan
Coyle, Whitney, 5427 Foxwood Dr., Crestwood, KY 40014
Czekanski-Moir, Brady, Univ. of Miami, 753 Navarre Ave, Coral Gables, FL 33134
Dagovitz, Damien, 2959 N Talman #3, Chicago, IL 60618
D'Angio, Leon, 110 Arlington Ave, Linwood, NJ 08221
Deem, Robert, 1170 N Milwaukee Ave, Chicago, IL 60622
Deslauriers, Glen, 3226 Mary St., Apt. # 9, Miami, FL 33133
Dickerman, Ed, 150 Glen Dr., Warwick, RI 02889

Draper, William, 3033 Grotto Walk, Ellicott City, MD 21042
Eastland, Grant, Washington State Univ., Physics, 530 SW Fountain St. Apt. B, Pullman, WA 99163
Fox, Michael, 2544 Roberts Ct., Altoona, WI 54720
Gibraski, Aaron, 8446 W. 50th St., Mccok, IL 60525
Gibson, Allyson, Washington Univ., Physics, Compton Hall 244, 1 Brookings Dr., St. Louis, MO 63130
Glogovsky, Scott, 700 Biltmore Way, Apt. 916, Coral Gables, FL 33134
Gong, Zheng, Northeastern Univ., ECE, 409 Dana Research Ctr., 360 Huntington Ave., Boston, MA 02115
Grossmann, Aurora, 9420 Hooper Rd., Athens, OH 45701
Hamilton, Scott, 503 W. State, West Lafayette, IN 47906
Hayashi, Kyohei, Future University-Hakodate, Systems Information Science, 116-2, Kamedanakano-cho, Hakodate, 041-8655, Japan
Heimbauer, Lisa, 1527 Pangborn Station Dr., Decatur, GA 30033
Hoeman, Stephanie, 2111 Naismith, Lawrence, KS 66046
Holt, Fred, 445 Waupelani Dr., F27, State College, PA 16801
Huang, Jingyuan, 5000 Forbes Ave., Psychology Dept., Pittsburgh, PA 15232
Iqbal, Asim, The Ohio State Univ., Mechanical Engineering, 108 Ctr. for Auto Research, 930 Kinnear Rd., Columbus, OH 43212
Jakjoud, Hicham, 460 Lettre D Massira 2, 40000 Marrakech, Morocco
Jasinski, Christopher, 55 Buttonwood Cir., Cheshire, CT 06410
Jensen, Frants, Langelandsgade 208, 2.tv., 8200 Aarhus N, Denmark
Jung, Youngsook, MIT, EECS, 50 Vassar St. Rm 36-511, Speech Communication Group, Cambridge, MA 02139
Kim, Kibeom, Univ. of Delaware, Computer Science, 161-10 Thorn Ln., Newark, DE 19711
Kodera, Masashi, Toyama Gofukusuehiro 2581, Joinasu-suehiro 303, Toyama, Japan
Krahforst, Cecilia, 606 East 11 St., Apt. 328, Greenville, NC 27858
Lasarcyk, Eva, Saarland Univ., Phonetics and Phonology, Campus C7.2, Postfach 151150, 66041 Saarbruecken, Germany
Laughlin, Trevor, RR4, Box 504, Linton, IN 47441
Leblanc, Jared, 514 Kevin Dr., Lafayette, LA 70507
Lee, Pyoung Jik, Hanyang Univ., Architectural Engineering, Room 603 Science and Technology Bldg., Seongdong-gu, Seoul, 133-791, Korea
Lewandowski, Natalie, Krefelder Str. 26, 70376 Stuttgart, Germany
Li, Feipeng, 1006 N Broadway Ave., Urbana, IL 61801
Mandru, Andrada, Univ. of Louisiana, Physics, PO BOX 44210, Lafayette, LA 70504
Mann, Douglas, 7430 SW 59th CT, Apt. C3, South Miami, FL 33143
Matoza, Robin, IGPP 0225, U.C. San Diego, La Jolla, CA 92093-0225
Mclaughlan, James, Inst. of Cancer Research, Joint Dept. of Physics, The Royal Marsden NHSFT, Downs Rd., Sutton, SM2 5PT, UK
McMahon, Kara, 901 Peoples Ave. Apt. 1, Troy, NY 12180
Menon, Ravishankar, UC San Diego, SIO, 9500 Gilman Dr., Mail Code 0238, La Jolla, CA 92093
Meyer, Lindsay, 5151 Highland Rd., Apt. 298, Baton Rouge, LA 70808
Mitsuya, Takashi, Queen's Univ., Psychology, 62 Arch St., Humphrey Hall, Kingston, ON K7K1P1, Canada
Nandur, Vuday, 2711 NE 25th St., Ocala, FL 34470
Newcomb, Wesley, GCC 223, 200 Campus Dr., Grove City, PA 16127
Nishimiya, Kojiro, Univ. Tsukuba, Japan Ibaraki, Tsukuba City, Japan
Olson, Derek, Vassar College, Physics, 124 Raymond Ave., Poughkeepsie, NY 12604
Olson, Stephen, 700 Toftrees Ave., Apt. 304, State College, PA 16803
Ormonde, Christopher, Boston Univ., Mechanical Engineering, 110 Cummington St., AME Dept., Boston, MA 02215
Ostrowski, Angela, RWU Box# 7842, One Old Ferry Rd., Bristol, RI 02809
Paluga, Andrew, 2151 Kudza Rd., West Palm Beach, FL 33415
Paul, Stephan, Servidco Corintians 283, A401, Pantanal, FloriansPolis, Sc, 88040-100, Brazil
Perez, Timothy, 37 Salamanca Ave., Apt. C, Coral Gables, FL 33134
Pinilla, Carlos, 6001 SW 70th St. Apt. 650, South Miami, FL 33143
Pinnow, Eleni, Binghamton Univ., Psychology, PO Box 6000, Binghamton, NY 13902
Plazas, Daniel, 18744 NW 89th Ct., Hialeah, FL 33018
Porupski, Dunja, Cvijete Zuzoric 51, Zagreb, Croatia
Radhakrishnan, Kirthi, Univ. of Cincinnati, Biomedical Engineering, 231 Albert Sabin Way, 6152 Medical Sciences Bldg., Cincinnati, OH 45267-0586
Rebillat, Marc, LIMSI-CNRS, Acoustic and Audio, Batiment 508, 91403 Orsay, France
Rilling, Matthew, 7603 Callbram Ln., Austin, TX 78736
Riner, Joshua, U.L. Lafayette, Physics, P.O. Box 41278, Lafayette, LA 70504
Ropp, Tessa-Jonne, 231 Rollingbrook Way, Catonsville, MD 21228
Russell, Benjamin, 16305 Enclave Village Dr., Tampa, FL 33647
Serbel, Ryan, 29 Temple St. Unit 14, Hartford, CT 06103
Smith, Ronald, 9842 SW 88th St., Apt.. A205, Miami, FL 33176
Sorenson, Andy, 129 Heather Hill Dr., Downingtown, PA 19335
Spinu, Laura, Univ. of Delaware, Linguistics and Cognitive Science, 46 E Delaware Ave., Newark, DE 19716
Stimpson, Timothy, Univ. of Miami, 8 Country Club Dr., West Simsbury, CT 06092
Stubbs, Christopher, Univ. of California, Earth Science, 1006 Webb Hall, Santa Barbara, CA 93106
Sturm, Jonathon, 341 Buttonwood Dr., Kissimmee, FL 34743
Tejada, Laura, 1919 E. Sycamore St. #228, Anaheim, CA 92805
Teran, Sergio, Univ. of Miami, 5638 Merrick Dr. # 23-R, Coral Gables, FL 33146
Tu, Jung-Yueh, #203, 668 Whiting Way, Coqitlam, BC V3J3S4, Canada
Tung, Yao-Sheng, Columbia Univ., Biomedical Engineering, Columbia Medical Campus, VC12-232 622 W 168th St., New York, NY 10032
Turo, Diego, Univ. of Salford, Acoustics Research Center, G11, Newton Bldg., Univ. of Salford, Greater Manchester, Salford, M5 4WT, UK
Vega, Sebastian, 2360 Quailroost Dr., Weston, FL 33327
Vishnubhotla, Srikanth, Univ. of Maryland, Electrical and Computer Engineer, 3180, A V Williams Bldg., Univ. of Maryland, College Park, MD 20770
Walters, Brian, Univ. of Miami, 1515 San Remo Ave. Apt. D8, Coral Gables, FL 33146
Wang, James, 3465 Rue Hutchison, Apt.. 401, Montreal, QC H2X 2G3, Canada
Wang, Ningyu, 1248 Arthur E. Adams Dr., Columbus, OH 43202
Weiss, Michael, 1515 San Remo, Coral Gables, FL 33146
Wigley, Emily, 2334 S. Beck Ln., Apt. 103, Lafayette, IN 47909
William, Peter, UNL, Electrical Engineering, 209N WSEC, Lincoln, NE 68588-0511
Wroblewski, Marcin, CUNY Graduate Center, Speech and Hearing, 365 5th Ave #7402, New York, NY 10016
Wu, Xianghua, Simon Fraser Univ., Linguistics, 8888 University Dr., 9201 Robert C. Brown Hall Bldg., Burnaby, BC V5A 1S6, Canada
Yoo, Jin, 15444 SW 8th Way, Miami, FL 33194
Yu, Jiao, 1 Coliseum Dr., Room 1044, NCPA, University, MS 38677
Zhao, Yuan, Stanford Univ., Linguistics, Margaret Jacks Hall, Bldg. 460, Dept. of Linguistics, Stanford, CA 94305
Zimmerman, Emily, 1505A Legend Trail Dr., Lawrence, KS 66047

New Corresponding Electronic Associates

Ajayi, Donatus, Plot 4 Blk 6 Site C North, Satellite Town, Lagos, Nigeria
Albu, Felix, Univ. Politehnica of Buch, Telecommunications, 1-3 Bd. Iuliu Maniu, 060028 Bucharest, Romania
Kang, Heawon, Na-dong, 205, LG dormitory, Shing-Pyung-dong, Gumi, Korea
Mohamad, Maheza Irna, 52 jalan simbang, taman perling, 81200 Johor Bahru, Johor, Malaysia
Park, Mi Sun, #101-402, Suksa Apt., Suksa-dong, Chuncheon, Korea

Reinstated

B. F. Beaudry, D. R. Jones, D. L. Josephson—*Members*
M. N. Coleman, M. D. Hall, A. E. Queenan—*Associates*

Members Elected Fellows

R. Berger, II, D. A. Berry, G. A. Bissinger, S. E. Boyce, A. R. Bradlow, J. A. Fawcett, D. M. Freeman, B. R. Gerratt, F. H. Guenther, K. R. Kluender, Y. W. Lam, M. Long, C. Lorenzi, R. S. McGowan, L. Mongeau, P. W. Moore,

T. R. T. Nightingale, B. E. Pfingst, J. W. Posey, D. L. Rice, S. Rosen, A. Sarvazyan, C. H. Shadle, M. Siderius, M. A. Stone, A. K. Syrdal, A. M. Thode, D. H. Whalen, J. Wolfe, L. M. Zurk

Associates Elected Members

L. T. Fialkowski, B. A. Houghton, K. S. Hyeong, M. J. Isakson, J. X. Jian, M. Kuster, S. Lee, H. Paek, R. Sauro, R. Silva, A. M. Thode, S. D. Thunder, M. W. Urban, K. D. Wallace, A. J. Wright

Electronic Associate to Student

C. H. Wiese

Student to Associate

T. R. Agus, I. Akhoun, F. G. Bechwati, J. H. Bostron, S. L. Deruiter, J. Escolano, R. M. G. Ferrari, E. Gaudrain, C. E. K. Haller, T. A. Hay, T. D. Khokhlova, K. Kotsidou, R. Kruse, S. McLennan, K. R. Richardson, M. A. Roe, R. V. Roeder, W. Saintval, N. H. Schiller, T. Shinya, P. Simko, E. Vannoni, M. C. Vigeant, A. J. Wagner

Corresponding Electronic Associate to Associate

D-G. Paeng

Resigned

C. Dym, Fellow

J. R. Demko, S. Gilman, T. Maeda, K. Nakayama, G. Neuweiler, G. H. Ruske, P. W. J. Van Hengel, M. Yoneyama, *Members*

A-H. Afaneh, B. I. Bachrach, J. M. Chowning, E. Diakoumakou, H. A. Fujii, M. Jaeger, J. A. Johnson, M. P. Karnell, P. Kellett, L. L. Klein, J. Li, E. Murray-Brown, Y. Ng, E. Osmundsen. L. D. Pedersen, *Associates*

E. Larsen, L. Wilson, *Student*

Deceased

W. Endres, L. A. Herstein, III, R. K. Hillquist, R. M. Hoover, N. Shapiro, *Fellows*

A. H. Carter, J. V. Ellison, M. A. Schoenberg, C. Tepper, E. M. Weiler, *Members*

910	Fellows
2143	Members
3363	Associates
111	Corresponding Electronic Associates
1142	Students
7669	

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]

Paul D. Schomer, Standards Director
Schomer and Associates, 2117 Robert Drive, Champaign, IL 61821 [Tel.: (217) 359-6602; Fax: (217) 359-3303; e-mail: Schomer@SchomerAndAssociates.com]

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.

Standards Meetings Calendar

National Standards Meetings

Accredited Standards Committees S1, Acoustics; S2, Mechanical Vibration and Shock; S3, Bioacoustics; S3/SC 1, Animal Bioacoustics; and S12, Noise will meet in May 2009. The U.S. TAGs to ISO/TC 43, ISO/TC 43/SC 1, ISO/TC 108 and its subcommittees, and IEC/TC 29 will also meet during the Standards Plenary meeting. Specific meeting dates and times follow:

- **Monday, 18 May 2009**
  - **4:30 p.m.–5:30 p.m.**—Meeting of Accredited Standards Committee **S2** on Mechanical Vibration and Shock
- **Tuesday, 19 May 2009**
  - **7:15 a.m.–8:45 a.m.**—ASA Committee on Standards **(ASACOS)**. Meeting of the Committee that directs the Standards Program of the Acoustical Society.
  - **9:00 a.m.–10:00 a.m.**—Standards Committee **Plenary** Group Meeting (PL)—To discuss national and international items relevant to Accredited Standards Committees S1, S2, S3, S3/SC1 and S12. This meeting will include the meetings of the U.S. TAGs to ISO/TC 43, ISO/TC 43/SC 1, ISO/TC 108 and its subcommittees, and IEC/TC 29.
  - **10:30 a.m.–11:30 a.m.**—Meeting of Accredited Standards Committee **S3** on Bioacoustics
  - **11:45 a.m.–1:00 p.m.**—Meeting of Accredited Standards Committee **S3/SC1** on Animal Bioacoustics
  - **2:45 p.m.–3:45 p.m.**—Meeting of Accredited Standards Committee **S1** on Acoustics
  - **4:00 p.m.–5:00 p.m.**—Meeting of Accredited Standards Committee **S12** on Noise

International Standards Meetings

30 March–3 April 2009—London
- ISO/TC 108/SC 2, Measurement and evaluation of mechanical vibration and shock as applied to machines, vehicles and structures

15–19 June 2009—Copenhagen
- ISO/TC 108/SC 5, Condition monitoring and diagnostics of machines

9–13 November 2009—Tokyo
- ICE/TC 29, Electroacoustics

16–20 November 2009—Seoul
- ISO/TC 43, Acoustics
- ISO/43/SC 1, Noise
- ISO/TC 43/SC 2, Building acoustics

STANDARDS NEWS FROM THE UNITED STATES

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 parentheses are for information only.

ASA (ASC S1) (Acoustical Society of America)

Reaffirmations

BSR/ASA S1.11-2004 (R200x), Specification for Octave-Band and Fractional-Octave-Band Analog and Digital Filters (reaffirmation and redesignation of ANSI S1.11-2004).

Provides performance requirements for analog, sampled-data, and digital implementations of bandpass filters that comprise a filter set or spectrum analyzer for acoustical measurements. It supersedes ANSI S1.11-1986 (R1998), American National Standard Specification for Octave-Band and Fractional-Octave-Band Analog and Digital Filters, and is a counterpart to International Standard IEC 61260:1995, Electroacoustics-Octave-Band and Fractional-Octave-Band Filters. (February 16, 2009)

BSR/ASA S1.26-1995 (R200x), Method for Calculation of the Absorption of Sound by the Atmosphere (reaffirmation and redesignation of ANSI S1.26-1995 (R2004))

Provides means to calculate atmospheric absorption losses of sound from any source for a range of meteorological conditions. Attenuation coefficients for pure-tone sounds are calculated by means of equations or tables over ranges of frequency and humidity, pressure and temperature of atmosphere. For sounds analyzed by fractional-octave-band filters, alternative methods are given in annexes to calculate attenuation caused by atmospheric absorption from that specified for pure-tone sounds. (February 16, 2009)

ASA (ASC S3) (Acoustical Society of America)

Revisions

BSR/ASA S3.2-1989 (R200x), Method for Measuring the Intelligibility of Speech over Communication Systems (revision and redesignation of ANSI S3.2-1989 (R1999))

 0001-4966/2009/125(3)/1833/3/$25.00

Includes measurement of speech intelligibility over entire communication systems, evaluation of the contributions of elements of speech communication systems, and evaluation of factors that affect the intelligibility of speech. Speech intelligibility over a communication system is measured by comparing the monosyllabic words trained listeners receive and identify with the words trained talkers speak into a communication system that connects the talkers with the listeners. (December 29, 2008)

IEEE (Institute of Electrical and Electronics Engineers)

Supplements

BSR/IEEE 1431-2004/Cor 1-200x, Standard Specification Format Guide and Test Procedure for Coriolis Vibratory Gyros—Corrigendum 1: Figure 1—Gyro Axes and Misalignment Angles (supplement to ANSI/IEEE 1431-2004)

Corrects the labeling on a few figures. (January 6, 2009)

InfoComm (InfoComm International)

New Standards

BSR/INFOCOMM 1M-200x, Audio Coverage Uniformity in Enclosed Listener Areas (new standard)

Delivers consistent audio coverage in sound reinforcement systems to improve the quality of experience for the audience, and a new metrology for determining this coverage. (February 16, 2009)

NSF (NSF International)

Revisions

BSR/NSF 46-200x (i18), Evaluation of components and devices used in wastewater treatment systems (revision of ANSI/NSF 46-2007)

Issue 18—To remove section 11.4.1 relating to noise levels of chlorination devices. (December 7, 2008)

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:

CEA (Consumer Electronics Association)

BSR/CEA 2019-200x, Testing and Measurement Methods for Audio Amplifiers (new standard)

Call for Members

ASA (ASC S3) (Acoustical Society of America)

BSR/ASA S3.2-1989 (R200x), Method for Measuring the Intelligibility of Speech over Communication Systems (revision and redesignation of ANSI S3.2-1989 (R1999))

BSR/ASA S3.46-200x, Methods of Measurement of Real-Ear Performance of Hearing Aids (revision and redesignation of ANSI S3.46-1997 (R2007))

ASA (ASC S12) (Acoustical Society of America)

BSR/ASA S12.9-Part 3-200x, Quantities and Procedures for Description and Measurement of Environmental Sound, Part 3: Short-Term Measurements with an Observer Present (revision of ANSI/ASAS12.9-Part 3-1993 (R2008))

BSR/ASA S12.74-200x, Airborne Sound Measurements and Acceptance Criteria in Shipboard Spaces (new standard)

Final Actions on American National Standards

The standards actions listed below have been approved by the ANSI Board of Standards Review (BSR) or by an ANSI-Audited Designator, as applicable.

ABMA (ASC B3) (American Bearing Manufacturers Association)

Reaffirmations

ANSI/ABMA 13-1987 (R2008), Rolling Bearing Vibration and Noise (Methods of Measuring) (reaffirmation of ANSI/ABMA 13-1987 (R1999))

AHAM (Association of Home Appliance Manufacturers)

Reaffirmations

ANSI/AHAM AC-2-2006 (R2008), Method for Sound Testing of Portable Household Electric Room Air Cleaners (reaffirmation of ANSI/AHAM AC-2-2006): 12/15/2008

ASA (ASC S2) (Acoustical Society of America)

Reaffirmations

ANSI/ASA S2.29-2003 (R2008), Guide for the Measurement and Evaluation of Vibration of Machine Shafts on Shipboard Machinery (reaffirmation and redesignation of ANSI S2.29-2003)

Withdrawals

ANSI S2.7-1982 (R2004), Balancing Terminology (withdrawal of ANSI S2.7-1982 (R2004)): 12/4/2008

ANSI S2.17-1980, Techniques of Machinery Vibration Measurement (withdrawal of ANSI S2.17-1980 (R2004)): 12/4/2008

ASA (ASC S3) (Acoustical Society of America)

Reaffirmations

ANSI/ASA S3.41-1990 (R2008), Audible Emergency Evacuation Signal (reaffirmation of ANSI S3.41-1990 (R2001))

ASA (ASC S12) (Acoustical Society of America)

New Standards

ANSI/ASA S12.67-2008, Pre-Installation Airborne Sound Measurements and Acceptance Criteria of Shipboard Equipment (new standard)

ASTM (ASTM International)

Reaffirmations

ANSI/ASTM F2174-2002 (R2008), Practice for Verifying Acoustic Emission Sensor Response (reaffirmation of ANSI/ASTM F2174-2002)

Project Initiation Notification System (PINS)

ANSI Procedures require notification of ANSI by ANSI-accredited standards developers of the initiation and scope of activities expected to result in new or revised American National Standards. This information is a key element in planning and coordinating American National Standards.

The following is a list of proposed new American National Standards or revisions to existing American National Standards that have been received from ANSI-accredited standards developers that utilize the periodic maintenance option in connection with their standards. Directly and materially affected interests wishing to receive more information should contact the standards developer directly.

InfoComm (InfoComm International)

BSR/INFOCOMM 1M-200x, Audio Coverage Uniformity in Enclosed Listener Areas (new standard)
One of the fundamental goals of sound system performance for both speech reinforcement and program audio is the delivery of consistent coverage in the listening area. A well-executed audio system design is one that allows all listeners to hear the system at approximately the same sound pressure level throughout the desired frequency spectrum range, no matter where positioned in the designated listening area. This Standard provides a procedure to measure the spatial coverage, and the criteria for use in the design and commissioning of audio systems. At this time, there is not intention to submit to ISO for adoption but this option shall be reserved. Project Need: To allow delivery of consistent audio coverage in sound reinforcement systems and thereby improve the quality of the experience for the audience. Also, to create a new metrology for determining this coverage. Stakeholders: Entertainment venues, houses of Worship, Educational institutions, judicial and municipal chambers.

ASA (ASC S3) (Acoustical Society of America)

BSR/ASA S3.46-200x, Methods of Measurement of Real-Ear Performance of Hearing Aids (revision and redesignation of ANSI S3.46-1997 (R2007))
Covers the terminology, procedures and essential equipment characteristics for the measurement of the acoustic output and acoustic gain of hearing aids, coupled to human ears, in a variety of acoustic environments. Both direct measures of sound pressure level (SPL) and insertion measurements are covered. Project Need: This standard has not been revised since it was issued in 1997. There have been changes in hearing aids, test stimuli and fitting methods since then. Stakeholders: Educators, hearing aid company trainers, software developers and engineers, equipment manufacturers.

ASA (ASC S12) (Acoustical Society of America)

BSR/ASA S12.9-Part 3-200x, Quantities and Procedures for Description and Measurement of Environmental Sound—Part 3: Short-Term Measurements with an Observer Present (revision of ANSI/ASA S12.9-Part 3-1993 (R2008))
Specifies the procedures: (a) for measurement of environmental sounds from a specific source or sources at a specified location; and (b) to eliminate the contributions of extraneous background sounds from source-specific measurements effectively. Sound pressure levels are measured with an observer present to record the data described in this part. Project Need: To update the existing American National Standard so as to promote harmonization with other national and international noise standards. Stakeholders: Noise control engineers, architects, land use planners, public officials.

BSR/ASA S12.74-200x, Airborne Sound Measurements and Acceptance Criteria in Shipboard Spaces (new standard)
Specifies procedures and instrumentation for the sound-pressure measurement of airborne sound in shipboard spaces. Exceptions or additions to the requirements of this standard may be granted or added by the purchaser. Overall noise in ship compartments is a combination of noise generated from all equipment installed in or near that compartment as well as other possible sources. Project Need: To describe measurements that, when taken according to standardized procedures and compared with acceptance criteria, will provide evidence of the overall acoustic adequacy of the ship's design and construction. Stakeholders: Naval architects, designers, engineers, shipbuilders, military and non-military purchasers of ships.

IEEE (Institute of Electrical and Electronics Engineers)

BSR/IEEE C37.082-200x, Recommended Practice for the Measurement of Sound Pressure Levels of Outdoor Power Circuit Breakers (revision of ANSI/IEEE C37.082-1982 (R2006))
Provides recommended practice and methods for the measurement of sound-pressure levels produced by outdoor power circuit breakers in a free-field environment. These methods may also be used indoors or in restricted fields, provided that precautions are observed in the measurement and the interpretation of the results. Project Need: To update an existing standard in order to include modern practices. Stakeholders: Utilities, industry, field engineers, manufacturers.

STANDARDS NEWS FROM ABROAD

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

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 3822-1/Amd1:2008, Acoustics—Laboratory tests on noise emission from appliances and equipment used in water supply installations—Part 1: Method of measurement—Amendment 1: Measurement uncertainty

FLUID POWER SYSTEMS (TC 131)

ISO 15086-3:2008, Hydraulic fluid power—Determination of the fluid-borne noise characteristics of components and systems—Part 3: Measurement of hydraulic impedance

MECHANICAL VIBRATION AND SHOCK (TC 108)

ISO 20283-2:2008, Mechanical vibration—Measurement of vibration on ships—Part 2: Measurement of structural vibration

IEC Standards

AUDIO, VIDEO AND MULTIMEDIA SYSTEMS AND EQUIPMENT (TC 100)

IEC 61606-3 Ed. 1.0 en:2008, Audio and audiovisual equipment—Digital audio parts—Basic measurement methods of audio characteristics—Part 3: Professional use

IEC 62087-BD Ed. 2.0 en:2008, Methods of measurement for the power consumption of audio, video and related equipment

ISO and IEC Draft International Standards

ISO

ACOUSTICS (TC 43)

ISO 3822-3/DAmd1, Acoustics—Laboratory tests on noise emission from appliances and equipment used in water supply installations—Part 3: Mounting and operating conditions for in-line valves and appliances—Draft Amendment 1 (2/26/2009)

MEASUREMENT OF FLUID FLOW IN CLOSED CONDUITS (TC 30)

ISO/DIS 17089-1, Measurement of fluid flow in closed conduits—Ultrasonic meters for gas—Part 1: Meters for custody transfer and allocation measurement (1/30/2009)

MECHANICAL VIBRATION AND SHOCK (TC 108)

ISO 2631-4/DAmd1, Statistical analysis method (3/2/2009)

IEC Draft Standards

29/671/FDIS, IEC 61094-2 Ed.2: Electroacoustics—Measurement microphones—Part 2: Primary method for the pressure calibration of laboratory standard microphones by the reciprocity technique (01/30/2009)

REVIEWS OF ACOUSTICAL PATENTS

Sean A. Fulop
Dept. of Linguistics, PB92
California State University Fresno
5245 N. Backer Ave., Fresno, California 93740

Lloyd Rice
11222 Flatiron Drive, Lafayette, Colorado 80026

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

Reviewers for this issue:

GEORGE L. AUGSPURGER, *Perception, Incorporated, Box 39536, Los Angeles, California 90039*
ANGELO CAMPANELLA, *3201 Ridgewood Drive, Hilliard, Ohio 43026-2453*
GEOFFREY EDELMANN, *Naval Research Laboratory, Code 7145, 4555 Overlook Avenue SW, Washington, DC 20375*
JEROME A. HELFFRICH, *Southwest Research Institute, San Antonio, Texas 78228*
DAVID PREVES, *Starkey Laboratories, 6600 Washington Avenue South, Eden Prairie, Minnesota 55344*
NEIL A. SHAW, *Menlo Scientific Acoustics, Inc., Post Office Box 1610, Topanga, California 90290*
ERIC E. UNGAR, *Acentech, Incorporated, 33 Moulton Street, Cambridge, Massachusetts 02138*
ROBERT C. WAAG, *Department of Electrical and Computer Engineering, University of Rochester, Rochester, New York 14627*

7,434,409

43.35.Ud PULSE TUBE COOLER HAVING 1/4 WAVELENGTH RESONATOR TUBE INSTEAD OF RESERVOIR

David R. Gedeon, assignor to Sunpower, Incorporated
14 October 2008 (Class 62/6); filed 23 August 2005

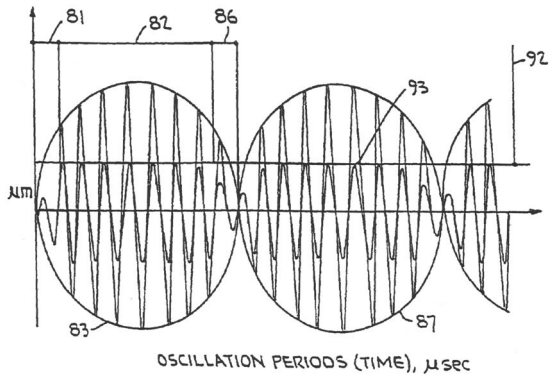
The need arises for all sorts of design tricks in acoustic refrigerator design. This brief patent describes one of those: the use of a quarter-wave termination in a Gifford–McMahon cryocooler design to eliminate the need for a large constant-pressure reservoir in the system. The reduced volume and heat capacity of this design are especially appealing to end users of these systems.—JAH

7,344,609

43.35.Zc ULTRASONIC IMPACT METHODS FOR TREATMENT OF WELDED STRUCTURES

Efim S. Statnikov, assignor to U.I.T., L.L.C.
18 March 2008 (Class 148/558); filed 1 December 2004

An ultrasonic impact device (UID) of superior performance for toughening the working surface of a metal object such as a vehicle brake shoe is claimed. Driving current 93 is sinusoidal at the ultrasonic resonance of the tool hammer bar. This drive current wave form is pulsed as envelope 83 and 87 such that rest periods, 86, etc., occur between drive periods to allow tool tip contact with the work piece to improve. Material deforming impacts occur at an ultrasonic frequency rate during the greater amplitude parts 83 and 87. During the moments of diminished drive current 86, the natural resonance frequency is sensed via its ultrasound ringing current zero crossing intervals. The drive current frequency is then readjusted, determining a



new drive frequency for next envelope pulse 87. This action (drive–ring–readjust–drive anew) optimizes bar resonance under load, thus providing the maximum possible ultrasound hammer amplitude for impact-hardening the surface of the work piece.—AJC

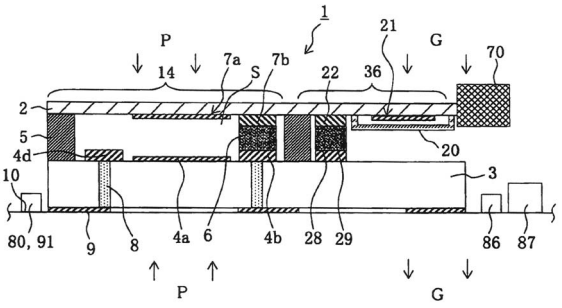
7,353,710

43.35.Zc PRESSURE SENSOR DEVICE WITH SURFACE ACOUSTIC WAVE ELEMENTS

Akira Oikawa *et al.*, assignors to Kyocera Corporation
8 April 2008 (Class 73/703); filed in Japan 27 November 2003

A transmitting tire pressure sensor 1 that is economical with electric power is claimed. Tire air pressure P depresses diaphragm 14 over closed space S. Surface acoustic wave (SAW) device 7a senses the diaphragm deflection while SAW device 4a mounted on backing 3 provides a reference SAW signal to compensate for environmental temperature changes. A radio transmitter and antenna (not shown) convey the signal to a receiver in the vehicle. Operating electric power supplied by a battery is switched on via a

 0001-4966/2009/125(3)/1836/10/$25.00



signal from transducer 21 mounted on bar 36 that is bent significantly by bob 70, but only when the vehicle is in motion and tire wheel is turning.—AJC

7,436,736

43.38.Bs HYDROPHONE ARRAY MODULE

John K. Schneider and Jack C. Kitchens, assignors to Ultra-Scan Corporation
14 October 2008 (Class 367/170); filed 13 August 2007

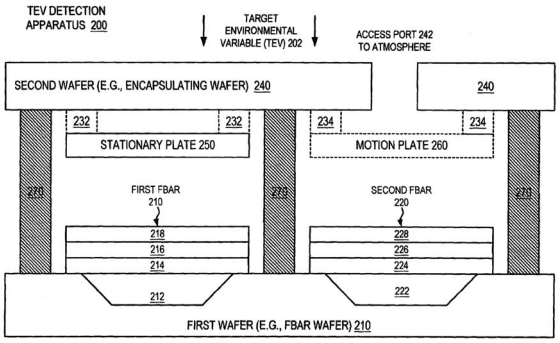
In order to facilitate the mass production of transducers, an array of small electrodes are used to create an electrostatic field (instead of labor intensive potting of a single, well-placed electrode).—GFE

7,427,819

43.38.Fx FILM-BULK ACOUSTIC WAVE RESONATOR WITH MOTION PLATE AND METHOD

Storrs T. Hoen *et al.*, assignors to Avago Wireless IP Pte Limited
23 September 2008 (Class 310/320); filed 4 March 2005

The authors disclose a design for a pressure or displacement transducer built around a film bulk acoustic resonator (FBAR). They show how the use of a pair of FBARs with just one of them exposed to the environment can be used to increase the stability (temperature and strain), and how a pentagonal cross-section for the FBAR can improve the rejection of lateral vibration modes in this component. The patent also discloses an accelerometer design based on the same resonator concept. It is clear that the motivation for this



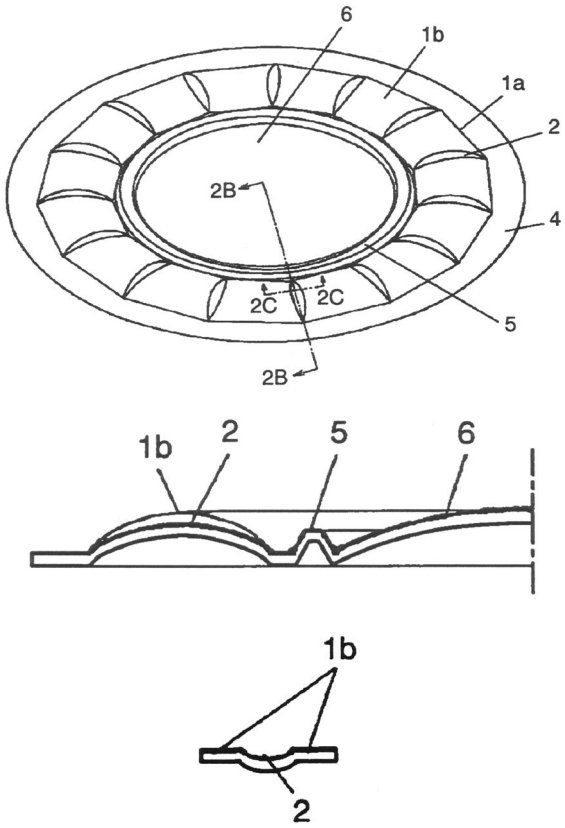
device is to get a sensor that couples directly to radio frequencies, but the concept seems to be very inefficient at coupling energy from the motion to the acoustic motion, resulting in a low coupling constant.—JAH

7,428,946

43.38.Ja SUSPENSION AND ELECTRO-ACOUSTIC TRANSDUCER USING THE SUSPENSION

Kazuki Honda *et al.*, assignors to Matsushita Electric Industrial Company, Limited
30 September 2008 (Class 181/171); filed in Japan 25 October 2002

Corrugated or half-roll loudspeaker cone suspensions are subject to hoop stresses that restrict the range of linear excursion and a number of alternative geometries have been patented. What is shown here is somewhat similar to an early Western Electric design in that individual cylindrical



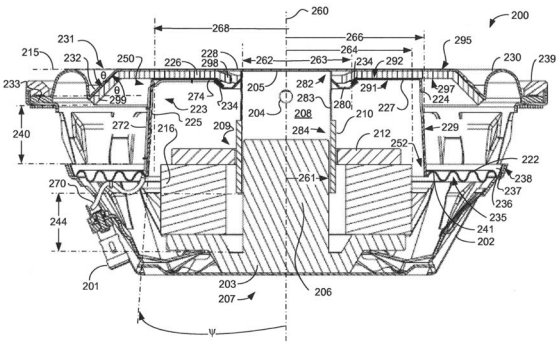
sections 1b are separated by deformable bridges 2. In this case the assembly is monolithic and the bridges are radial dimples molded into the generally convex surface. Sections 2B and 2C clarify the geometry.—GLA

7,433,485

43.38.Ja SHALLOW SPEAKER

Brad Michael Diedrich and Chad A. Kautz, assignors to Mitek Corporation, Incorporated
7 October 2008 (Class 381/404); filed 13 February 2008

The major feature of this shallow speaker is a large-diameter centering spider 222 that sits outside the magnetic assembly instead of on top. For whatever reason, the invention as illustrated is not nearly as shallow as it might be—there is a lot of wasted space between the planar diaphragm and the top of the voice coil.—GLA

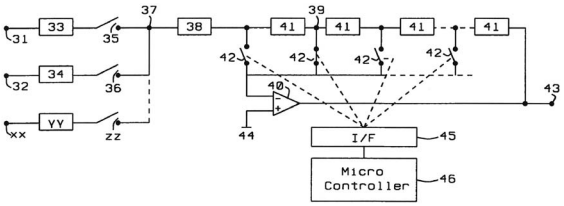


7,433,477

43.38.Lc AUDIO SUPERPOSITIONING

Edgar Sexauer and Klaus Graef, assignors to Dialog Semiconductor GmbH
7 October 2008 (Class 381/119); filed in the European Patent Office 16 April 2002

Instead of a summing amplifier followed by a gain control stage, the gain of summing amplifier 40 is adjusted by changing the value of its feedback resistor. Is it possible that no prior art exists?—GLA

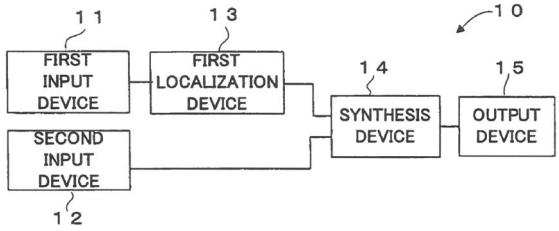


7,433,479

43.38.Lc HEADPHONE APPARATUS

Koichi Sato, assignor to Pioneer Corporation
7 October 2008 (Class 381/309); filed in Japan 20 December 2002

Some game boxes allow two players to compete via the Internet and simultaneously carry on a conversation using headsets with microphones. In such an arrangement, the sound effects of the game are normally reproduced by loudspeakers. To avoid annoying others, however, one or both players may prefer to hear everything through the headset; but if the two signals are simply mixed, it may be difficult to distinguish between them. This patent teaches that if stereo headsets are used, then at least one signal can be synthetically localized to make its identification easier.—GLA

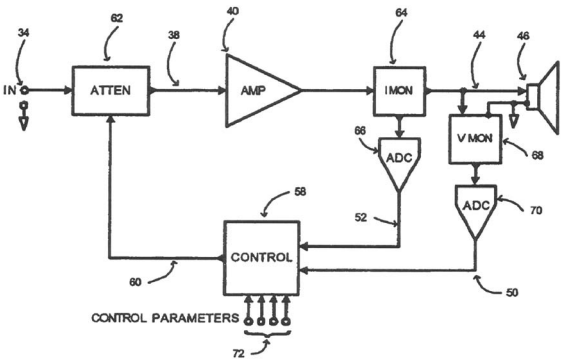


7,436,967

43.38.Lc APPARATUS AND METHOD OF LIMITING POWER APPLIED TO A LOUDSPEAKER

Brian Neunaber, assignor to QSC Audio Products, Incorporated
14 October 2008 (Class 381/55); filed 23 March 2005

This is a continuation of U.S. patent no. 6,940,981 which was filed in 2003. Several variants are described but they are all concerned with protecting a loudspeaker against thermal overload, and they all involve simultaneous sampling of voltage and current. The two values are processed in some fashion to estimate power or voice coil temperature, and the input signal is attenuated when a threshold is exceeded. As described, the various embodiments use digital processing but this is not required by the patent claims. Some hidden assumptions and simplifications are made, but these may well be of little practical importance. However, the patent also sets forth a number of statements that are simply false, i.e., volt-amperes does not equate to watts, a loudspeaker's impedance curve does not retain the same shape when its voice coil resistance increases, and there is no such thing as "instantaneous power."—GLA

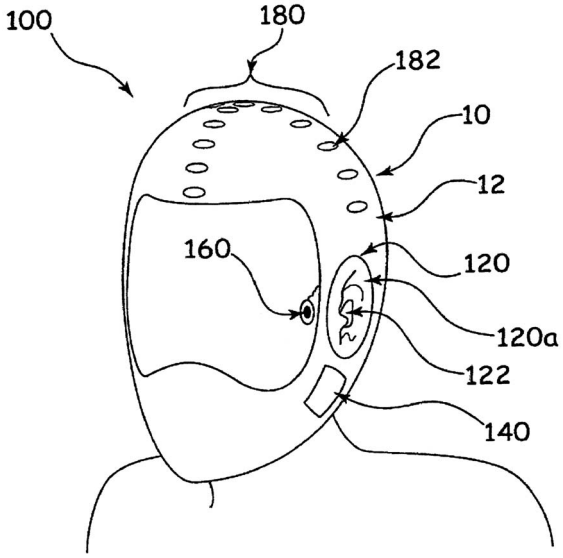


7,430,300

43.38.Md SOUND PRODUCTION SYSTEMS AND METHODS FOR PROVIDING SOUND INSIDE A HEADGEAR UNIT

Frederick Vosburgh and Walter C. Hernandez, assignors to Digisenz LLC
30 September 2008 (Class 381/376); filed 17 November 2003

Here is a bit of free advice to the inventor who is rushing to patent a red hot idea: before writing up endless details of possible variants and filling pages of text with questionable theory, make a simple model to see if it works. Turning to the patent at hand, anyone who has worn a sports or safety helmet knows that external sounds are substantially attenuated. To remedy this situation, one might mount a pair of microphones on the outside of the helmet and transmit their signals to a corresponding pair of earphones inside the helmet. Moreover, to be really scientific about this, each microphone 122 might be seated in a molded replica of a human outer ear 120. (Why not just drill a couple of holes and dispense with the electronics?) But that's not all. The molded ears can be augmented or replaced by an array of microphones 180 whose signals are suitably analyzed and processed to yield a simulation of the free-air sound field. Conspicuously missing from the patent is any discussion of physical contact with sports helmets or wind noise in motorcycle helmets.—GLA



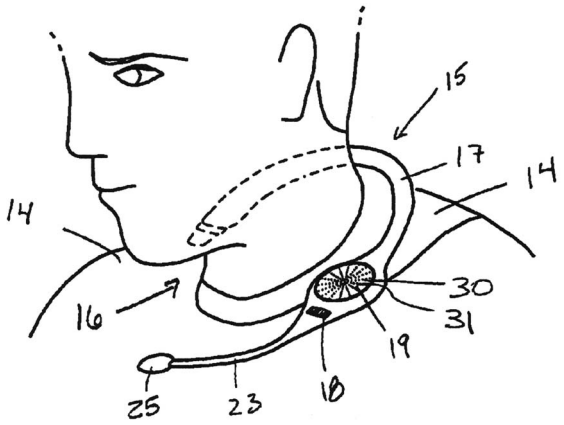
7,428,429

43.38.Si HANDS-FREE PERSONAL COMMUNICATION DEVICE

Christopher C. Gantz *et al.*, assignors to SYNC1 LLC
23 September 2008 (Class 455/575.1); filed 5 January 2005

This is a continuation of U.S. patent no. 6,934,567, originally filed in 2002. It discloses an interesting headset design that requires no headband. Instead, the assembly drapes across the user's shoulders, positioning loudspeaker 30 and microphone 25 at appropriate locations. Yoke 17 is deformable, allowing the assembly to rest on either the left or right shoulder. A

connecting cable emerges from the far end of the yoke (dotted line). Also stored in the far end is a removable earbud that can be taken out and used instead of the loudspeaker.—GLA

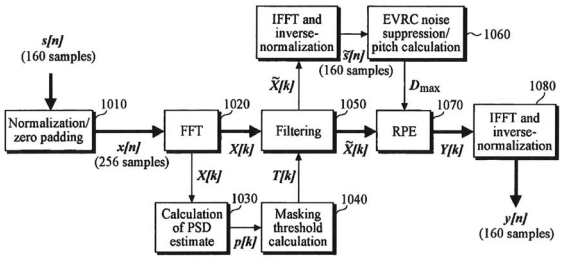


7,430,506

43.38.Si PREPROCESSING OF DIGITAL AUDIO DATA FOR IMPROVING PERCEPTUAL SOUND QUALITY ON A MOBILE PHONE

Young Han Nam *et al.*, assignors to RealNetworks Asia Pacific Company, Limited
30 September 2008 (Class 704/207); filed in Republic of Korea 9 January 2003

To make efficient use of available bandwidth, cellular telephone networks encode audio signals dynamically, employing algorithms based on human speech. During normal pauses between words, for example, the encoder classifies the signal as noise rather than information, and the encoding rate drops to its lowest value (1/8). Unfortunately, as explained in this patent, encoding optimized for speech can mangle music. Percussive signals are classified as noise, and annoying dropouts occur. Worse yet, the longer music is played, the more dropouts are generated. A sophisticated method of preprocessing is disclosed which fools the encoder into identifying music as speech. "By this method, the number of pauses in a music signal is reduced, and the perceptual sound quality of the music is improved."—GLA

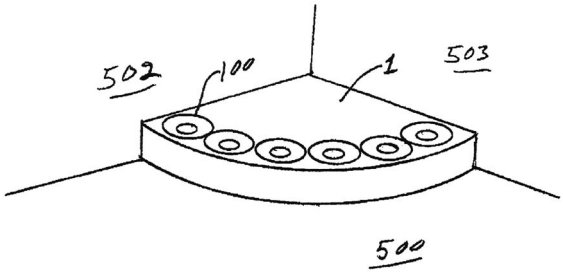


7,409,071

43.38.Tj LARGE-DIAMETER ARCUATE SPEAKER

Nick Bromer, Marietta, Pennsylvania
5 August 2008 (Class 381/335); filed 15 July 2003

A means of converting the "wind" portion in an arcuate array is disclosed. "The pulsating motion also produces changes in the air pressure, as the particle velocity can be divided into two parts that differ in their relationship to the air pressure. I (the inventor) call these parts sound, and wind." Huh? An interesting discussion follows that says, in part, the sound part follows a $1/r$ relation but the wind part follows a $1/(r\times r)\times k$ formula, k being the wavenumber. Moreover, since one transducer is common practice, the number of loudspeakers needs to be "equal to a perfect square $n\times n$ of a number n ($n\times n=4,9,16\ldots$)." Despite the contortions in the prosaic description and the number of figures supplied being less than that referenced in the text, a practical device may result.—NAS

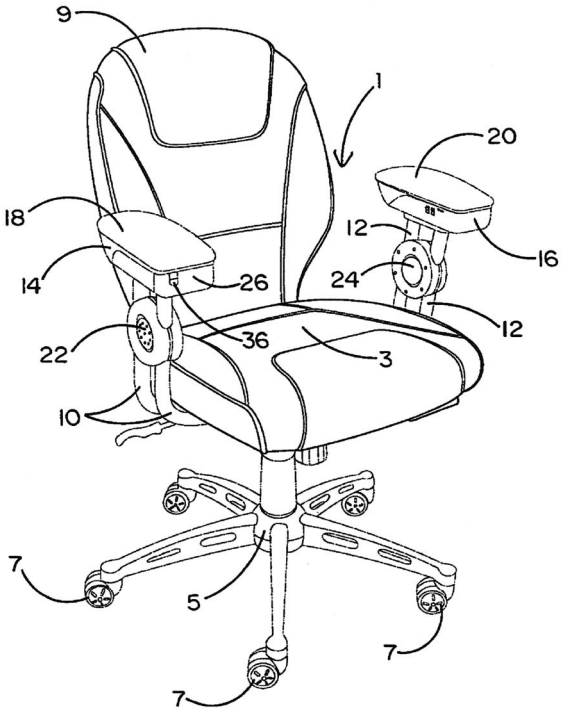


7,431,392

43.38.Vk CHAIR HAVING BUILT-IN AUDIO SPEAKERS AND A SLIDE-OUT COMPARTMENT FOR A PORTABLE DIGITAL STORAGE AND PLAYBACK DEVICE

George Tamara, assignor to True Seating Concepts, LLC
7 October 2008 (Class 297/217.4); filed 26 September 2006

This is not the first stereo chair to be patented. The two novel features appear to be the locations of the loudspeakers 22, 24 and a slide-out compartment 26 for an MP3 player. The patent places a great deal of emphasis on the slide-out compartment—the term appears in the title, abstract, and claims, although one independent claim refers merely to open and closed positions. Power for the electronic components can be supplied from a built-in battery pack or an AC power cord that emerges from opening 36.—GLA



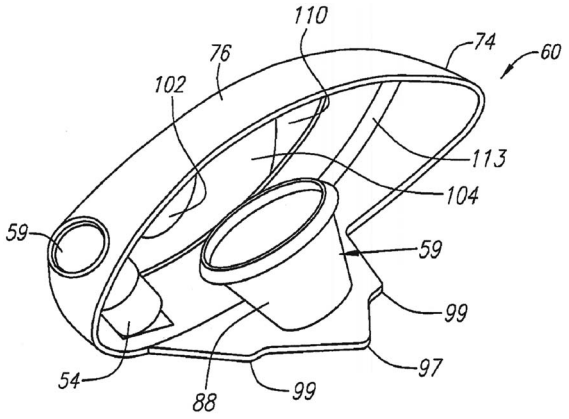
7,402,112

43.40.Ey MULTIPLE MATERIAL GOLF CLUB HEAD

J. Andrew Galloway, assignor to Callaway Golf Company
22 July 2008 (Class 473/224); filed 15 October 2007

A means is disclosed wherein the sound characteristics of a golf club

head 60 is modified by component 59 so that the sound of the golf club head is altered when it impacts a golf ball. Data in the patent indicates that the changes in sound level can vary from −1.5 dBA to +1.2 dBA, and the center frequency pitch can be lowered from 100 to 800 Hz in comparison to that of an ordinary club.—NAS

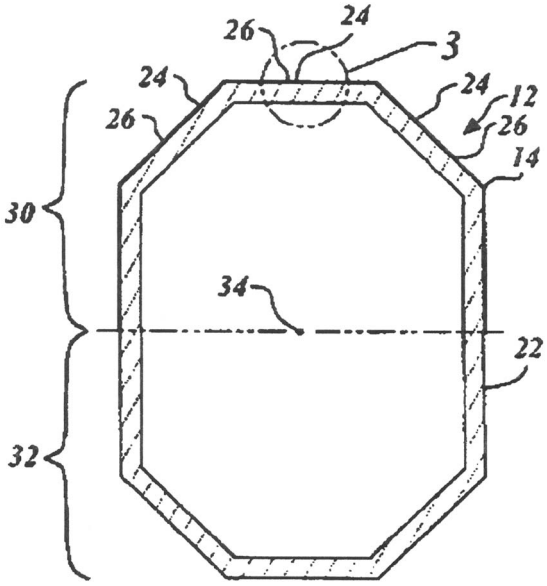


7,404,775

43.40.Tm HANDLE FOR A LACROSSE STICK

David Morrow and Andrew Maliszewski, assignors to Warrior Lacrosse, Incorporated
29 July 2008 (Class 473/513); filed 3 March 2005

Coating 26 is "coupled" to outer surface 22 of tubular member 14, a part of lacrosse stick 12 (or any device that requires a handle where the orientation of same needs to be determined by tactile stimuli). The coating, depending on its composition, can decrease vibration while also providing the tactile stimuli, as it need not be applied to the entire perimeter 22. The thickness of the coating can vary from the head end to the bottom end (which can provide additional tactile information), and it can be applied with varying smoothness or roughness, to suit the desires of a player.—NAS

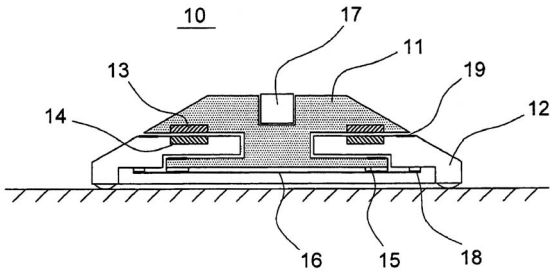


7,412,066

43.40.Tm MICROPHONE

Peter Muren and Trygve Frederik Marton, assignors to Tandberg Telecom AS
12 August 2008 (Class 381/177); filed in Norway 25 August 2003

Audio teleconference assemblies are often placed on tables. Impacts to the table from human interaction can cause shocks to the microphone assembly resulting in unpleasant noises at the teleconference "far end." Various and sundry elastic vibration isolation systems have been devised, but this system uses "magnetic levitation." Microphone 17 is mounted in substantially heavy top part 11 which is levitated above base 12 by magnet sets 13 and 14. The magnet sets are placed so that like poles are facing one another which generates a repelling magnetic force. Neat idea.—NAS

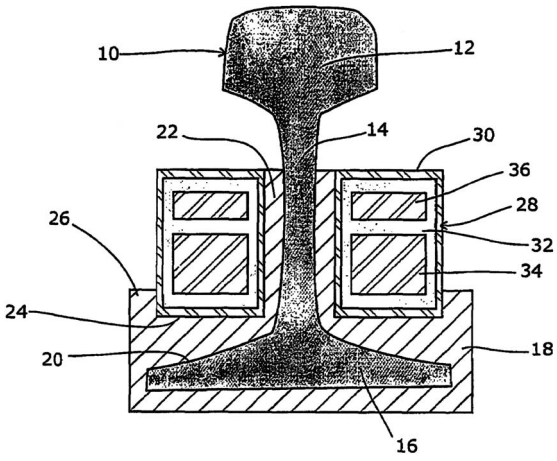


7,427,035

43.40.Tm RAIL DAMPER

David Farrington and Christopher John Cedric Jones, assignors to Corus UK Limited
23 September 2008 (Class 238/382); filed in United Kingdom 5 March 2003

A nylon clip 18 has an aperture 20 within which the foot of the rail fits. The damper 28 consists of an elongated hollow rigid steel section 30 that is filled with a rubber or rubber-like material, embedded in which are masses 34 and 36. The masses are sized so that their resonant frequencies are in the range of the rail's vibration frequencies.—EEU



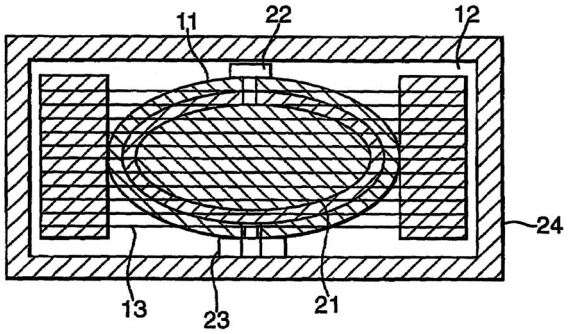
7,345,953

43.40.Yq FLEXTENSIONAL VIBRATION SENSOR

Roger Ian Crickmore *et al.*, assignors to Qinetiq Limited
18 March 2008 (Class 367/163); filed in United Kingdom 8 November 2002

A sensitive fiber optics seismic acceleration sensor is claimed where

mass 21 is attached by bolt 22 to flex tensional ring 11, in turn attached to supporting enclosure 24 by attachments 23. Acceleration along the 23 to 22 direction will cause mass 21 to pull ring 11 inward, causing it to also elongate laterally along its major axis, thereby stretching fiber optics filament 13 wound around it. Light passing through the length of the fiber filament will cause a delay that can be measured as proportional to acceleration. This sensor is insensitive to acceleration in either of the two lateral directions.—AJC



7,429,053

43.40.Yq SUBFRAME STRUCTURE AND VIBRATION DAMPER FOR THE SAME

Akira Katagiri *et al.*, assignors to Tokai Rubber Industries, Limited
30 September 2008 (Class 280/124.109); filed in Japan 9 September 2004

An automotive vehicle's wheel and suspension assembly is resiliently supported from a subframe, which is resiliently connected to the vehicle. For certain control purposes, such as the activation of antilock braking systems, the state of motion of the vehicle needs to be determined. This patent describes dynamic load sensors for this purpose that are incorporated in resilient connectors between the subframe and vehicle frame. These sensors employ Hall effect elements to measure the loads along three axes.—EEU

7,430,439

43.40.Yq USER HAND DETECTION FOR WIRELESS DEVICES

Jason Griffin and Steven Fyke, assignors to Research In Motion Limited
30 September 2008 (Class 455/552.1); filed 31 January 2005

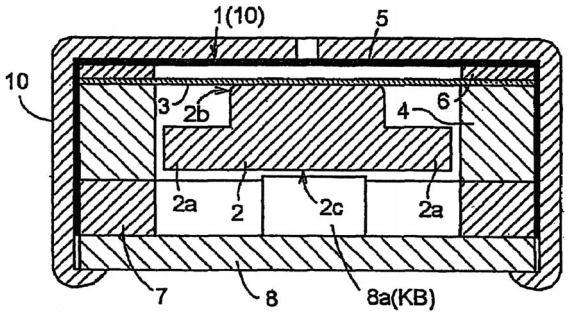
Whether a wireless device, such as a cell phone, is being held in a user's hand (versus being stored in a holster, for example) is determined by activating a vibrator in the device for a brief period, measuring the resulting acceleration, and comparing the measured acceleration to a stored signal.—EEU

7,430,915

43.40.Yq VIBRATION SENSOR

Mamoru Yasuda *et al.*, assignors to Hosiden Corporation
7 October 2008 (Class 73/652); filed in Japan 2 September 2003

This single-axis sensor is intended to be sensitive, rugged, and relatively easy to assemble. It consists of an electrode 1 that is rigidly fixed inside a housing 10 and an electrode membrane 3 that carries a mass 2. Vibrations are sensed in terms of changes in the capacitance between the two electrodes resulting from motion along an axis perpendicular to the membrane. A ring 9 protects the membrane from excessive axial excursions; for example, if the sensor is dropped. The extensions 2a at the base of the mass 2 and the ring surrounding the mass prevent excessive lateral and torsional deflections due to lateral impacts. Lips at the bottom of the plastic housing 10 enable convenient "snapping together" assembly of the device.—EEU



7,437,274

43.40.Yq METHOD AND APPARATUS FOR OBJECTIVE MEASUREMENT OF NOISE

Francois Charette *et al.*, assignors to Ford Motor Company
14 October 2008 (Class 702/184); filed 12 August 2002

For quality control purposes an automotive vehicle is placed on a platform, which subjects the vehicle to vibrations representative of those resulting from the vehicle's driving over various surfaces. These vibrations may give rise to noise, such as squeaks and rattles. Rather than relying on subjective evaluation by a listener, the arrangement described in this patent uses sound data acquired via a microphone and compares the data, analyzed in various ways, with predetermined criteria.—EEU

7,430,912

43.50.Fe RANDOM INCIDENT ABSORBER APPROXIMATION

Donald C. Albin, Jr., assignor to International Automotive Components Group North America, Incorporated
7 October 2008 (Class 73/599); filed 28 December 2005

Described is an obvious and commonly used method to measure sound absorption as a function of incident angle. It is specifically geared toward automotive noise reduction.—GFE

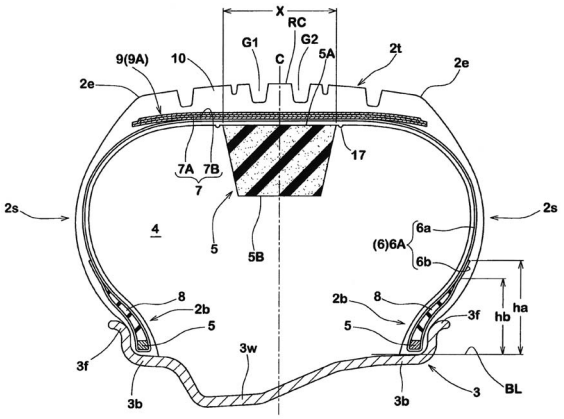
7,347,239

43.50.Gf TIRE NOISE REDUCING SYSTEM

Naoki Yukawa and Tadao Matsumoto, assignors to Sumitomo Rubber Industries, Limited
25 March 2008 (Class 152/209.18); filed in Japan 27 January 2004

A tire noise damper 5B is claimed, where plastic foam ring 5B is fitted into tire cavity 4 between retaining ribs 17 cast into the interior tire surface. Ring 5-5B volume can be between 0.4% to 20% of the tire cavity. Tires tested have a cavity volume resonance frequency between 125 and 200 Hz.

In one case, a damper reduced a 220 Hz resonance amplitude by about 6 dB, and reduced the broadband noise by 2 dB—AJC

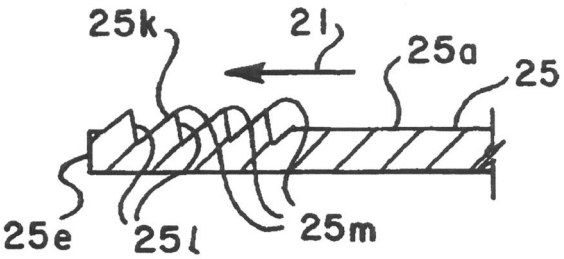


7,351,041

43.50.Gf FAN WITH REDUCED NOISE GENERATION

Robert B. Uselton *et al.*, assignors to Lennox Industries Incorporated
1 April 2008 (Class 416/236 R); filed 10 February 2005

A method of reducing high frequency (whistling) noise emitted by a moving fan blade 25 by means of skived roughness 25k-l-m on the trailing edge pressure side 25a of fan blade 25 is claimed. By turning boundary layer 21 turbulent by this means, it is hoped to prevent the coherent shedding of vortices at the fan blade trailing edge 25e. Theoretical expressions for sizing the roughness scale are provided, but no supporting test data is included.—AJC

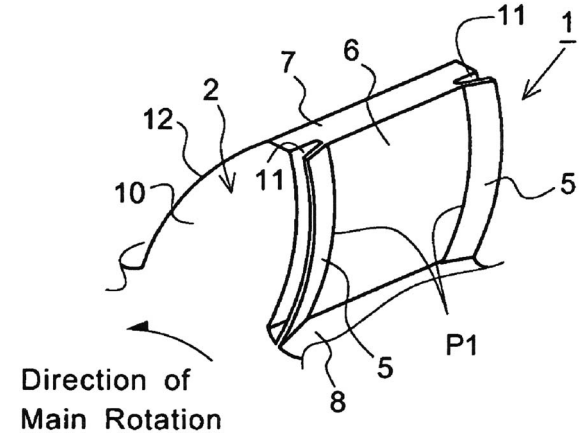


7,406,892

43.50.Gf PLASTIC MOLDED GEAR, AND INTERMITTENT ROTATION TRANSMISSION DEVICE AND GEAR TRAIN USING SAME

Yoichi Takeuchi *et al.*, assignors to Enplas Corporation
5 August 2008 (Class 74/440); filed in Japan 3 September 2003

Protruding portion 5 is formed at the ends of tooth flanks 6 of each tooth 2. The protruding portion 5 is elastically deformed so as to absorb backlash. A good tutorial of prior art tooth design for backlash control is included in the text.—NAS



7,434,657

43.50.Gf ACOUSTIC FOAM SOUND REDUCER FOR VACUUM POWER UNIT

Darrell V. Nieschwitz and Andrew L. Gabric, assignors to H-P Products, Incorporated
14 October 2008 (Class 181/231); filed 27 October 2005

A sound absorber design is detailed, for the purpose of quietening a vacuum motor.—GFE

7,434,659

43.50.Gf ACOUSTIC SEPTUM CAP HONEYCOMB

Earl Ayle, assignor to Hexcel Corporation
14 October 2008 (Class 181/292); filed 4 April 2005

A thin lightweight sound absorber using a honey comb shape is detailed for the purpose of reducing jet engine noise.—GFE

7,427,823

43.58.Hp PIEZOELECTRIC VIBRATION ELEMENT, PIEZOELECTRIC VIBRATOR, PIEZOELECTRIC OSCILLATOR, FREQUENCY STABILIZATION METHOD, AND METHOD OF MANUFACTURING THE PIEZOELECTRIC VIBRATOR

Tsuyoshi Ohshima *et al.*, assignors to Epson Toyocom Corporation
23 September 2008 (Class 310/344); filed in Japan 3 February 2005

This patent teaches a method for stabilizing piezoelectric resonators using monolayer adsorption of chemicals on the resonator material and electrodes. The principle presented appears to be extremely simple: One puts a drop of polydimethysiloxane (PDMS) together with the resonator in a sealed cavity and heats it to a prescribed temperature for a prescribed time. The authors disclose just how long it takes for their resonators to stabilize, but, in general, it can be accomplished in a few hours. The frequency drops as a

monolayer of PDMS is transferred, and then it stabilizes and the resonator is ready for use. Unfortunately, little information is given about the types of materials used for the resonators and what the ultimate stability of the devices is—does the PDMS keep migrating?—JAH

7,433,266

43.60.Gk ACOUSTIC SOURCE LOCALIZATION SYSTEM AND APPLICATIONS OF THE SAME

Ákos Lédeczi *et al.*, assignors to Vanderbilt University
7 October 2008 (Class 367/129); filed 15 September 2005

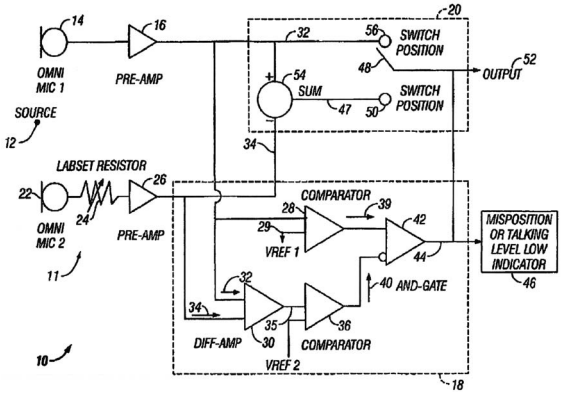
A method is presented which is purported to overcome the problems with localization of sniper/gun shot/explosions in an urban environment. Yet, it ultimately falls back on common time-of-arrival processing.—GFE

7,346,176

43.60.Qv AUTO-ADJUST NOISE CANCELING MICROPHONE WITH POSITION SENSOR

Robert J Bernardi *et al.*, assignors to Plantronics, Incorporated
18 March 2008 (Class 381/92); filed 11 May 2001

Noise canceling microphone system 10 is claimed that determines whether the microphone is properly positioned. When the talker source 12 is near microphone 14 and further from microphone 22, comparator 28-36 determines that acceptable acoustic activity is occurring and causes switch 48 to connect differential sum output 47-50 to output 52. If talker source 12 is remote from the microphones, microphones 14 and 22 receive similar sound levels, causing the difference 47 to be virtually zero. Comparator 28-36, on sensing this, causes switch 48 to connect microphone 14 output 32-56 directly to output 52 and to turn on display 46 to warn the user to position the microphone set closer to source 12.—AJC



7,434,834

43.60.Qv OCCUPANT PROTECTION SYSTEM FOR VEHICLE

Takashi Mihara and Yukiyasu Ueno, assignors to Denso Corporation
14 October 2008 (Class 280/735); filed in Japan 10 March 2004

To prevent the erroneous deployment of airbags during a presumed rear-end collision, accelerometer sensors are augmented with sonar sensors on the bumper (currently used for parking collision prevention).—GFE

7,428,312

43.66.Ts METHOD FOR ADAPTING A HEARING DEVICE TO A MOMENTARY ACOUSTIC SITUATION AND A HEARING DEVICE SYSTEM

Hilmar Meier *et al.*, assignors to Phonak AG
23 September 2008 (Class 381/312); filed 18 November 2003

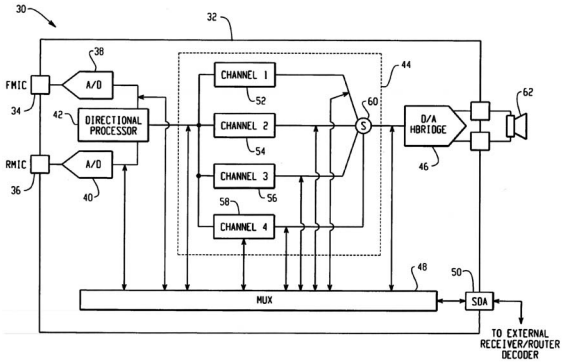
The momentary acoustical environment is identified by a hearing device, and its signal processing parameters, including those other than volume level, are automatically updated accordingly to preset rules. The wearer may also adjust these parameters manually via wireless control from a remote device or with an input unit, e.g., buttons, integrated into the hearing device.—DAP

7,430,299

43.66.Ts SYSTEM AND METHOD FOR TRANSMITTING AUDIO VIA A SERIAL DATA PORT IN A HEARING INSTRUMENT

Stephen W. Armstrong and Brian D. Csermak, assignors to Sound Design Technologies, Limited
30 September 2008 (Class 381/312); filed 12 April 2004

A bi-directional serial data port facilitates communication between a hearing device and external device. Processed audio signals picked up by an ear canal microphone are transmitted to the external device to allow the hearing aid dispensing professional to monitor the audio presented to the hearing aid wearer as signal processing parameters are changed.—DAP



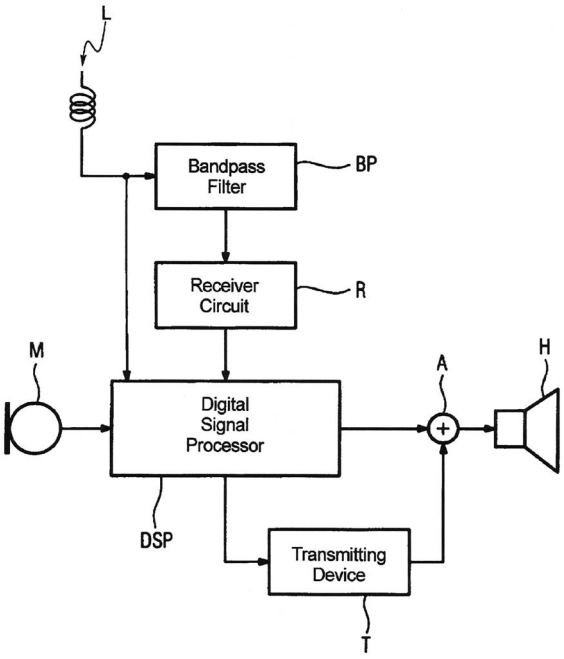
7,433,480

43.66.Ts HEARING AID WITH WIRELESS TRANSMISSION SYSTEM, AND OPERATING METHOD THEREFOR

Torsten Niederdränk, assignor to Siemens Audiologische Technik GmbH
7 October 2008 (Class 381/312); filed in Germany 1 December 2003

To facilitate wireless communication in a hearing aid without having to add an antenna, the telecoil and hearing aid receiver are used for their usual functions of receiving inductive audio signals and converting electrical signals to acoustic energy, respectively, and also to transmit and receive

electromagnetic signals at frequencies above the audible spectrum in order to communicate with an external device.—DAP

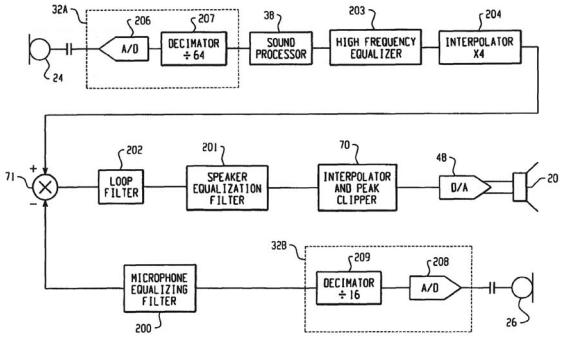


7,433,481

43.66.Ts DIGITAL HEARING AID SYSTEM

Stephen W. Armstrong *et al.*, assignors to Sound Design Technologies, Limited
7 October 2008 (Class 381/312); filed 13 June 2005

To mitigate occlusion resulting from the hearing aid wearer's own voice, two microphones are used in a directional hearing aid with digital signal processing. The output of the microphone that is placed in the ear canal, which is the occlusion signal, is equalized and subtracted from the processed intended audio signal picked up by the other microphone.—DAP



7,433,490

43.70.Gr SYSTEM AND METHOD FOR REAL TIME LIP SYNCHRONIZATION

Ying Huang *et al.*, assignors to Microsoft Corporation
7 October 2008 (Class 382/100); filed 16 May 2006

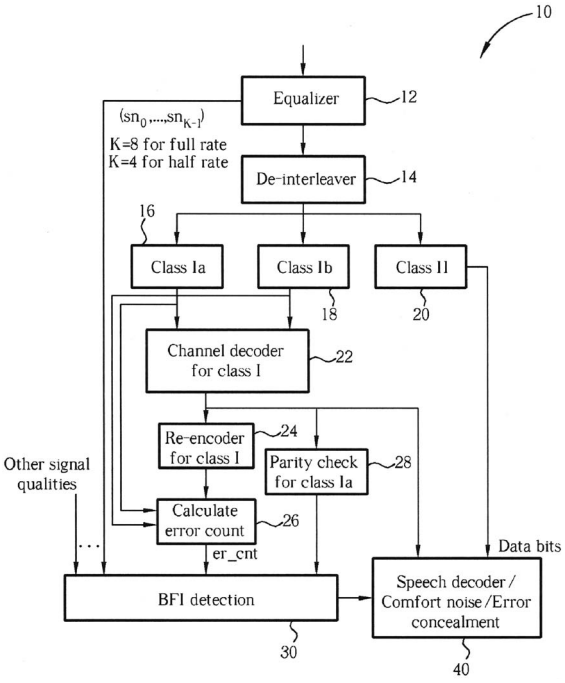
A method to synchronize lip shapes on a sketched face with an input audio signal is put forward.—GFE

7,434,117

43.72.Gy METHOD AND APPARATUS OF DETERMINING BAD FRAME INDICATION FOR SPEECH SERVICE IN A WIRELESS COMMUNICATION SYSTEM

Pei-Shiun Chung and Wei-Nan Sun, assignors to MediaTek Incorporated
7 October 2008 (Class 714/704); filed 28 October 2005

Corruption in transmitted blocks of speech information in a wireless communication system is detected and quantified as a signal quality by averaging the burst-by-burst signal-to-noise ratio. If the number of bit errors in the received speech block is too large due to discontinuous transmission, a bad frame indication for that block is generated. Depending on the level of signal quality, the speech block will be decoded, the errors will be concealed, or noise will be injected in place of the corrupted speech.—DAP



7,437,299

43.72.Gy CODING OF STEREO SIGNALS

Ronaldus Maria Aarts and Roy Irwan, assignors to Koninklijke Philips Electronics N.V.
14 October 2008 (Class 704/500); filed in the European Patent Office 10 April 2002

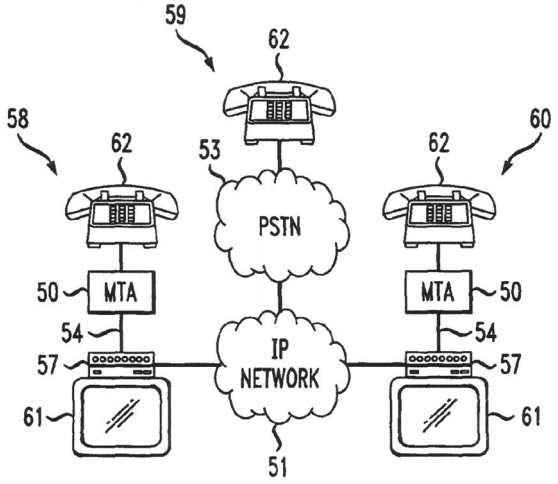
Reduction of bit rate at a given sound quality is said to be obtained by transforming a multichannnel signal (e.g., stereo) into a principal signal that contains most of the signal energy and at least one lower-energy residual signal. The transformation may comprise mapping the left and right signals of a stereo signal into different angles of rotation corresponding to the amount of signal variance, and changing adaptively the number of bits and bit rate allocated to the principal and residual signals.—DAP

7,433,818

43.72.Ne SUBSCRIBER TERMINAL FOR PROVIDING SPEECH-TEXT ENCODING AND TELEPHONY SERVICE

Charles David Caldwell *et al.*, assignors to AT&T Corporation
7 October 2008 (Class 704/235); filed 1 February 2006

This invention is said to eliminate the need for a live operator to transcribe speech into text using a Teletype machine in order to provide a third-party telecommunication relay service (TRS) for severely hearing-impaired persons. A subscriber terminal has a speech recognizer that is activated via dual tone multiple frequency (DTMF) tones from a user, a network communication interface to receive speech packets, a visual display interface, a telephone interface, and a processor to decode and display the speech as text.—DAP



7,437,295

43.72.Ne NATURAL LANGUAGE PROCESSING FOR A LOCATION-BASED SERVICES SYSTEM

Ashton F. Pitts III *et al.*, assignors to Accenture LLP
14 October 2008 (Class 704/275); filed 25 April 2002

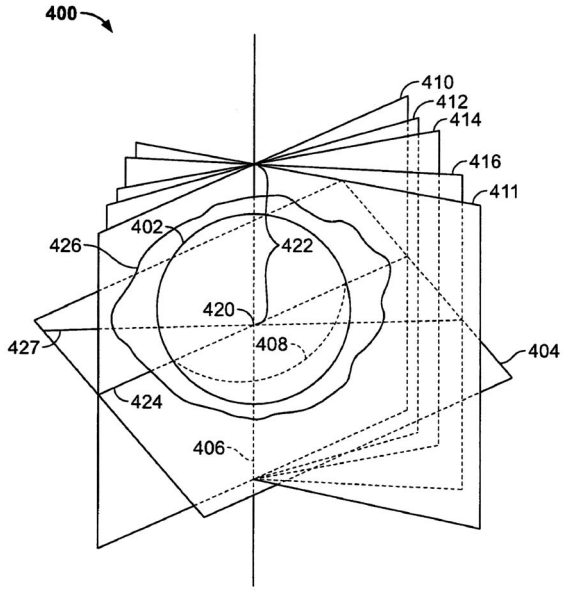
This patented system for speech recognition is geared toward the specific needs and requirements of use in a personal mobile wireless device. The recognition server software runs in a base system rather than in the user's device. It does not attempt do identify 100% of the words in the utterance. Instead, each word that is identified triggers a word-specific grammar. All of the grammars thus selected are combined to form an initial mapping of the utterance. All available grammatical information plus GPS-based location information is then used to identify a range of products and services which might fit the user's inquiry. A response is then constructed in either speech or text and sent to the user's device.—DLR

7,428,334

43.80.Vj METHODS AND SYSTEMS FOR 3D SEGMENTATION OF ULTRASOUND IMAGES

Armin Schoisswohl and Johannes Ruisz, assignors to General Electric Company
23 September 2008 (Class 382/173); filed 27 August 2004

Ultrasonic echoes encompassing the volume of an object are segmented from other echoes. In the segmentation process, a reference cross section and multiple other cross sections through the object echoes are defined. Points at the edge of the object in the reference cross section are determined and a contour that intersects the cross sections is found. The contour is adjusted until the contour points contain essentially all of the echoes from the object.—RCW



7,431,698

43.80.Vj APPARATUS AND METHOD FOR CONTROLLING AN ULTRASOUND PROBE

Reinhold Bruestle, assignors to GE Medical Systems Global Technology Company, LLC
7 October 2008 (Class 600/459); filed 13 January 2004

Memory and control circuitry are included in an ultrasound probe to connect a set of signal lines to different elements in a movable transducer array.—RCW